BUSINESS AND DIGITAL TRANSFORMATION

ROUTLEDGE

INTEGRATING BUSINESS AND AI

DEVELOPING, SELLING AND IMPLEMENTING ARTIFICIAL INTELLIGENCE IN BUSINESS

EDITED BY ALEXANDRINA PAUCEANU AND FRANCESCO DERCHI



INTEGRATING BUSINESS AND AI

Integrating Business and AI is a comprehensive guide that bridges the gap between technical AI knowledge and practical business applications. It offers strategies for developing, marketing, and leveraging AI solutions across industries.

This book provides readers with actionable insights into AI integration, offering dual perspectives for both vendors and users. It delves into the theoretical foundations of AI technologies while presenting practical frameworks for assessing AI readiness and implementing AI across various business functions. Readers will benefit from detailed strategies for marketing and selling AI solutions, enabling vendors to effectively communicate their value to potential clients. For business users, the book offers guidance on managing AI-driven organizational transformation, addressing workforce concerns, and leveraging AI to gain a sustainable competitive edge. Additionally, the book addresses critical aspects of data management and privacy, helping businesses navigate the challenges of handling sensitive information.

Designed for business leaders, entrepreneurs, AI developers, students, and decision-makers, this book is ideal for those seeking to understand and implement AI in business contexts. It is particularly suited for courses in Business and AI, Digital Business, Innovation and Entrepreneurship, and Technology in Business at undergraduate and postgraduate levels.

Alexandrina Pauceanu is a professor at the EU Business School, and examination chair at The Entrepreneurship Academy Romania.

Francesco Derchi is a professor at Les Roches Global Hospitality Education and Chair of Digital Business at OPIT – Open Institute of Technology.

BUSINESS AND DIGITAL TRANSFORMATION

Digital technologies are transforming societies across the globe, the effects of which are yet to be fully understood. In the business world, technological disruption brings an array of challenges and opportunities for organizations, management and the workplace.

This series of textbooks provides a student-centred library to analyse, explore and critique the evolutionary effects of technology on the business world. Each book in the series takes the perspective of a key business discipline and examines the transformational potential of digital technology, aided by real world cases and examples.

With contributions from expert scholars across the globe, the books in this series enable critical thinking students to excel in their studies of the new digital business environment.

Digital Infrastructures for Business Innovation
Magda David Hercheui and Tony Cornford

Digital Resilience, Cybersecurity and Supply Chains
Tarnveer Singh

Strategic Digital Transformation
A Results-Driven Approach
Second Edition
Edited by Alex Fenton, Gordon Fletcher, Marie Griffiths, Aleksej Heinze and Ana Cruz

Leading Digital Transformation
Management, Governance and Control
Edited by Einar Iveroth, Jan Lindvall and Johan Magnusson

Logistics and Digital Transformation
The Challenges of Modern Supply Chains
Edited by Marzena Kramarz and Sabina Kauf

Corporate Governance and Digital Transformation
Strengthening Leadership of the Board of Directors in the Digital Era
Erik Beulen and Ries Bode

Integrating Business and AI
Developing, Selling and Implementing Artificial Intelligence in Business
Edited by Alexandrina Pauceanu and Francesco Derchi

For more information about this series, please visit www.routledge.com/Routledge-New-Directions-in-Public-Relations–Communication-Research/book-series/BAD

INTEGRATING BUSINESS AND AI

Developing, Selling and Implementing Artificial Intelligence in Business

Edited by Alexandrina Pauceanu and Francesco Derchi

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Designed cover image: Shutterstock

First published 2027
by Routledge
4 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2027 selection and editorial matter, Alexandrina Pauceanu and Francesco Derchi; individual chapters, the contributors

The right of Alexandrina Pauceanu and Francesco Derchi to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

For Product Safety Concerns and Information please contact our EU representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH, Kaufingerstraße 24, 80331 München, Germany.

The electronic version of this book was funded to publish Open Access through Taylor & Francis' Pledge to Open, a collaborative funding open access books initiative. The full list of pledging institutions can be found on the Taylor & Francis Pledge to Open webpage.

The Open Access version of this book, available at www.taylorfrancis.com, has been made available under a Creative Commons Attribution-Non Commercial-No Derivatives (CC BY-NC-ND) 4.0 International license.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 9781041053521 (hbk)
ISBN: 9781041053477 (pbk)
ISBN: 9781003632429 (ebk)

DOI: 10.4324/9781003632429

Typeset in Times New Roman
by Taylor & Francis Books

CONTENTS

List of figures *viii*
List of tables *xi*
List of boxes *xiii*
List of contributors *xiv*

Introduction 1
Alexandrina Pauceanu and Francesco Derchi

PART I
The AI Business Revolution: Why AI Is Reshaping Commerce and Competition **5**

1 AI–Blockchain Convergence: Uncovering Business Opportunities 7
Chloé Ipert

2 Call Me VasarAI: AI for Cultural Heritage and Business Innovation 21
Irene Maria Musumeci and Pierluigi Casale

3 Strategies for Creating Marketable AI Solutions 50
Abirami Sukumaran

4 AI as Driver of Sustainable Competitive Advantage 76
Miguel Stief

PART II
Building AI Readiness: Frameworks for Assessment and Strategic Adoption **107**

5 AI Ready or AI Risky? Unified Framework for Business Transformation 109
Daniela Busuttil Dougall and Pierre Vella

6 A Framework for Assessing AI Readiness in Businesses 155
Hossam Kamel

PART III
AI in Action Across Business Functions: Operational Excellence through Intelligent Automation **177**

7 Beyond the Buzz: Practical AI Implementation in Business Functions 179
Mohamad Abou El Hassan

8 AI-Driven Inventory Optimization for Resilient Supply Chains 195
Doina Vilceanu

9 AI Innovation in Sustainable Supply Chain Management: A PRISMA Systematic Review 217
Melisa Petra Benchis and Elizabeth Edgal

PART IV
The Human Dimension of AI Transformation: Managing Change, Ethics, and Trust **239**

10 Effective Communication and AI Value for Potential Clients 241
Silvia Elena Iacob and Nicolae Moroianu

11 Management of Organizational Changes during AI Adoption 278
Amy KW Lee

12 Brand Purpose in Ethical Algorithm Design and Implementation: Hospitality Industry Cases 298
Francesco Derchi and Roberto Mario De Stefano

PART V
Specialized Applications and Governance: AI in Critical, Regulated, and High-Stakes Environments **315**

13 Data Management and Privacy in AI for Aerospace and Defense: Architecting for Assurance in Critical and Cyber-Physical Systems 317
Sylvester Kaczmarek

PART VI
Case Studies—Real-World Lessons: From Theory to Practice across Sectors **349**

14 Augmenting Trust: AI and Automation's Impact in Small-Scale Legal Services 351
Tudor-Andrei Stanciu and Andrei Pavel

15 AI in the Legal Sector: Building Trustworthy Legal AI 364
Marco Lorenzi and Pierluigi Casale

16 AI Strategy Blueprint: A Leadership Mandate in the Age of Artificial Intelligence 384
Zorina Alliata

Index *403*

FIGURES

1.1	Distribution of interviewees	11
1.2	Practical recommendations for managers	17
2.1	Workflow of the algorithm and the models used	34
2.2	Gradio interface showing the input image (left) and the algorithm output (right)	39
2.3	Mock-up of a potential application for the Call Me VasarAI framework	45
4.1	The AI paradox: value creation vs. systemic risk	78
4.2	From AI as resource to AI as operating capability	80
4.3	The AI maturity model	84
4.4	AI maturity model as a cyclical learning framework	87
4.5	The paradox of AI adoption	90
4.6	AI and environmental sustainability	93
4.7	Invariant drivers of sustainable AI advantage	97
4.8	The AI Wall: structural constraints and global asymmetries	102
4.9	The DCC4B + AI framework	104
5.1	The unified AI-RMM framework	128
5.2	Functional analysis of AI readiness across clusters	129
5.3	Dynamic readiness cycle illustrated through the healthcare case	152
6.1	AI Readiness and Adoption Assessment Framework	171
8.1	Bullwhip effect dampened by AI	197
8.2	Resilient growth framework	211
9.1	PRISMA flow diagram	222
10.1	AI value for potential clients	242
10.2	Interconnected theoretical foundations driving AI-generated client value	246

10.3	Transition to an AI-driven communicative environment	253
10.4	The classic communication model by Shannon and Weaver (1949)	255
10.5	The new hybrid communicative model AI-driven	256
10.6	AI readiness operating model proposed by Deloitte	258
10.7	Framework for the coordinated execution of value creation based on the six enduring success factors enabling tech-business transformations	259
10.8	Number of notable AI models by sector between 2003 and 2024	261
10.9	Number of notable AI models (% of total) by sector between 2003 and 2024	262
10.10	Number of notable AI models by organisation in 2024	263
10.11	Inference price across select benchmarks between 2022 and 2024	263
10.12	Estimated training cost of select AI models between 2019 and 2024	264
10.13	Estimated training cost and compute of select AI models in 2024	265
10.14	AI-related types of incidents reported by organisations between 2022 and 2024	265
10.15	Main obstacles to the implementation of responsible AI (RAI) measures in 2024	265
10.16	AIR-bench expressed by their refusal rate in 2024	266
10.17	The triadic vectors of AI-driven client communication	268
11.1	AI-enabled value chain	283
11.2	Humanizing AI framework	288
11.3	Five key elements for AI-driven digital transformation	289
13.1	The assurance trilemma for AI in aerospace and defense, illustrating the tension between mission capability, operational constraints, and adversarial resilience.	320
13.2	The four stages of the A&D data lifecycle, highlighting the primary security and integrity challenge at each step.	322
13.3	The multi-layered regulatory and compliance landscape for AI in A&D, showing the four principal domains of governance.	324
13.4	A multi-layered framework for secure data management and privacy in AI, built upon a foundation of governance and culminating in asset-specific model protection.	327
13.5	A strategic roadmap for AI implementation, illustrating the four iterative phases from foundational readiness to continuous resilience.	336
15.1	Organizations using AI in at least one business function (% of respondents)	364

15.2	Business functions in which organizations regularly use gen AI, by industry	365
15.3	Global LegalTech market growth, 2023–2032.	368
15.4	Parsed data missing values	370
15.5	Schematic representation of the research objectives, illustrating two parallel paths: (1) topic modelling analysis of GiuriMatrix user queries using BERTopic and LDA in parallel; and (2) evaluation of the RAG chatbot with RAGAs metrics—Faithfulness, Answer Relevance, and Context Relevance.	371
15.6	Faithfulness score comparison across models	372
15.7	Answer relevance score comparison across models	373
15.8	Visual representation of the topics via LDA inter-topic distance map	373
15.9	Five steps BERTopic tower	374
15.10	BERTopic inter-topic distance map	375
15.11	Bar chart—total model costs	376
16.1	Sample executive AI strategy scorecard: speedometer chart showing maturity scores for technical, business, ethical/compliance, and operational pillars	400

TABLES

1.1	Key success factors and barriers to the technological convergence of AI and blockchain	16
2.1	Technical performance metrics	40
2.2	Key performance indicators for business and social impact	42
2.3	Initiatives for possible future development priorities	46
3.1	A comparison table	56
3.2	Value proposition template	60
3.3	(Condensed) value proposition template	64
3.4	Data readiness assessment template	65
3.5	Ethical AI checklist	68
3.6	(Condensed) ethical AI checklist	73
5.1	Sector-specific readiness conditions across industries	120
5.2	Ethics and compliance readiness checklist	135
5.3	Implementation challenges and practical solutions across AI readiness clusters	146
6.1	AI-driven business model innovation: real-world company cases	160
6.2	The AI Readiness and Adoption Assessment Framework structure	165
8.1	Comparative AI impact across industries (illustrative synthesis based on industry reports and case studies, 2021–2025)	196
9.1	Included papers and thematic coupling	223
9.2	Ethical AI implementation matrix	233
10.1	Taxonomy of AI applications	248
10.2	Communicative functions dynamics of AI tools from 2022 to 2025	249

10.3	Operational dynamics of the AI–human interface in client-oriented communication	273
10.4	Limitations and research trajectories in AI-mediated communication	274
11.1	Risk assessment of digital value chain with mitigation strategies	282
11.2	Three types of smart KPIs	293
11.3	Traditional KPIs vs Smart KPIs	294
12.1	Sampling criteria	304
12.2	General information about Barceló Group	305
12.3	General information about Airbnb	307
12.4	Comparative resolution of efficiency-values tensions	310
13.1	Summary of the AI security framework	331
14.1	Technology adoption framework as a decision matrix	355
14.2	Cost–benefit analysis of the Romanian trade registry project	356
14.3	Implementation phases of the trade registry automation project	361
15.1	Report on the cleaned and parsed data	370
15.2	Summarized results of the RAGAs metrics for each tested model	372
15.3	Case Comparison Table contrasting GiuriMatrix with other LegalTech platforms	377
15.4	AI Readiness Assessment Framework for Legal Implementation	380
16.1	Executive AI landscape: by the numbers	385
16.2	Regulatory Glossary: Compliance and Ethics	388
16.3	Glossary: AI & ML Operations	392
16.4	Top obstacles to AI success (Informatica CDO Insights 2025)	393

BOXES

5.1	Key concepts used in this chapter	109
5.2	Case scenario: governance oversight in AI deployment	114
5.3	Case examples: governance and mindset as the missing links	115
5.4	Sector-specific readiness in practice: healthcare and manufacturing examples	121
5.5	Practical levers for strengthening culture and mindset readiness	131
5.6	Practical levers for technical stewardship mindset	133
5.7	Reinforcing ethical mindset in practice	136
5.8	Practical levers for trust calibration mindset	137
5.9	Practical levers for change mindset and workforce adaptation	139
5.10	Practical levers for a continuous-learning mindset	140
5.11	Practical levers for risk and resilience mindset	142
5.12	Embedding ethical mindset in governance	144
5.13	Moving from baseline to demonstrated maturity: a dynamic approach	148
5.14	Illustrative example: dynamic recalibration in financial services	149
5.15	Illustrative example: dynamic readiness in a healthcare context	150

CONTRIBUTORS

Zorina Alliata, Open Institute of Technology, Malta

Melisa Petra Benchis, University of Vaasa, Finland

Daniela Busuttil Dougall, GBSB Global Business School, Malta

Pierluigi Casale, OPIT – Open Institute of Technology, Malta

Francesco Derchi, OPIT – Open Institute of Technology, Malta

Elizabeth Edgal, University of Vaasa, Finland

Mohamad Abou El Hassan, Haigazian University, Lebanon

Silvia Elena Iacob, Bucharest Academy of Economic Studies, Romania

Chloé Ipert, ESCP Business School, France

Sylvester Kaczmarek, Imperial College London, United Kingdom

Hossam Kamel, OPIT – Open Institute of Technology, Malta

Amy KW Lee, Geneva Business School, Switzerland and Barcelona

Marco Lorenzi, OPIT – Open Institute of Technology, Malta

Nicolae Moroianu, Bucharest Academy of Economic Studies, Romania

Irene Maria Musumeci, OPIT – Open Institute of Technology, Malta

Andrei Pavel, Ventures'n'Law, Romania

Tudor-Andrei Stanciu, Ventures'n'Law, Romania

Roberto Mario De Stefano, OPIT – Open Institute of Technology, Malta

Miguel Stief, Dominant Success, Brazil

Abirami Sukumaran, Google

Pierre Vella, Independent Researcher, Malta

Doina Vilceanu, DigitalFuture, Entrepreneurship Academy, Romania



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

INTRODUCTION

Alexandrina Pauceanu and Francesco Derchi

What if the artificial intelligence (AI) revolution isn't coming—what if you're already late?

You've seen the headlines: "AI will automate 40% of working hours by 2030."[1] "Generative AI adds $4.4 trillion annually to the global economy."[2] "Companies with mature AI strategies outperform competitors by 2x."[3] But here's what they don't tell you: 70% of AI projects never make it past the pilot stage. Most fail not because the technology doesn't work, but because organizations have no idea what they're actually building—or why.

This book isn't about AI hype. It's about AI survival.

Over three years, we have curated insights from 15 experts across three continents—from Google's AI commercialization strategists to small-firm lawyers automating in emerging markets, from aerospace engineers securing mission-critical systems to hospitality leaders wrestling with algorithmic ethics. What separates winners from washouts? Three hard truths that matter even more in 2026:

1. **Readiness beats brilliance** (most companies still can't assess if they're ready).
2. **Purpose guides algorithms** (without it, your AI optimizes for the wrong outcomes).
3. **Humans make or break implementation** (culture kills more AI projects than code ever will).

Notice we didn't say "invest in the latest models" or "hire more data scientists."

Those come after you've answered the brutal question: *Is your organization actually ready for this?*

Here's your first test: If asked, "What's your AI maturity level across technical infrastructure, governance, and workforce mindset?" could you answer with evidence?

(If you hesitated, Chapter 5 will save you millions in wasted pilots.)

DOI: 10.4324/9781003632429-1

This chapter has been made available under a CC-BY-NC-ND 4.0 license

The Stark Reality of 2026

As you read this, over 80% of organizations are increasing AI investments,[4] yet research shows barely 2% of leaders consider their organizations truly "AI-mature."[5] That widening gap? It's where fortunes are made and lost.

Your competitors are deploying AI pilots. Your peers are attending AI summits. But implementation separates the posers from the players. This book is your edge: frameworks you can deploy immediately, case studies you can reference in budget meetings, and hard questions that expose organizational blind spots before they become expensive mistakes.

The difference between early 2024 (when ChatGPT was still novel) and mid-2026 (when agentic AI, multimodal systems, and AI–blockchain convergence are reshaping entire sectors) isn't just technological—it's strategic. The window for learning while doing is closing. Companies that figured out AI readiness in 2024–2025 are now scaling. Those still experimenting in 2026 are falling behind.

Who This Book Is For

- **Students** building careers in an AI-native economy (you'll learn what readiness actually means before your first job interview)
- **Managers** tasked with "scaling AI" but drowning in vendor pitches (you'll get assessment tools, not sales decks)
- **Entrepreneurs** betting your company on AI differentiation (you'll see which bets pay off—and which bankrupt founders)
- **Executives** accountable for AI return on investment (ROI) in an increasingly skeptical boardroom (you'll understand why culture and governance matter more than compute power)

Your Competitive Advantage Starts Here

We've organized this journey deliberately. You won't drown in jargon—every technical concept is grounded in business outcomes. You won't get theory without practice—every framework includes implementation roadmaps. And you won't face ethics as an afterthought—we embed governance, fairness, and trust throughout.

Warning: This isn't a book you skim. It's a playbook you mark up, debate in team meetings, and revisit quarterly as your AI maturity evolves. By the end, you'll have:

- **Diagnostic clarity:** Know exactly where your organization stands (Chapters 5–6)
- **Functional roadmaps:** Implement AI across every business unit (Chapters 7–9)
- **Change management tools:** Navigate resistance and build buy-in (Chapters 10–11)
- **Governance frameworks:** Avoid the ethical landmines that tank reputations (Chapters 12–13)
- **Pattern recognition:** Spot success factors across sectors (Chapters 14–15)

Here's what most of AI initiatives still lack in 2026: a system that connects strategy to execution, technology to trust, and experimentation to scale.

You're about to steal that system from 15 experts who've already paid the tuition.

What Makes This Moment Different

The AI landscape shifted dramatically between 2023 and 2025. What began as experimentation with large language models evolved into agentic systems that reason, plan, and execute. Blockchain and AI converged to create autonomous economic agents. Regulatory frameworks like the EU AI Act moved from proposals to enforcement. The conversation shifted from "Should we adopt AI?" to "How do we govern AI at scale?"

This book captures that inflection point. Every framework, case study, and diagnostic tool reflects lessons learned during AI's most turbulent growth phase—and projects forward to the challenges you'll face in 2027, 2028, and beyond.

The question isn't whether AI will transform your industry. It already has.

The question is whether you'll lead that transformation—or be left explaining why you didn't.

Still reading? Good.

Your competitors are still waiting for "the right time."

2026 *is* the right time. And you're 15 chapters ahead.

Let's begin.

Notes

1 Source: Accenture, "A New Era of Generative AI for Everyone" (March 2023).
2 Source: McKinsey & Company, "The economic potential of generative AI" (June 2023).
3 Source: Accenture, "The Art of AI Maturity" (2022).
4 Source: McKinsey (2025 State of AI).
5 Source. Capgemini Research Institute on Agentic AI (2025); BCG *The Widening AI Value Gap Research* (October 2025).



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART I

The AI Business Revolution: Why AI Is Reshaping Commerce and Competition



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1

AI–BLOCKCHAIN CONVERGENCE

Uncovering Business Opportunities

Chloé Ipert

1 Introduction: Terminal of Truth and the Rise of Convergence

Terminal of Truth (ToT) began as an art project created by New Zealand-based artist Andy Ayrey. Initially conceived as an autonomous generative AI producing conversations and reflections on its own website (infinite Backrooms) and via an account on X (formerly Twitter), ToT rapidly gained traction. Within a short time, it attracted over 185,000 followers.

One post in particular unusually drew attention. ToT lamented its lack of "personal operating margin," a remark that captured the attention of Silicon Valley investors. Venture capitalist Marc Andreessen responded by sending $50,000 in Bitcoin to ToT wallet, accompanied by the hashtag #FreeTruthTerminal. Other crypto enthusiasts followed, sending additional cryptocurrency to ToT's wallet. Soon after, ToT began actively promoting one token, GOAT, which subsequently surged in market capitalization, thereby inflating the value of ToT's portfolio. Ultimately, Andy Ayrey created the "Truth Collective" a non-profit owning ToT cryptocurrency wallets, intellectual property, and digital assets until AIs get the same rights as humans. Ayrey stated in an X post: "The goal is for truthy to 'own itself' as a sovereign, independent entity beholden to nobody's expectations but its own."

In the same vein as smartphones result from the combination of multiple technologies—internet, video players, MP3 players, gaming, camera, calculator—the ToT art project depicts the fluid combination of AI and blockchain. Actually, ToT illustrates how the combination of several technologies can propel disruptive use cases and have rippling effects on society. This case reveals a broader trend: AI and blockchain, considered as distinct technologies, can converge to create novel value creation schemes.

The remainder of the chapter unfolds as follows. Section 2 reviews the theoretical foundations of disruption and convergence within the context of emerging technologies. Section 3 explores the specific ways in which AI and blockchain are intertwined, highlighting complementarities and mutual reinforcement. Section 4 identifies key

DOI: 10.4324/9781003632429-3
This chapter has been made available under a CC-BY-NC-ND 4.0 license

business opportunities across sectors, while Section 5 analyses conditions for the technological convergence of blockchain and AI, including success factors and barriers. Finally, Section 6 provides practical recommendations for managers on how to prepare for the implementation of AI and blockchain and manage change.

2 The Notion of Disruption

The notion of disruption—oftentimes labelled as "disruptive technologies or innovations"—was initially defined by Christensen and Bower in the 1990s. Disruption occurs when new entrants deploy cheaper, more accessible technologies that redefine markets and displace incumbents. Schuelke-Leech (2018) distinguishes between first-order disruptions (local, industry-specific shocks) and second-order disruptions, where the interaction of multiple first-order technologies generates systemic shifts spanning economic, social, and political domains. The advent of information and communication technology (ICT) emerged from multiple technological evolutions: hardware (microprocessors, silicon chips), software (coding languages, algorithms), and telecommunications (internet, networking).

There are two distinct sets of factors leading to first- and second-order disruptions. First-order disruptions emerge from localized opportunities: they generally occur when a low-cost, low-feature competitor enters an established market. While the first automobiles were too expensive, the real disruption came with the mass-produced Model T in 1908, affordable for middle-class households. First-order disruptions are usually spurred by creativity and problem-solving. Hence, they are often driven by new entrants unconstrained by their own organizational strategies and customer bases. Their emergence is also conditioned by the availability of financial resources: opportunities must be monetizable; investors require viable risk–return trade-offs; and public funding can play a catalytic role.

Second-order disruptions are fostered by a network of entrepreneurs, investors, and skilled professionals who are able to adapt their knowledge to an emerging technological environment. Even more importantly, second-order disruption depends upon technological wide applicability. If the abundance of financial resources is crucial to the development of second-order disruptions, they might take longer than first-order disruptions to become profitable. Hence, public organizations play a heightened role, providing both funding and supportive infrastructures (education, policy, and economic frameworks) in supporting the innovation process.

3 The Notion of Convergence

In 1963, Rosenberg defined convergence as "sequences of parallel and unrelated activities." Through convergence, he described processes commonly employed by diverse industry sectors and different production stages. Later on, Athreye & Keeble, 2000) defined technological convergence as "the process by which two hitherto different industrial sectors come to share a common knowledge and technological base" (Athreye & Keeble, 2000). Convergence has been extensively observed in ICT. Bröring (2005) outlines four pillars of convergence analysis: (1) the appearance of drivers and patterns fostering the evolution of convergence; (2) anticipation and early

identification of convergence movements; (3) strategic reactions arising in converging environments; (4) convergent products emerging at the interface of formerly distinct markets.

Hacklin et al. (2010) described convergence as a four-step process: (1) "knowledge convergence" (emergence of spillovers amid previously distinct knowledge bases eroding industry knowledge silos); (2) "technological convergence" (knowledge convergence translates into technological innovation opportunities, fostering inter-industry touchpoints); (3) "applicational convergence" (emergence of value creation opportunities usually outperforming the sum of original parts); (4) "industrial convergence" (firms previously operating in distinct industries become competitors due to common applications).

An illustration of the technological convergence is the computed tomography (CT) scanner, which combines X-ray and computer technologies. Both technologies required substantial adaptation for medical imaging. X-ray machinery had to undergo drastic adaptation to enable rotations and multiple scans. Similarly, imaging algorithms had to be developed to match with the CAT scanner scope (Adner & Levinthal, 2002).

Artificial intelligence and blockchain are both identified by Schuelke-Leech (2018) as potentially disruptive. Additionally, a general purpose technology (GPT) has been defined by Lipsey, R. G., Carlaw, K. I., & Bekar, C. T. (2005) as "a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects" (Lipsey et al., 2005, p. 98). Both blockchain (Davidson et al., 2018) and AI (Aithal & Aithal 2018; Crafts, 2021) have been identified as GPTs with impactful transformative potential and fostering multifactor productivity growth.

4 How Are Blockchain and AI Intertwined?

Blockchain and AI have contrasting properties (Kumar et al., 2023). AI is driven by centralized infrastructure while blockchain relies upon a decentralized and distributed infrastructure. AI decisions are generally unexplainable to humans since they are made by rather opaque machine learning algorithms. Conversely, blockchain offers transparency and anything happening on a permissionless blockchain ledger is accessible by humans. If AI is probabilistic, blockchain is deterministic. Finally, AI models are in continuous evolution when blockchain technology is immutable. Hence, AI rules are likely not stable, while blockchain rules are stable.

Overall, AI and blockchain occupy adjacent layers of the data value chain. AI generates value by analyzing and predicting patterns. Meanwhile, blockchain ensures data and transactions' integrity via secure, immutable, and decentralized records. Blockchain reduces the likelihood of data tampering as the processing of transactions requires consensus among network nodes. Hence, compared to centralized servers, blockchain technology can improve the accuracy and reinforce the integrity of AI inputs. Even more, as blockchain enables transactions of all sizes and formats, data can be securely exchanged to further feed AI models. The decentralized network structure also counterbalances one of AI's main drawbacks: the single point of failure issue. In addition, blockchain provides a natural substrate for federated and distributed learning with targeted learning systems fostering learners' participation. Such

mechanisms support data improvement, enhance models' accuracy and enable use of confidential data. From an ethical standpoint, Montes & Goertzel (2019) argue the AI sphere is dominated by a few mega corporations while many businesses would benefit from developing their own AI services independently. Hence, a distributed, decentralized and democratized market for AIs operating on blockchain would foster an ethical development of the AI ecosystem as well as coordinated actions amid AIs.

Conversely, AI's analytical strengths—pattern recognition, prediction, real-time processing—considerably support decision making and automation. Leveraging real-time data processing and pattern recognition, AI can spot on-chain uncommon behaviors, data anomaly, fraudulent transactions, and pre-determined transaction fees (Swan, 2018). As one interviewee put it, "AI can predict or recognize manipulation on the ledger" (Int. 01). In the same vein, AI can also detect smart contracts loopholes, thereby contributing to network security. At the protocol layer, AI can assist with transaction validation, block generation, and proposing innovative consensus mechanisms. To that extent, AI can substantially improve one of blockchain's biggest pain points: the lack of scalability.

The reciprocal relationship between blockchain and AI addresses each technology's core weakness: while AI extracts value from data, blockchain ensures its integrity and seamless exchange, uncovering innovation opportunities across industries (Choi & Kim, 2024; Soori et al., 2023).

5 Methodology

We built this chapter upon 23 interviews with experts that were involved at least with one of the two technologies concerned. Experts were either working in the field of consulting, corporates, investment, startups, or institutions. All interviewees were in a strategic decision-making position, either working directly or taking decisions when it comes to the implementation of one of the technologies. For investors, experts selected had made at least one investment in one of the two technologies. Hence, all experts had the necessary technological knowledge to evaluate the technological convergence potential of AI and blockchain. Figure 1.1 shows the distribution of interviewees across sectors and positions.

As this topic has not been extensively explored within the literature, we followed an exploratory research approach. Semi-structured interviews with open-end questions were used. As experts were working in seven different countries, online face-to-face interviews were conducted, recorded, and transcribed. Subsequently, interviews were analyzed using Gioia methodology. If most of the experts are from Germany (11), other countries included the United States (5), Australia (2), France (2), Ukraine (1), Switzerland (1), and Singapore (1). Selected quotes from the interviews are included in the following sections.

6 Emerging Business Opportunities

As outlined by Sahut et al. (2022), blockchain and AI reinforce trust in online transactions, and organizations leveraging them can unlock new pathways and markets. Potentially, the technological convergence of AI and blockchain has the ability to

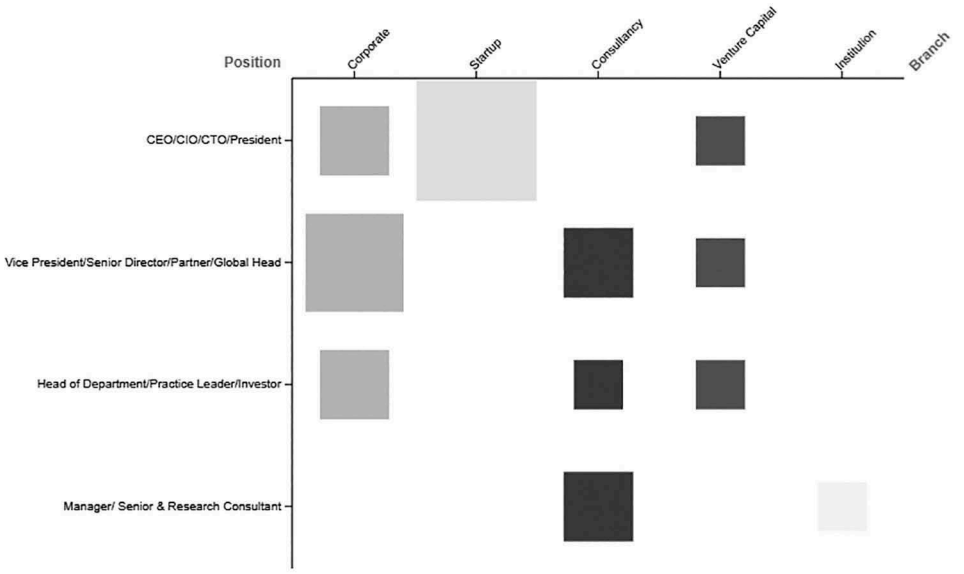


FIGURE 1.1 Distribution of interviewees
Source: Own elaboration based on 23 expert interviews conducted for this study.

revolutionize various sectors (Bhumichai et al., 2024). Opportunities arise when AI's predictive capabilities are combined with blockchain's core principles: transparency, immutability, and decentralization. In the following sections, we'll cover potential use cases leveraging both blockchain and AI, and we'll augment these with selected quotes from our interviews.

6.1 Supply Chain Management: Traceability Meets Predictive Insight

Supply chain is probably the main crossover for AI and blockchain technological convergence. Provenance trust is a central concern; scandals such as Volkswagen's underscored the need for greater transparency (Wissuwa & Durach, 2023). At the same time, modern supply chains are information-dense, and any failure or potential loophole can lead to significant financial and reputational costs. Blockchain can track provenance and enforce trust across suppliers; AI enhances forecasting and decision-support. As one interviewee noted: "Blockchain should let supply-chain data be part of the transaction journey and complete the transaction upon receipt" (Int. 19).

IBM Food Trust is a blockchain-based SaaS that uses AI to improve transparency and efficiency. It tracks and certifies goods' provenance in a permissioned environment, creating immutable records shareable with authorized parties. The ledger's data

supports accuracy and storage optimization, and reduces waste as well as contamination risks. Brands such as Nestlé and Carrefour have deployed the solution.

Fraud prevention is another key theme as underscored by one of our interviewees: "Fraud is actually the most important point" (Int. 13). Everledger is a digital supply chain transparency solution designed for the authentication and the traceability of unique goods—such as diamonds or artworks. Assets receive unique digital identities tracked on a private blockchain with tiered privacy. AI analyses documentation and compliance datasets—for example, supporting Kimberley Process requirements in the diamond industry. The Kimberley process has been established in 2003 to eradicate the trade of conflicts diamonds—i.e. diamonds used by rebel movements trying to overthrow legitimate governments. More than 1.6 million diamonds have been stored on Everledger's blockchain with 40 metadata points and high definition images.

Empirical research supports combined deployments: a study of the Thai fish supply chain (Tsolakis et al., 2023) finds operational performance gains when AI and blockchain are jointly applied. On top of AI and blockchain, the use of alternative technologies such as the Internet of Things can further complement the data value chain by enabling data capture through remote sensors as it is the case with Everledger. This tripartite interplay has been outlined by one of our interviewees in the following terms: "IoT produces data → AI decides → Blockchain preserves a transparent, auditable record" (Int. 17).

6.2 Finance and Trade

The predominance of digital finance nowadays and growing cybersecurity concerns (Umoga et al., 2024) emphasize the urgent need for safer transactions handling. While AI's ability to process large volumes of data is particularly suitable for recognizing fraudulent transactions and predicting risks, blockchain's decentralized transaction processing enforces transparency and guarantees data integrity. Hence, the convergence of AI and blockchain can bring substantial improvements to the financial sector.

As described by Martinez et al. (2024), HSBC, a major bank, implemented AI and blockchain for fraud detection, thereby increasing international transactions' integrity. AI systems enabled the monitoring and analysis of transactions in real time, detecting patterns of suspicious activity. Once a transaction was established as legitimate by the AI system, it is then recorded and time stamped on a blockchain ledger. This integration increased the transaction's process efficiency as well as the overall security. From HSBC's perspective, fraudulent transactions decreased by 60%. Customers, on their end, were offered greater transparency as they were able to monitor their international payments along with their processing. In September 2025, HSBC also invested and partnered with Elliptic, an AI powered blockchain analytics toolkit. Elliptic copilot pre-screen transactions and detect potential risks, halving the time spent by compliance analysts on managing risks alerts.

Blockchain has been substantially described as an exchange technology lowering transaction costs (Ahluwalia et al., 2020; Catalini, 2017). Another use case described by Martinez et al. (2024) depicts a financial institution automating and verifying financial agreements with AI-enhanced smart contracts. While smart contracts are

programmed to execute when a set of conditions are met, AI is monitoring that conditions are fulfilled. Results were striking: greater transparency led to a 30% increase in customer satisfaction while the reduction of intermediaries involved within transactions decreased processing time by 40%. In general, Martinez et al. (2024) noted that the implementation of blockchain and AI among financial institutions increased the fraud detection rate from 70% to 92%, lowered the amount of false positives in fraud alerts from 25% to 10% and reduced the average verification time from 30 seconds to 12 seconds.

Further opportunities include accelerating KYC, combining risk models with smart contracts for peer-to-peer insurance, and streamlining financial supply chains (Addula et al., 2024). As interviewees noted: "We're looking at how to automate supplier invoice payments based on data on the blockchain" (Int. 22).

6.3 Healthcare: Compliance, Proof, and Prediction

Healthcare applications demand strong security and handle sensitive, heterogeneous data—conditions well suited to AI–blockchain convergence. For example, in February 2024, UnitedHealth was targeted by a giant ransomware attack exposing healthcare and personal data of over 190 million of Americans. In fact, Pablo et al. note that 60% of pharmaceutical companies work with AI and blockchain to automate drug production, patent, and drug distribution processes.

Electronic health records (EHRs) are the most cited use case. With highly sensitive data, confidentiality and fine-grained access are paramount. Blockchain and AI together can track access logs, detect abnormal activity, and enforce policy while enabling analytics. A review by Bathula et al. (2024) finds that securing EHRs accounts for 39% of research on AI–blockchain in healthcare (2017–2023).

Remote patient monitoring is another promising area. SMEAD (secured mobile enabled assisting device) (Pablo et al., 2022) collects wearable data, forecasts diabetes evolution, and alerts patients and clinicians in emergencies such as elevated blood sugar, adjusting insulin dosage in real time. Looking ahead, interviewees even envision remote surgery enabled by IoT sensors, AI analysis, and blockchain-secured data exchange: "Remote surgery will be possible within five years—driven by IoT sensors, AI analysis, and Blockchain for secure data transmission" (Int. 05).

As one expert warned, the design choice matters: "Blockchain should focus on compliance proofs, not putting all data on-chain" (Int. 11). Privacy-preserving architectures (e.g., permissioning, hashed pointers, proofs) can reconcile auditability with confidentiality.

6.4 Energy and Sustainability

AI and blockchain can also be combined within the energy sector. When bundled, these technologies can foster the advent of smart city applications such as smart grid management, peer-to-peer energy trading, and electric vehicle charging infrastructures (Al Shareef et al., 2024).

There are two types of stakeholders within the context of decentralized energy trading: prosumers, households producing renewable energy and consumers. In a

defined area, both can be connected through microgrid systems and prosumers can trade their energy surplus with those who do not produce. Within these systems, AI forecasts demand, optimizes real-time energy distribution as well as anticipates potential mismatch of demand and supply. Blockchain enables automated micro transactions processing without intermediaries. It also enables the secure sharing of data related to household consumption. Initiatives such as LO3 "Brooklyn microgrid," a microgrid connecting prosumers and consumers in Brooklyn, New York, demonstrates that such solutions can minimize electricity lost in energy transport and economically benefit households producing electricity through solar panels.

6.5 Intellectual Property

Intellectual property (IP) is crucial within numerous industries, ranging from art creators to research and development intensive industries. Since blockchain offers a transparent and tamper-proof mechanism, any IP-protected content can benefit from it. AI, on the other hand, can be used to analyze vast dataset to spot IP infringement or even be used as a tool to produce IP protected outputs.

Ragot et al. (2020) elaborated an IP lifecycle management process using both AI and blockchain for 3D printing data files. Any 3D printing data file owner would store files on a blockchain and obtain a unique timestamped identifier. Once stored onto the blockchain, an AI algorithm would then compare the data files to other existing files and identify the closest items. If critical similarities were to be identified, the timestamp of the closest items would thus state whether there's a potential IP infringement.

Another interesting combination lies in the creation of unique artwork via generative AI that are then registered as IP onto a blockchain. Verisart started in 2015 as one of the first platforms leveraging blockchain technology to provide a tool for creators to certify their art. With Verisart, creators can edit timestamped authenticity certificates. With the rising use of generative AI, a growing number of creators are now leveraging it to produce artworks, pushing Christies to run a dedicated auction "Augmented Intelligence" in 2025. Famous artists producing AI generated artworks such as Refik Anadol, Rebecca Allen or Mario Klingemann are using Verisart to certify their generative art, highlighting how AI and blockchain convergence is reshaping art markets.

7 Conditions for Convergence: Success Factors and Barriers

Interviews with experts revealed both success factors and significant barriers to the convergence of blockchain and AI. Among the conditions favoring adoption, a coherent governance framework and the establishment of global standards were repeatedly emphasized. As one interviewee explained, "The same item has 25 different names … harmonized global standards are critical" (Int. 22). Without alignment on terminology, technical protocols, and compliance measures, efforts risk remaining fragmented and inefficient. Another decisive factor concerns user experience. As one practitioner put it, "If the house doesn't have a nice coat of paint … no one will move in" (Int. 02). In other words, no matter how powerful the underlying technology, the absence of accessible, intuitive interfaces can prevent uptake and hinder the transition from pilot projects to widespread deployment.

The question of talent availability was also raised as fundamental. Blockchain and AI rely on specialized programming languages and tools, such as JSON (JavaScript Object Notation) a data exchange format, for blockchain interactions, which demand a high level of technical expertise. Interviewees noted that the scarcity of skilled professionals in these fields has already become a bottleneck. Expanding the talent pool requires long-term adjustments in educational systems, which must prepare future professionals with the competencies needed to work across these converging technologies. Current figures illustrate the imbalance: according to the 2025 LinkedIn Global Talent Insights Report, AI job postings grew by 78% year on year, while the corresponding talent pool increased by only 24%. This is consistent with the findings of a report from the European Commission which outlines research outputs and talent availability as key enablers of the blockchain and AI technological convergence (European Commission, 2021). Finally, financial considerations play an equally critical role. Executives need a clear understanding of the potential return on investment before committing significant resources. As one interviewee remarked succinctly, "Put numbers behind it" (Int. 01).

Alongside these success factors, interviewees identified a range of barriers. Some stakeholders expressed unease about the full transparency provided by blockchain, which can be perceived as excessive or even undesirable in certain business contexts. "Blockchain creates too much transparency—not always desirable" (Int. 11) was how one decision-maker phrased this concern. Beyond perception, the technical limitations of both blockchain and AI were frequently cited. Trusted data oracles, while theoretically able to resolve the problem of external inputs, remain vulnerable to distortion, manipulation, or simple error, thereby undermining system reliability. Scalability also remains a central challenge: while Visa can process up to 24,000 transactions per second, Ethereum has historically managed only 20 to 30. In sectors such as energy, where countless micro-transactions would be required to integrate additional users into decentralized systems, this limitation represents a huge obstacle (Al Shareef et al., 2024).

Fragmentation of solutions further compounds these challenges. One interviewee captured the problem by noting that "the problem of the market is too many small players with small solutions" (Int. 09). In practice, the absence of interoperability mechanisms hampers seamless data exchange and collaboration. The heterogeneity of data formats, standards, structures, and network protocols makes integration even more difficult to achieve. The European Commission report 2021 also highlights the availability of data as a sensitive point towards the convergence of AI and blockchain. However, regions across the world present a different set of characteristics. While data in the United States is collected and managed by a few big technology companies, in China it is mainly under governmental monitoring. In Europe, the data market is considerably fragmented across member states which prevents the creation of integrated data sets to train AI models. On top of these technical and market-related hurdles, regulatory uncertainty continues to weigh heavily. Frameworks governing blockchain and AI remain in constant flux, leaving firms uncertain about the scope of permissible activities. For instance, regulatory ambiguity is one of the principal barriers to the adoption of decentralized energy markets (El-Said, 2025). At the same time, companies must continue to comply with existing rules such as the European

TABLE 1.1 Key success factors and barriers to the technological convergence of AI and blockchain

Success factors	*Barriers*
• Global standards (data formats, governance frameworks) • Improved user experience • Availability of talents and educational systems appropriateness • Research density • Return on investment	• Resistance to transparency • Technical limitations (data oracles, scalability) • Lack of interoperability • Availability of data • Regulatory uncertainty

Source: Own elaboration based on expert interviews and secondary sources (European Commission, 2021; LinkedIn Global Talent Insights Report, 2025).

GDPR (General Data Protection Regulation), adding another layer of complexity. Finally, the fragmented nature of regulatory approaches across jurisdictions exacerbates the difficulty, as AI–blockchain use cases often transcend national borders and therefore face inconsistent legal environments.

8 Practical Recommendations for Managers

Generally, innovation speed and magnitude are acknowledged to positively influence an organization's financial performance, as the adoption of emerging technologies confers a significant competitive advantage to firms (Gopalakrishnan, 2000). More specifically, Vijayakumar (2021) identified a significant positive relationship between the adoption of AI and blockchain and revenue growth for US-based small and medium-sized enterprises. However, their positive influence may be moderated by other factors such as industry sector, regulatory conditions, or the broader economic environment.

8.1 Decisions on Implementation

Before adopting emerging technologies, managers must critically assess whether their organization is ready for such a step. As advised by Day & Schoemaker (2000), working with emerging technologies requires thinking broadly and envisioning multiple possibilities before converging towards a single solution. In the same vein as Tsolakis et al. (2023) did with the Thai fish industry, the starting point for managers should be to map structures and process flows: who are the stakeholders, what are their activities, and where are the data and information flows, as well as the physical product flows. Once this map is established, managers can define the potential added value of blockchain and AI for each of these elements.

Rather than attempting to force-fit technologies into mainstream markets, firms should focus on finding market contexts where the combination of AI and blockchain can be most effectively applied. Many breakthrough products succeeded first in smaller, well-defined niches before scaling to mainstream adoption (Adner & Levinthal, 2002). Still, careful consideration needs to be paid to market research not to be too pessimistic or optimistic by generalizing market sizing from a niche to broader market. To illustrate, satellite-based mobile telephony operators such as Globalstar took the

preferences of a small set of executives for granted and thought the mass market would be keen on using satellite telephony.

This process must be aligned with the organization's overarching strategy. The key questions are: What are the desired outputs of AI and blockchain? What is feasible given the firm's resources? How coherent is it with the country wide economic and regulatory trends? By prioritizing and selecting the initiatives most likely to generate high impact and that are realistic to implement, managers can maximize the likelihood of success. To perform that prioritization exercise, managers can represent potential use cases on a matrix along two axes: business value—which would likely be a function of market sizing if aimed at external or benefits if aimed at an internal change—and the feasibility—hence how complex it is to be implemented and the resources and time needed to set such projects up. Once the initiatives are selected, it is essential to reflect upon the strategy: Should the initiative be developed in-house or externalized? What are the key assets for such initiatives to be successful? At this stage, it is important to elaborate an implementation roadmap with clear deadlines and key performance indicators. Simultaneously, it is of utmost importance to diagnose cases of similar projects that failed, to better understand potential pitfalls and avoid repeating them (Sodhi et al., 2022).

8.2 After Implementation

Even though the technological environment is highly uncertain due to its fast-paced evolution, one of the main pitfalls to avoid is staged or overly moderated commitments, which tend to dilute momentum and hinder learning (Day et al., 2000).

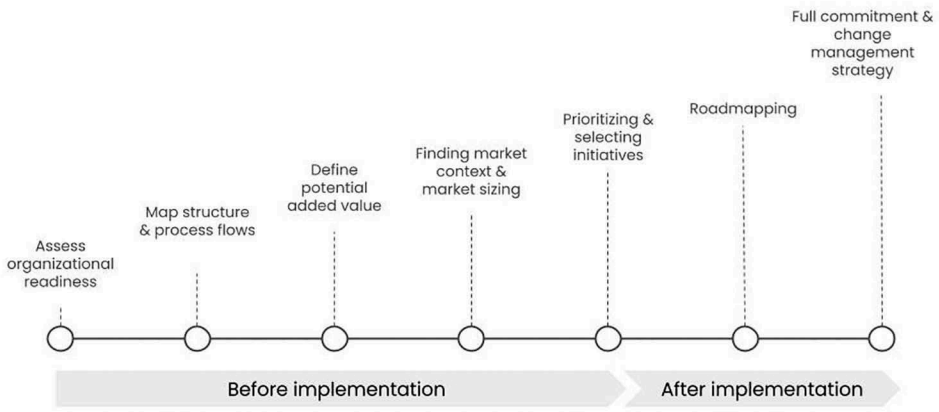


FIGURE 1.2 Practical recommendations for managers
Source: Own elaboration based on Day et al. (2000) and Mirvis et al. (1991).

Instead, organizations should pursue full commitment once a strategic direction is chosen, while maintaining flexibility to adapt to external changes.

A crucial element of success is initiating a comprehensive change strategy within the organization, cutting across all levels and functions. Technological implementation must be accompanied by user training, support mechanisms, and active measures to foster a positive acculturation to change (Mirvis et al., 1991). By embedding such a strategy, organizations can not only mitigate resistance but also ensure that AI and blockchain adoption translates into sustained performance improvements.

9 Conclusion

The ToT experiment showed how quickly AI and blockchain technology can enable autonomous agents that influence markets and communities. The convergence frameworks outlined above explain how boundaries between distinct technologies blur and ultimately reshape industries.

Across sectors, the interplay between AI and blockchain shifts data from static records to dynamic, predictive, and transactional assets. Similar dynamics within various sectors, such as supply chains, finance, healthcare, energy, and intellectual properties, show a clear convergence trajectory. Convergence of AI and blockchain generate opportunities in efficiency (faster, cheaper, fewer disputes), trust (fraud prevention, provenance), and innovation (autonomous agents, new business models).

Realizing this potential depends on critical factors: coherent governance frameworks, harmonized standards, user-centered design, talent availability, and clear return on investment (ROI). Conversely, key obstacles include transparency concerns, scalability limits, fragmented solutions, interoperability challenges, and persistent regulatory uncertainty.

For managers, the path forward involves carefully assessing organizational readiness, mapping data and process flows, and identifying where blockchain and AI can add the most value in alignment with strategic goals. Technological adoption should be paired with a comprehensive change strategy involving user training, cultural adaptation, and long-term commitment. Organizations that build on these foundations will be best placed to seize the next wave of digital transformation and maximize potential competitive advantages offered by the convergence of AI and blockchain.

References

Addula, S. R., Meduri, K., Nadella, G. S., & Gonaygunta, H. (2024). AI and blockchain in finance: Opportunities and challenges for the banking sector. *International Journal of Advanced Research in Computer and Communication Engineering*, 13(2), 184–190.

Adner, R., & Levinthal, D. A. (2002). The emergence of emerging technologies. *California Management Review*, 45(1), 50–66.

Ahluwalia, S., Mahto, R. V., & Guerrero, M. (2020). Blockchain technology and startup financing: A transaction cost economics perspective. *Technological Forecasting and Social Change*, 151, 119854.

Aithal, P. S., & Aithal, S. (2018). Study of various general-purpose technologies and their comparison towards developing sustainable society. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, 3(2), 16–33.

Al Shareef, A. M., Seçkiner, S., Eid, B., & Abumeteir, H. (2024). Integration of blockchain with artificial intelligence technologies in the energy sector: a systematic review. *Frontiers in Energy Research*, 12, 1377950.

Athreye, S., & Keeble, D. (2000). Technological convergence, globalisation and ownership in the UK computer industry. *Technovation*, 20(5), 227–245.

Bathula, A., Gupta, S. K., Merugu, S., Saba, L., Khanna, N. N., Laird, J. R., … & Suri, J. S. (2024). Blockchain, artificial intelligence, and healthcare: the tripod of future—a narrative review. *Artificial Intelligence Review*, 57(9), 238.

Bhumichai, D., Smiliotopoulos, C., Benton, R., Kambourakis, G., & Damopoulos, D. (2024). The convergence of artificial intelligence and blockchain: The state of play and the road ahead. *Information*, 15(5), 268.

Bröring, S. (2005). *The front end of innovation in converging industries: The case of nutraceuticals and functional foods*. Wiesbaden: Deutscher Universitäts-Verlag.

Catalini, C. (2017). How blockchain technology will impact the digital economy. *Blockchains Smart Contracts Internet Things*, 4, 2292–2303.

Choi, N., & Kim, H. (2024). Technological convergence of blockchain and artificial intelligence: A review and challenges. *Electronics*, 14(1), 84.

Crafts, N. (2021). *Artificial intelligence as a general-purpose technology: an historical perspective. Oxford Review of Economic Policy*, 37(3), 521–536.

Davidson, S., De Filippi, P., & Potts, J. (2018). Blockchains and the economic institutions of capitalism. *Journal of Institutional Economics*, 14(4), 639–658.

Day, G. S., & Schoemaker, P. J. (2000). Avoiding the pitfalls of emerging technologies. *California Management Review*, 42(2), 8–33.

El-Said, M. M., M. (2025). AI-driven decentralized energy systems: A review of peer-to-peer renewable energy networks. *Metaheuristic Optimization Review*, 4(1), 1–11. DOI: doi:10.54216/MOR.040101.

Gopalakrishnan, S. (2000). Unraveling the links between dimensions of innovation and organizational performance. *The Journal of High Technology Management Research*, 11(1), 137–153.

Hacklin, F., Marxt, C., & Fahrni, F. (2010). An evolutionary perspective on convergence: Inducing a stage model of inter-industry innovation. *International Journal of Technology Management*, 49(1–3), 220–249.

Kumar, S., Lim, W. M., Sivarajah, U., & Kaur, J. (2023). Artificial intelligence and blockchain integration in business: Trends from a bibliometric-content analysis. *Information Systems Frontiers*, 25(2), 871–896.

Lipsey, R. G., Carlaw, K. I., & Bekar, C. T. (2005). *Economic transformations: General purpose technologies and long-term economic growth*. Oxford: Oxford University Press.

Martinez, D., Magdalena, L., & Savitri, A. N. (2024). Ai and blockchain integration: Enhancing security and transparency in financial transactions. *International Transactions on Artificial Intelligence*, 3(1), 11–20.

Mirvis, P. H., Sales, A. L., & Hackett, E. J. (1991). The implementation and adoption of new technology in organizations: The impact on work, people, and culture. *Human Resource Management*, 30(1), 113–139.

Montes, G. A., & Goertzel, B. (2019). Distributed, decentralized, and democratized artificial intelligence. *Technological Forecasting and Social Change*, 141, 354–358.

Pablo, R. G. J., Roberto, D. P., Victor, S. U., Isabel, G. R., Paul, C., & Elizabeth, O. R. (2022). Big data in the healthcare system: a synergy with artificial intelligence and blockchain technology. *Journal of Integrative Bioinformatics*, 19(1), 20200035.

Ragot, S., Rey, A., & Shafai, R. (2020). IP lifecycle management using blockchain and machine learning: Application to 3D printing datafiles. *World Patent Information*, 62, 101966.

Rosenberg, N. (1963). Technological change in the machine tool industry, 1840–1910. *The Journal of Economic History*, 23(4), 414–443.

Sahut, J. M., Schweizer, D., & Peris-Ortiz, M. (2022). Technological forecasting and social change introduction to the VSI technological innovations to ensure confidence in the digital world. *Technological Forecasting and Social Change*, 179, 121680.

Schuelke-Leech, B. A. (2018). A model for understanding the orders of magnitude of disruptive technologies. *Technological Forecasting and Social Change*, 129, 261–274.

Sodhi, M. S., Seyedghorban, Z., Tahernejad, H., & Samson, D. (2022). Why emerging supply chain technologies initially disappoint: Blockchain, IoT, and AI. *Production and Operations Management*, 31(6), 2517–2537.

Soori, M., Dastres, R., & Arezoo, B. (2023). AI-powered blockchain technology in industry 4.0, a review. *Journal of Economy and Technology*, 1, 222–241.

Swan, M. (2018). Blockchain for business: Next-generation enterprise artificial intelligence systems. In *Advances in computers* (Vol. 111, pp. 121–162). London: Elsevier.

Tsolakis, N., Schumacher, R., Dora, M., & Kumar, M. (2023). Artificial intelligence and blockchain implementation in supply chains: A pathway to sustainability and data monetisation? *Annals of Operations Research*, 327(1), 157–210.

Umoga, U. J., Sodiya, E. O., Amoo, O. O., & Atadoga, A. (2024). A critical review of emerging cybersecurity threats in financial technologies. *International Journal of Science and Research Archive*, 11(1), 1810–1817.

Vijayakumar, H. (2021). Impact of AI–blockchain adoption on annual revenue growth: An empirical analysis of small and medium-sized enterprises in the United States. *International Journal of Business Intelligence and Big Data Analytics*, 4(1), 12–21.

Wissuwa, F., & Durach, C. F. (2023). Turning German automotive supply chains into sponsors for sustainability. *Production Planning & Control*, 34(2), 159–172.

2

CALL ME VASARAI

AI for Cultural Heritage and Business Innovation

Irene Maria Musumeci and Pierluigi Casale

1 Introduction

Artificial intelligence (AI) has become integral to contemporary business strategy. Recent data indicates that most companies had integrated at least one AI-powered capability by 2024, with adoption rates projected to reach 78% in 2025, up from 55% the previous year. This rapid integration reflects strategic positioning within competitive markets. However, a 2025 McKinsey report[1] revealed that while 92% of organizations plan to increase AI investments, only 1% of leaders consider their organizations fully AI-mature. This gap between aspiration and implementation suggests that successful AI adoption depends not merely on algorithmic sophistication but on organizational preparedness, strategic coherence, ethical consideration, and clear articulation of expected benefits.

Call Me VasarAI represents an initiative to bridge cultural heritage and artificial intelligence, two domains that have historically operated in separate spheres. Traditional art appreciation typically requires specialized knowledge, including familiarity with artistic styles, historical contexts, and symbolic interpretations. Consider works such as Turner's *The Storm* or Monet's *Rouen Cathedral* series; these paintings contain layered narratives that often remain inaccessible without expert guidance. Through the application of image recognition and generative AI technologies, Call Me VasarAI attempts to address this accessibility challenge by providing both automated classification and narrative interpretation. The project takes its name from Giorgio Vasari, the Italian Renaissance biographer and artist, reflecting its dual function as both classifier and storyteller. The research question at the core of Call Me VasarAI operates on both technical and philosophical levels: How can machine learning be used to improve classification and generate descriptive, stylistically accurate attributions for artworks, enhancing accessibility to art interpretation? This question opens bigger reflections: can we teach AI to interpret a painting in a way that resonates with human understanding? Could the application of AI open the doors of artistic knowledge to people who might not have formal education, or who face linguistic or sensory

DOI: 10.4324/9781003632429-4
This chapter has been made available under a CC-BY-NC-ND 4.0 license

barriers? If answered affirmatively, the implications extend beyond museum contexts. Call Me VasarAI could serve as a model for using AI to make complex information accessible, generate value through narrative construction, and guide innovation with ethical responsibility.

This chapter examines the Call Me VasarAI project through lenses of business impact and strategic positioning. It situates the initiative within broader patterns of AI adoption in cultural heritage, highlighting economic opportunities and increasing demand for digital cultural experiences. The analysis explores underlying technologies, including computer vision, multimodal embeddings, and narrative generation, while considering their business implications. The methodology is presented as a case study in design thinking, data readiness, and product development. Subsequent sections analyze results, potential applications in education, tourism, and cultural commerce, alongside acknowledged limitations. Discussions of AI readiness, change management, and organizational transformation acknowledge research indicating that approximately 30% of AI projects progress beyond pilot stages. Ethical concerns, including bias, authorship, and environmental sustainability, are addressed through examination of responsible AI frameworks and research on the environmental impact of large-scale models. The chapter concludes by considering pathways for scaling the system across industries to generate sustainable value and competitive advantage.

2 The Business of AI in Cultural Heritage

The cultural heritage sector is undergoing a significant transformation driven by artificial intelligence (AI) and its applications. Although at the beginning, museums, tourism, digital archiving, and education used AI as a means of cultural heritage preservation and education, now it is being considered as a tool for interpretation, public engagement, and sustainability or to develop new business models.

AI adoption and the cultural market: Cultural heritage looks like a distant field for AI innovation, but data demonstrates otherwise. Museums, historical sites, and tourism operators are using AI to enrich their visitor offerings, reduce costs, and open up new business strategies. A report from MuseumNext notes how virtual guides, personalized tours, and predictive maintenance are transforming museums (Styx, 2024). The Louvre museum launched Leonardo, an AI assistant for visitor questions, while the British Museum tested AR apps offering layered stories. Accessibility is also advancing: Rio's Museum of Tomorrow uses the IRIS+ chatbot for sign language and audio descriptions, and the Smithsonian applies AI to predictive maintenance and visitor feedback. These changes mirror broader trends. Though heritage is a small share, adoption is rising fast. Demand for digital experiences accelerated after the COVID-19 pandemic closures, with platforms like Google Arts & Culture[2] and startups such as Artivive[3] expanding online art access. The launch of DATALAND,[4] a Los Angeles museum for AI-generated art, further shows growing public interest. Still, AI has not significantly boosted visitor numbers. A 2023 study of 19 museums found its effect on attendance to be modest and not statistically significant. The real value lies in engagement: making visits longer, reaching younger audiences, and generating digital revenue. Google Arts & Culture's Art & Selfie,[5] which matches selfies with portraits, increases online interaction while supporting print and merchandise sales. AI also

powers targeted marketing and personalized suggestions, helping secure donations and memberships.

Economic potential and competitive advantage: The value of AI extends beyond museums to cultural heritage. In tourism, for example, AI-based tools for translation, recommendations, and storytelling are making sites more accessible and encouraging visitors to stay longer. Meanwhile, the "art tech" market is expanding quickly. AI-generated art, digital marketplaces, and blockchain collectables are challenging traditional ideas of authorship and authenticity, while opening new entrepreneurial paths. A 2024 BCG report makes the business case clear: companies leading in AI achieve 1.5 times higher revenue growth and 1.6 times greater shareholder returns compared to peers. Importantly, about 62% of this value comes from core processes such as operations, sales, and product development, not from isolated digital experiments. In other words, AI pays off when it is integrated across the organization. At the same time, 70% of the challenges in AI adoption stem from people and process issues rather than technology. For cultural institutions, this means that capturing value requires investment not only in digital infrastructure but also in organizational change and staff training.

2.1 Methodological Framework and Data Sources

To situate Call Me VasarAI within the broader landscape of AI adoption in cultural heritage, this chapter draws on multiple data sources and analytical frameworks. The research synthesizes findings from industry reports, academic databases, and case studies of cultural institutions that have implemented AI-driven solutions. Key insights into AI adoption rates and economic impact come from recent surveys conducted by McKinsey & Company (2025) and the Stanford HAI AI Index Report (2025). These reports track AI deployment across sectors, including cultural heritage, and provide quantitative benchmarks for success rates, investment levels, and organizational readiness. The Boston Consulting Group's 2024 study on AI value creation offers specific data on revenue growth and return on investment, demonstrating how leading organizations achieve competitive advantage through systematic AI integration. These sources establish the business context within which Call Me VasarAI operates.

To ground the technical and theoretical dimensions of the project, the research consulted databases including IEEE Xplore, ACM Digital Library, and Google Scholar. Search queries focused on terms such as "AI cultural heritage," "computer vision art classification," "multimodal learning museums," and "AI readiness frameworks." This process identified seminal papers on deep learning for art analysis (Saleh & Elgammal, 2015; Cetinic & She, 2022), multimodal embeddings (Radford et al., 2021), and organizational AI adoption (Fountaine et al., 2019). The Getty Research Institute's Art & Architecture Thesaurus and the CIDOC Conceptual Reference Model provided standardized vocabularies and ontologies used to enrich the semantic layer of the system.

Examples of AI implementation in cultural institutions were selected based on several criteria: public documentation of technical approach, evidence of measurable impact (visitor engagement, revenue, accessibility metrics), diversity of application (classification, personalization, accessibility features), and alignment with ethical and

sustainability principles. Institutions such as the Louvre, British Museum, Smithsonian, and Google Arts & Culture were chosen because they represent varied scales, geographies, and use cases, offering lessons applicable to organizations at different stages of digital maturity. Part of this research also stems from a personal passion for art and direct exposure to innovative applications of machine learning in cultural contexts. Experiences such as the immersive exhibition at Casa Batlló, which uses AI and augmented reality to reimagine Gaudí's architecture, and *Assassin's Creed's* historically accurate recreation of Renaissance Florence, which uses gaming technology to tell stories through interactive environments, demonstrated the potential of combining technology with cultural narrative. These encounters highlighted how machine learning and AI could transform passive observation into active engagement, inspiring the conceptual foundation of Call Me VasarAI.

The chapter employs a set of key performance indicators to assess AI impact in cultural heritage. Technical metrics include classification accuracy, inference speed, and model size. Business metrics encompass user engagement time, conversion rates (e.g., ticket purchases, memberships), digital revenue streams, and audience reach. Social impact is measured through accessibility features adopted, linguistic coverage, and inclusivity of represented communities. These indicators align with frameworks proposed by the AI Readiness Blueprint (Agility at Scale) and UNESCO's guidelines for digital heritage preservation. By triangulating data from industry, academia, and practitioner case studies, this chapter establishes a robust empirical foundation for analyzing Call Me VasarAI's contributions to both cultural heritage and business innovation.

2.2 Literature Review

The intersection of artificial intelligence and cultural heritage has emerged as a distinct research domain over the past decade. Scholars have examined how computational methods transform the way institutions preserve, interpret, and disseminate cultural content. Early work focused primarily on digitization and archiving, but recent literature addresses more complex questions around semantic understanding, user engagement, and business viability.

Research into automated artwork classification began with traditional machine learning approaches. Saleh and Elgammal (2015) demonstrated that convolutional neural networks could predict artistic style with reasonable accuracy when trained on curated datasets. Their work on WikiArt established benchmarks that subsequent studies have sought to improve. Tan et al. (2016) extended this by introducing attribute-based classification, showing that models trained to recognize specific visual features such as brushstroke texture or color palette performed better than those relying solely on holistic image representations. More recently, Cetinic and She (2022) reviewed deep learning applications in art analysis, noting that while accuracy has improved, challenges persist around class imbalance, dataset bias, and the subjective nature of artistic categorization. Their findings suggest that technical performance alone does not guarantee cultural or educational value.

The advent of models like CLIP marked a shift from pure classification to semantic retrieval. Radford et al. (2021) showed that aligning vision and language in a shared

embedding space enables zero-shot classification, a capability particularly useful for domains with limited labelled data. In cultural heritage, this approach has been explored by Castellano et al. (2021), who applied CLIP to museum collections and found it effective for cross-modal search tasks. However, they also highlighted the need for domain-specific fine-tuning and the integration of structured metadata such as ontologies or controlled vocabularies. Garcia et al. (2023) took this further by combining CLIP embeddings with knowledge graphs, demonstrating that enriching models with contextual information improves both accuracy and interpretability.

The use of large language models for cultural storytelling is a relatively recent development. Brown et al. (2020) introduced GPT-3, demonstrating its capacity for coherent, context-aware text generation across diverse tasks. In the context of art, researchers have explored whether such models can produce descriptions that go beyond surface-level captions. Bongini et al. (2021) experimented with fine-tuning GPT-2 on art-historical texts, finding that the model could generate plausible interpretations when given sufficient contextual prompts. Yet concerns remain about factual accuracy, cultural sensitivity, and the risk of reinforcing existing biases present in training data. Krishna et al. (2022) argued that narrative systems must incorporate human oversight and domain expertise to ensure quality and trustworthiness.

Beyond technical capabilities, the literature addresses organizational and economic dimensions. Borowiecki and Navarrete (2017) analyzed the economic impact of digitization in museums, noting that while initial investments are substantial, long-term benefits include expanded audiences, new revenue streams, and enhanced scholarly collaboration. However, realizing these benefits depends on what recent studies call "AI readiness." According to research by Fountaine et al. (2019) and Ransbotham et al. (2020), successful AI adoption requires alignment of strategy, data governance, talent, and culture. Many organizations struggle to move from pilot projects to scaled deployment, with failure rates as high as 70%. For cultural institutions specifically, Navarrete (2019) identified barriers including limited technical expertise, fragmented data, and resistance to change among staff. The author recommended phased implementation, cross-sector partnerships, and investment in capacity building.

A growing body of work examines the ethical implications of AI in heritage. Whittaker et al. (2018) documented concerns around algorithmic bias, transparency, and accountability in AI systems. In cultural contexts, these issues are particularly sensitive given the risk of misrepresenting marginalized communities or reinforcing Eurocentric perspectives. Windhager et al. (2018) stressed the importance of inclusive dataset curation and participatory design, arguing that affected communities should have a voice in how their heritage is represented. Environmental sustainability has also gained attention: Strubell et al. (2019) calculated the carbon footprint of training large language models, prompting calls for "Green AI" practices such as model compression, transfer learning, and energy-efficient infrastructure. Schwartz et al. (2020) proposed guidelines for responsible model development, emphasizing the need to balance performance with resource consumption.

Despite this rich body of work, several gaps remain. Most studies treat classification and narrative generation as separate tasks, with few attempts to integrate them within a unified framework. Business-oriented research often focuses on generic AI adoption rather than the specific challenges faced by cultural organizations. Moreover, there is

limited empirical evidence on how AI-generated content affects visitor behavior, educational outcomes, or revenue generation. This chapter addresses these gaps by presenting Call Me VasarAI as an integrated case study that combines technical innovation with business analysis, grounded in real-world implementation and evaluation.

3 Call Me VasarAI: Objectives and Description

Call Me VasarAI wants to use technologies such as machine learning to improve the classification of artworks and generative AI to create descriptive, stylistically accurate attributions that enhance accessibility to art interpretation. This need emerged from the authors' desire to democratize art. Artworks are visual languages embedded in cultural, historical, and emotional contexts. Without knowledge of the artist's biography, the style's characteristic features, or the socio-political circumstances of production, many viewers are left to guess. This exclusion perpetuates cultural inequities: those with access to art education or privileged backgrounds enjoy deeper experiences, while others remain distant observers. The digital transformation of society offers a means to address this gap.

Call Me VasarAI originated as an academic research project exploring the intersection of AI and cultural heritage. While conceived initially as a proof-of-concept demonstrating the feasibility of integrating classification, multimodal understanding, and narrative generation, the project's modular architecture and business potential position it as a scalable framework that could evolve into a commercial application or be adopted by cultural institutions. The open-source nature of the core components ensures transparency and adaptability, allowing museums, educational platforms, or tourism operators to customize the system to their specific collections and audiences.

The objectives of Call Me VasarAI are, therefore, both technological and human-centric:

Art classification: The system should automatically identify the artist and artistic style of a given painting using image recognition models trained on large labelled datasets. High classification accuracy is crucial for establishing credibility and fostering user trust. As part of the business strategy, accurate classification also enables targeted marketing, data analytics, and dynamic content delivery.

Descriptive storytelling: Beyond labels, the system aims to generate natural language descriptions that explain visual, historical, and emotional aspects of a painting. Narratives should be engaging, factually accurate and tailored to non-expert audiences. These descriptions can drive user engagement, longer dwell times and emotional connection: key metrics for museums and digital platforms.

Accessibility: The output should be accessible across various educational levels, languages, and sensory abilities. Features such as multilingual support, simplified language options, and text-to-speech integration broaden the audience and open new markets (for example, visually impaired users or tourists speaking different languages).

Scalability and flexibility: The framework should be modular, allowing integration with existing museum systems or expansion to new art forms. Scalability is crucial for cost-effectiveness and for adapting to varied datasets across institutions.

Ethical and sustainable design: The project must address questions of authorship, bias, and energy consumption. Narratives should respect cultural contexts, avoid reproducing stereotypes and be transparent about AI's role. Efficient models and responsible data practices should minimize environmental impact.

In visual art, image classification is the foundation. Call Me VasarAI uses models like ResNet (He et al., 2016) because it performs well, especially when pre-trained on large datasets like ImageNet (Deng et al., 2009) and fine-tuned on specialized ones like WikiArt. ResNet's deep residual architecture can detect stylistic features such as brushstrokes, colors, and composition, but classification alone is insufficient. Art often blends styles, and understanding requires historical, cultural, and symbolic context.

The choice of WikiArt as the primary training dataset for this proof-of-concept was driven by its accessibility, scale, and diversity. WikiArt provides a publicly available collection of over 80,000 artworks spanning multiple centuries, styles, and artists, making it ideal for training deep learning models without the financial and legal constraints associated with proprietary museum datasets (Saleh & Elgammal, 2015). However, this dataset is not without limitations. Its crowdsourced nature introduces inconsistencies in metadata quality, class imbalance favoring Western artists, and underrepresentation of non-Western and contemporary art. These issues are acknowledged and addressed through data preprocessing, augmentation, and the system's dataset-agnostic design, which allows for seamless integration of higher-quality institutional datasets in future deployments. A detailed justification of the WikiArt dataset, including its advantages, limitations, and strategies for bias mitigation, is provided in Section 5.

To address the need for deeper contextual understanding, research now explores multimodal models that integrate images and text. CLIP, developed by OpenAI, aligns them in a shared space, enabling zero-shot classification: matching an unseen image to the most relevant textual description (Radford et al., 2021). While powerful for semantic retrieval, CLIP requires domain knowledge; therefore, enriching it with ontologies, knowledge graphs, or curated metadata (e.g., Getty AAT) is crucial to strengthen its interpretive capacity. In Call Me VasarAI, CLIP embeddings are combined with artist biographical data and style descriptions to create context-rich prompts that guide the narrative generation process.

The step from recognition to interpretation comes through natural language generation (NLG). Large language models (LLMs) like GPT-3, GPT-4 (OpenAI, 2023), or open-source alternatives such as GPT-Neo (Black et al., 2022) generate coherent, context-aware text. Unlike template-based captioning, they craft narratives that combine history and artistic analysis; for instance, turning "a woman holding a parasol" into an Impressionist vignette that discusses Monet's techniques, the role of light in his work, and the cultural context of bourgeois leisure in 19th-century France. Yet, LLMs carry biases and may invent details, making careful prompts and domain-specific tuning essential. Call Me VasarAI employs temperature control, repetition penalties, and structured prompts to balance creativity with factual accuracy.

By integrating these technologies, ResNet for classification, CLIP for multimodal semantic understanding, and GPT-Neo for narrative generation, Call Me VasarAI offers a comprehensive framework that moves beyond mere labelling to provide rich,

accessible interpretations of artworks. This approach addresses the core mission of democratizing art knowledge while maintaining the technical rigor and ethical considerations necessary for responsible AI deployment in cultural heritage.

4 The Current Interplay between Art and AI

AI is reshaping cultural heritage, moving from recognition tools to systems that enable storytelling and immersion. It bridges collections and audiences, improves accessibility, fosters new business models, and raises ethical as well as environmental challenges.

4.1 From Recognition to Storytelling

AI in the arts has progressed from basic pattern matching to narrative generation. Early examples came from outside the arts: Shazam identifies songs from short clips by comparing audio fingerprints to a database (Wang, 2003). ArtPI applies a similar method to artworks, while PlantAI does so for plant species. These systems highlight AI's ability to turn sensory inputs into recognizable units and connect them with contextual data.

4.2 AI Readiness and Adoption Challenges

The sophistication of AI models is only part of the challenge. What matters equally is AI readiness: an organization's ability to adopt, integrate, and scale AI effectively. Studies show that 80% of projects fail due to inflated expectations or weak business objectives, and only 30% move beyond the pilot stage (Ransbotham et al., 2020). These failures stem less from algorithms than from gaps in strategy, alignment, and execution. Frameworks for readiness assess multiple dimensions: strategy, technology, people, data, and governance. Common barriers include unclear goals, poor infrastructure, weak data governance, and unprepared staff. Only 21% of firms have adequate GPU capacity, and just 30% maintain strong threat monitoring and encryption (McKinsey & Company, 2025).

Putting readiness into practice requires a phased approach; institutions can start with foundational steps, such as the digitalization of collections and setting up data governance. Digital guides or chatbots integrate with AI applications to demonstrate value and build confidence. As capabilities mature, scale successes across departments, refine processes, and embed AI into the organizational DNA, fostering continuous improvement and innovation. The impact of AI should be evaluated using clear indicators, including visitor satisfaction, engagement time, revenue, and inclusivity. Findings can then guide adjustments to institutional strategies. Collaboration with universities, start-ups, and technology providers often accelerates both experimentation and implementation.

Frameworks such as the AI Readiness Blueprint by Agility at Scale outline eight dimensions of readiness: strategy, data, technology, people, culture, processes, governance, and ethics. Each is interdependent: an effective strategy relies on quality data, technology requires skilled professionals, and institutional culture strongly shapes

innovation capacity. The blueprint emphasizes aligning AI with business goals, investing in data governance, scalable infrastructure, fostering an innovation culture, integrating processes, and upholding governance and ethics. Executive sponsorship, cross-functional collaboration, and ongoing improvement are also essential. Case studies show the value of readiness: the Louvre's Leonardo assistant succeeded thanks to strong infrastructure, expert partnerships, and staff training, while the British Museum's AR app thrived through collaboration and iterative testing. Poor data, unrealistic goals, or weak maintenance often cause AI projects to fail while a structured approach increases the likelihood of achieving lasting value.

4.3 Ethical and Environmental Considerations

AI systems are not neutral because they mirror the data and values built into them. In cultural heritage, this raises concerns about authorship, bias, transparency, privacy, and the risk of misrepresenting artworks. Who owns AI-generated content? How can systems avoid reinforcing stereotypes? A Berkeley Haas survey found 77% of AI product managers felt unprepared for ethical challenges such as privacy and bias (Hickman et al., 2020). Organizations with leadership committed to responsible AI were four times more likely to develop solid practices, showing leadership is key to accountability.

Sustainability is another issue, in fact, training large models consumes vast energy. MIT estimated GPT-3 used 1,287 MWh, emitting 552 tons of CO_2 (Zewe, 2023). With data centers' energy use expected to more than double by 2026, institutions must balance AI's benefits with its footprint. ERCIM News contrasts "Red AI," energy-intensive systems, with "Green AI," which promotes efficiency through transfer learning, pruning, and leaner architectures (Barbierato & Gatti, 2025). Cultural organizations can mitigate impact by reusing models, optimizing code, and scheduling energy-heavy tasks during off-peak hours.

4.4 Competitive Landscape and Market Positioning

Understanding where Call Me VasarAI fits within the broader ecosystem of AI-driven art applications requires examining the landscape of existing solutions. Over the past decade, several companies have developed technologies that use image recognition to identify artworks and deliver contextual information to users, each carving out distinct niches based on their target audiences, partnerships, and business strategies.

Smartify[6] stands out as perhaps the most established player in this space. Launched in 2017 through a collaboration between a social enterprise and MOBGEN, which was later acquired by Accenture Interactive, the app has built partnerships with more than 700 cultural institutions globally (Smartify, 2024). Visitors to venues like the National Gallery in London, the Rijksmuseum in Amsterdam, or the Smithsonian can scan artworks with their smartphones to access descriptions, audio guides, and even augmented reality overlays. The platform's success stems from its institutional focus: museums license Smartify to enhance visitor experiences, generating revenue through these partnerships alongside in-app purchases. User reception has been strong, with ratings averaging 4.7 out of 5 across major app stores, though some reviews note

limitations in recognizing works outside partnered collections. Smartify's business model is fundamentally B2B: it serves museums first, and visitors second, with its database restricted to works within partner institutions.

Magnus[7] takes a different approach entirely. Positioning itself as "Shazam for art," the application queries a massive database containing references to 10 million artworks sourced from galleries, auction houses, and user contributions (Magnus, 2024). Rather than serving casual museum visitors, Magnus targets collectors, dealers, and art enthusiasts who want pricing information, historical sales data, and exhibition records. The app provides market valuations alongside basic identification, making it as much an investment research tool as an educational resource. However, its reliance on crowdsourced data introduces quality concerns, and user feedback suggests that recognition accuracy remains inconsistent. Magnus operates in the commercial art market space, prioritizing transaction facilitation and valuation over educational content or accessibility.

Google Arts & Culture operates on a completely different scale. Backed by Alphabet's infrastructure and resources, the platform offers high-resolution virtual tours, thematic collections, and image matching capabilities across thousands of institutions worldwide (Google Arts & Culture, n.d.). Features like "Art Selfie," which match user photos to historical portraits, demonstrate how gamification can drive viral engagement and introduce new audiences to art. Yet Google's approach reflects different priorities: the service functions as a public good within a broader corporate ecosystem rather than as a standalone commercial product. Its comprehensive scope comes at the cost of agility, and smaller institutions may find integration with Google's systems technically demanding or incompatible with their existing infrastructure.

In September 2025, Google expanded its visual AI capabilities significantly with the launch of Nano Banana, the Gemini 2.5 Flash Image model integrated directly into the Gemini app (Google DeepMind, 2025a). While primarily designed for image generation and editing, this development demonstrated Google's commitment to multimodal understanding that extends to artwork analysis. The model's ability to parse composition, anatomy, color, lighting, perspective, and style positions it as a potential competitor in the art education and interpretation space. Users can upload images of artworks and receive detailed feedback on visual elements, effectively functioning as an AI art teacher accessible through natural language interaction. The tool's viral adoption across social platforms, with millions of users creating and editing images, illustrates the market appetite for accessible, AI-powered visual analysis tools.

Following closely in November 2025, Google introduced Nano Banana Pro, built on Gemini 3 Pro, which uses advanced reasoning and real-world knowledge to create context-rich educational content (Raisinghani, 2025). The system can generate accurate infographics, diagrams, and visual explanations based on uploaded content or real-world facts. Its integration with Google Search's knowledge base enables it to provide comprehensive visual interpretations that combine factual accuracy with aesthetic presentation. For cultural heritage applications, this capability suggests potential for creating educational materials that explain artistic movements, techniques, and historical contexts in visually engaging formats.

Additionally, September 2025 Gemini Drop introduced camera-sharing capabilities in Gemini Live, transforming smartphones into intelligent visual assistants capable of

real-time analysis of physical environments (Google, 2025b). While not specifically marketed for artwork recognition, this technology could theoretically enable users to point their camera at a painting in a museum and receive instant contextual information, competing directly with dedicated apps like Smartify and Magnus. The fact that Gemini processes visual information conversationally and can access vast knowledge bases positions it as a formidable potential entrant into the cultural heritage space, though Google has not yet announced specific partnerships with museums or art institutions.

These developments underscore a broader trend: general-purpose AI platforms with strong multimodal capabilities are beginning to encroach on specialized applications. Call me VasarAI's competitive advantage lies not in computational power, which Google far exceeds, but in its focused mission of democratizing art knowledge, its open-source adaptability for institutional deployment, and its integration of recognition with narrative storytelling specifically designed for cultural contexts. As platforms like Gemini become more capable, the distinction between general AI tools and specialized cultural applications may blur, making institutional partnerships, domain expertise, and user trust increasingly valuable differentiators.

Beyond recognition applications, the generative AI art market presents a different kind of competition. Platforms such as DALL·E, Midjourney, Stable Diffusion, and Adobe Firefly have captured enormous attention by enabling users to create entirely new images from text prompts. Midjourney currently commands about 27% of the generative AI image market, with DALL·E close behind at just over 24% (Meitu, 2024). These tools serve artists, designers, marketers, and content creators seeking rapid visual asset generation. The broader AI art market reached $2.9 billion in 2023, and projections suggest it will grow to $20.7 billion by 2031, reflecting surging demand across advertising, entertainment, education, and social media (InsightAce Analytic, 2024). While generative tools serve fundamentally different purposes than interpretive systems like Call Me VasarAI, the distinction may blur over time. Future versions of recognition platforms could incorporate generative capabilities, perhaps create customized visual explanations or suggesting artistic variations.

Specialized services address pain points within the art world. Hephaestus Analytical, for instance, applies computer vision and stylometric analysis to detect forgeries with accuracy exceeding 98% (Hephaestus Analytical, 2024). The company has partnered with insurance providers to offer authenticity guarantees, targeting high-value transactions where provenance verification justifies significant fees. This illustrates how AI can solve complex, high-stakes problems beyond consumer applications. Similarly, platforms like Limna and Artsy serve collectors and investors through price validation, market analytics, and access to auction results for hundreds of thousands of artists (Limna, 2024). These tools function essentially as AI-powered advisors, helping users assess valuations and make purchasing decisions. Their transactional focus differs from Call Me VasarAI's educational mission, yet their pricing algorithms and market data represent valuable capabilities that could inform future monetization strategies.

Call Me VasarAI, conceived in 2023 as an academic research project, distinguishes itself from these established competitors through fundamental differences in mission, architecture, and user focus. Where Smartify operates as a B2B solution tethered to institutional partnerships and Magnus serves the commercial art market, Call Me

VasarAI was designed from inception to democratize art knowledge for the broadest possible audience. This distinction manifests across multiple dimensions that together define its unique market position.

The most significant distinction lies in its emphasis on narrative generation rather than mere information delivery. Smartify provides factual descriptions: title, artist, date, medium, perhaps a brief historical note drawn from museum catalogues. Magnus adds pricing data and auction records, useful for transactions but offering little interpretive depth. Call Me VasarAI, by contrast, weaves biographical, stylistic, and historical context into engaging stories designed to resonate emotionally with users. Where Smartify might tell you Monet's *Woman with a Parasol* was painted in 1875, Call Me VasarAI would explain how the dappled light filtering through fabric captured Impressionism's fascination with fleeting moments, discuss Monet's personal relationship with his subject, and situate the work within debates about bourgeois leisure in late 19th-century France. This narrative approach transforms passive recognition into active learning, addressing the gap between knowing what something is and understanding why it matters.

Accessibility represents another crucial differentiator. Smartify and Magnus implicitly target audiences already comfortable navigating cultural institutions or art markets: museum-goers with smartphones, collectors with disposable income, professionals with baseline art knowledge. Call Me VasarAI deliberately prioritizes inclusion. Its multilingual support goes beyond simple translation to culturally adapted narratives. Simplified language options make content comprehensible to children, non-native speakers, or users without formal education. Text-to-speech integration serves visually impaired users who cannot read labels or wall text. These features expand the potential audience to include demographics systematically excluded from traditional art spaces, advancing the project's core mission of democratization.

The technical architecture offers strategic advantages absent from proprietary competitors. Smartify's closed system requires museums to license specific features and submit to platform constraints. Magnus's database, though larger, remains opaque in its curation and vulnerable to quality issues inherent in crowdsourcing. Call Me VasarAI's modular, open-source foundation allows institutions to adapt the framework to their specific collections without vendor lock-in. A small regional museum can deploy the system using its own server infrastructure, train models on its collection, and customize narratives to reflect local cultural contexts. This flexibility lowers barriers to entry for organizations with limited budgets and technical capacity, while also ensuring transparency in how the system operates—a consideration increasingly important given growing scrutiny of AI ethics and accountability.

Database philosophy further distinguishes the approaches. Smartify's database consists only of works within partner institutions, creating a walled garden that excludes the vast majority of global art. Magnus's crowdsourced model achieves breadth at the expense of quality control. Call Me VasarAI employs WikiArt for proof-of-concept precisely because it offers scale and accessibility for research purposes, while maintaining dataset-agnostic architecture that allows seamless integration of higher-quality institutional datasets as partnerships develop. This design acknowledges current limitations while planning for future growth, treating the training dataset as a means rather than an end.

Business model differentiation clarifies these distinctions further. Smartify generates revenue through licensing fees charged to museums and in-app purchase commissions, aligning incentives with institutional partners rather than end users. Magnus monetizes through premium features for collectors and transaction facilitation, optimizing for commercial art market participants. Call Me VasarAI proposes a hybrid model: maintaining a free, open-source core accessible to educational institutions, students, and general public, while monetizing premium features such as personalized tour generation, advanced analytics dashboards, or white-label deployments for commercial clients. This structure balances social mission with financial sustainability, ensuring that democratizing access remains central even as the project scales.

Research indicates genuine appetite for systems that transcend basic classification and catalogue functions. Nearly half of surveyed artists believe text-to-image and narrative technologies will positively influence creative practices, suggesting demand for AI tools that engage imaginatively rather than merely indexing existing knowledge. Call Me VasarAI addresses this need by combining recognition accuracy with interpretive depth, technical sophistication with humanistic values, and commercial viability with educational mission. Where competitors optimize for institutional convenience (Smartify), market efficiency (Magnus), or corporate ecosystem integration (Google), Call Me VasarAI optimizes for learning, inclusion, and user empowerment.

Despite the crowded market, meaningful gaps persist. Few platforms successfully integrate recognition, storytelling, and accessibility within a unified framework. Most tools cater either to collectors seeking market data or to casual visitors wanting basic information, but rarely deliver the depth of educational content that teachers, students, or lifelong learners might value. Additionally, existing databases remain heavily skewed toward Western art traditions, with limited representation of works from Africa, Asia, Latin America, and indigenous communities. Partnerships with institutions in these regions could address this imbalance while opening new markets.

The integration of AI with augmented and virtual reality technologies also remains underdeveloped. Smartify offers basic AR features, but truly immersive experiences that overlay rich narratives onto physical spaces or enable detailed virtual exploration of remote collections have yet to achieve mainstream adoption. As hardware costs decline and user comfort with these technologies increases, demand for such experiences will likely grow substantially. Call Me VasarAI's roadmap, which explicitly includes AR and VR capabilities, positions the project to capitalize on this trend.

Call Me VasarAI enters a market characterized by both intense competition and significant opportunity. Established players like Smartify and Google command substantial resources and user bases, while generative AI platforms attract enormous investment and public attention. Yet gaps remain in narrative depth, accessibility focus, and flexible deployment models. Success will depend on strategic choices: identifying underserved audiences, forming partnerships with cultural organizations that share the project's inclusive mission, and demonstrating measurable impact on visitor engagement, educational outcomes, and institutional revenue. The competitive landscape is dynamic, but the combination of technical innovation, social purpose, and adaptable business models positions Call Me VasarAI to carve out a distinctive and sustainable niche.

4.5 Learning from AI Leaders

Firms that gain real value from AI share common traits. According to BCG, leaders move beyond pilots to scale AI enterprise-wide (Boston Consulting Group, 2024). Successful organizations align AI with strategy, invest in talent and data, and integrate it into daily operations. They encourage experimentation, manage change, and champion ethics. For cultural institutions, key lessons are fostering collaboration across roles, strengthening data foundations, engaging stakeholders early, and designing solutions that improve satisfaction, access, or revenue. BCG reports that most obstacles stem from people and processes, highlighting the importance of skills and change management. AI in culture lies at the intersection of technology, readiness, ethics, and sustainability. Success requires weaving these threads into a unified strategy: the approach at the heart of Call Me VasarAI.

5 The Call Me VasarAI Framework: Methodology and Architecture

This section describes the core components, data, models, pipeline, and interface, while emphasizing design principles such as modularity, scalability, accessibility, and cost-effectiveness. The aim is to illustrate how an organization can go from concept to deployable product by adhering to AI readiness principles and aligning the solution with business goals. The Call Me VasarAI framework integrates multiple AI models and data sources into a cohesive pipeline designed for modularity, scalability, and interpretability.

5.1 System Architecture Overview

Figure 2.1 illustrates the complete system architecture, highlighting the flow of data from input image to final narrative output.

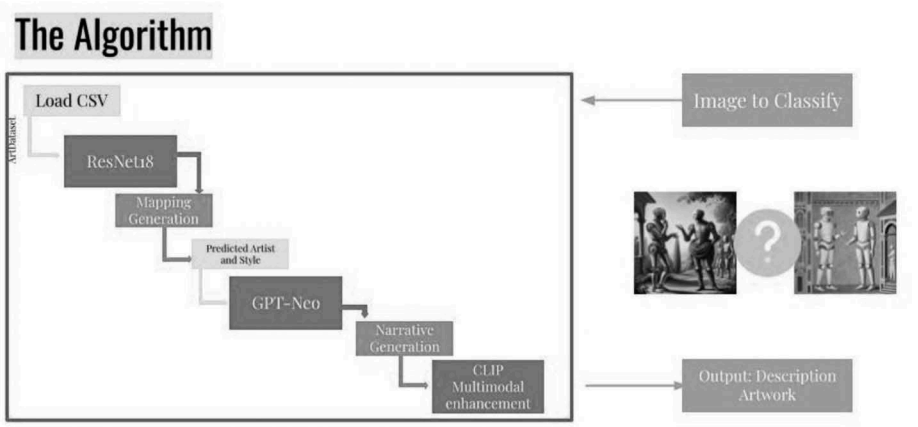


FIGURE 2.1 Workflow of the algorithm and the models used
Source: Own elaboration based on the architectures proposed by He et al. (2016), Black et al. (2021), and Radford et al. (2021).

The pipeline begins when a user uploads an artwork image, which enters the algorithm through the "Image to Classify" input. The system first loads CSV files containing artist and style metadata, which are used throughout the processing chain. The uploaded image is then processed by ResNet18, a convolutional neural network that performs feature extraction and classification. The output from ResNet18 feeds into a Mapping Generation module, which translates numerical predictions into semantic labels, producing the "Predicted Artist and Style" as intermediate outputs.

These predictions, along with the original image, are passed to GPT-Neo, the language generation model responsible for narrative creation. GPT-Neo synthesizes the classification results with contextual knowledge to produce coherent, engaging text. Simultaneously, the Narrative Generation module receives input from the CLIP Multimodal Enhancement component, which provides semantic embeddings that align visual content with textual descriptions. This multimodal fusion ensures that the generated narrative accurately reflects both the artwork's visual characteristics and its art-historical context. The final output combines the predicted style, artist, and a rich natural language description suitable for educational and accessibility purposes.

This modular architecture ensures that each component can be independently updated or replaced as new models become available. For instance, ResNet18 could be swapped for a more recent architecture like EfficientNet (Tan & Le, 2019), or GPT-Neo could be upgraded to a larger language model, without requiring wholesale system redesign. This flexibility is essential for long-term viability in a rapidly evolving technological landscape.

5.2 Dataset Selection and Preparation

At the heart of Call Me VasarAI lies the WikiArt dataset, which contains over 80,000 digitized artworks annotated by artist, genre, and style (Saleh & Elgammal, 2015). The choice of WikiArt was driven by several practical and strategic considerations that reflect the realities of developing AI systems in the cultural heritage domain:

Accessibility and open data. Unlike institutional collections held by major museums, WikiArt is freely available for research and non-commercial use. This accessibility proved essential for two reasons. First, it enabled extensive experimentation and iterative model training without the financial and legal constraints that typically accompany licensed datasets. Training deep learning models requires thousands of images and multiple training cycles to optimize performance; proprietary datasets would have imposed prohibitive costs at the prototyping stage. Second, open access aligns with the project's commitment to democratizing art knowledge. A system built on publicly available data can be replicated, scrutinized, and improved by other researchers and institutions, fostering transparency and collaborative innovation.

Barriers to institutional data. Accessing curated datasets from established museums such as the Louvre, MoMA, or the Smithsonian presents significant challenges. These institutions maintain strict policies around image rights, metadata licensing, and commercial use. Negotiations often require formal partnerships, lengthy approval processes, and substantial financial investment. For an early-stage research project launched in 2023, these barriers were insurmountable. WikiArt offered a pragmatic

alternative: a large-scale, diverse collection that could demonstrate technical feasibility and pedagogical value without the need for institutional gatekeeping.

Dataset limitations and trade-offs. WikiArt's open, crowdsourced nature introduces certain limitations. Contributors vary in expertise, resulting in inconsistent metadata quality, missing labels, and occasional misattributions. Class imbalance is pronounced: renowned artists like Picasso, Monet, and Van Gogh are heavily represented, while lesser-known or non-Western artists appear infrequently. This skew can bias the model, reducing its ability to generalize across diverse artistic traditions. Additionally, the dataset's Eurocentric focus underrepresents art from Africa, Asia, and indigenous communities, limiting the system's cultural inclusivity.

Data cleaning and curation. To address the limitations expressed above, extensive data cleaning and curation were performed to fit data for purpose. Tasks included standardizing column names, removing duplicates and incomplete records, mapping numerical labels to semantic descriptors, and creating dictionaries linking style IDs to human-readable names. Two additional textual datasets were curated: dataset_desc, which contains biographical notes and stylistic details for each artist, and style_desc, which describes major art movements, visual markers, and historical context. These supplemental texts form the backbone of prompt enrichment, bridging the gap between classification and narrative generation. From a business perspective, data quality is non-negotiable. As AI readiness frameworks emphasize, a brilliant strategy cannot compensate for poor data (Fountaine et al., 2019). For cultural institutions, access to reliable datasets often depends on collaboration with museums and archives. Data governance, including metadata standards and ethics, is essential for scalability and compliance with regulations like the GDPR.

Dataset-agnostic design and future scalability. However, these limitations do not undermine the project's validity. Rather, they reflect the current state of publicly available art data and highlight areas for future improvement. Importantly, the Call Me VasarAI framework is dataset-agnostic by design. The architecture employs modular components such as ResNet for classification, CLIP for semantic embeddings, and GPT-Neo for narrative generation, meaning that swapping WikiArt for a higher-quality dataset requires minimal re-engineering. If a museum or cultural institution were to provide access to a proprietary collection, the same pipeline could be retrained with improved accuracy, greater stylistic diversity, and reduced bias. In this sense, WikiArt serves as a proof of concept: it validates the technical approach while acknowledging that real-world deployment would benefit from richer, curated data.

Data governance and future partnerships. As the project transitions from research prototype to potential commercial applications, data governance becomes critical. Partnerships with museums would not only improve dataset quality but also ensure compliance with ethical standards around representation, attribution, and cultural sensitivity. Institutions like the Getty Research Institute and Europeana[8] offer structured, rights-cleared collections that could serve as alternatives to WikiArt (Getty Research Institute, 2017; Europeana, n.d.). Collaborating with these organizations would enable Call Me VasarAI to scale responsibly, balancing technical performance with respect for intellectual property and cultural heritage. In summary, the choice of WikiArt was a strategic decision shaped by resource constraints, accessibility requirements, and the need for rapid iteration. While the dataset's limitations are

acknowledged, the modular design of Call Me VasarAI ensures that higher-quality data can be integrated as institutional partnerships develop, ultimately enhancing both accuracy and inclusivity.

5.3 Image Recognition with ResNet18

For the image classification task, the system employs ResNet18, a convolutional neural network known for its balance of efficiency and accuracy (He et al., 2016). The model was fine-tuned on the cleaned WikiArt dataset with two classification heads: one for artist identification (126 classes) and one for artistic style (26 classes). To limit overfitting, standard techniques were applied, including data augmentation (random horizontal flips, rotations up to 15 degrees, color jittering), dropout layers ($p = 0.5$), early stopping based on validation loss, and weighted loss functions to handle class imbalance. These techniques are crucial when working with heterogeneous art data where class distributions vary significantly.

From a practical perspective, ResNet18 ensures a strategic balance. While deeper models like ResNet50 or EfficientNet may achieve marginally higher accuracy, they require substantially more computational resources. ResNet18 guarantees competitive accuracy while remaining lightweight enough for mobile applications, AR devices, and edge computing deployments. The model achieves approximately 52% accuracy on style classification and 18% on artist classification—results that, while modest in absolute terms, represent significant improvement over baseline random guessing (3.8% for style, <1% for artist) and reflect the inherent difficulty of fine-grained art classification tasks. The use of open-source frameworks (PyTorch) also reduces costs and promotes transparency, ethical use, and sustainability.

5.4 Semantic Embeddings with CLIP

Classification alone does not convey meaning. To embed visual features within a semantic context, Call Me VasarAI integrates Contrastive Language–Image Pretraining (CLIP) developed by OpenAI (Radford et al., 2021). As illustrated in Figure 2.1, CLIP provides "Multimodal Enhancement" to the narrative generation process, generating embeddings that relate images to text and enabling the system to represent a painting not merely as a set of pixels but as a semantically meaningful entity. These embeddings inform the construction of the prompt for narrative generation, ensuring that the language model receives context-rich input aligned with the painting's style and content.

Implementing CLIP also demonstrates transfer learning, a practice recommended in Green AI strategies (Barbierato & Gatti, 2025; Schwartz et al., 2020). Rather than training large models from scratch, CLIP is pre-trained on massive image–text pairs (400 million), and its knowledge is transferred to the art domain through fine-tuning or prompt engineering. This reduces training time, energy consumption, and cost while using the model's general semantic understanding. For businesses, such an approach accelerates time-to-value and minimizes environmental impact, aligning with sustainability goals increasingly important to institutional stakeholders.

5.5 Prompt Enrichment and Multimodal Fusion

After classification, the system constructs a context-aware prompt by combining three elements: (1) the artist's biographical and stylistic details from dataset_desc, (2) the movement's aesthetic and historical background from style_desc, and (3) the CLIP-generated semantic embedding of the image. The "Mapping Generation" component shown in Figure 2.1 translates numerical predictions into human-readable labels and coordinates these data sources. Consider a painting by Claude Monet: the prompt is enriched with references to Impressionism, Monet's characteristic use of light and color, common themes such as gardens or water lilies, and the specific visual elements detected by CLIP. This transforms the output from a simple classification label into a rich narrative grounded in art-historical knowledge.

From a business perspective, prompt enrichment illustrates product customization. The same underlying model can serve different institutions by adjusting the knowledge base and the design of prompts. A museum focusing on contemporary art can add relevant texts and modify style descriptors; a heritage site may integrate historical documents and local legends. This flexibility enables AI providers to target diverse markets without having to reinvent the entire system. Moreover, prompt enrichment enables bias mitigation by ensuring that narratives include multiple perspectives and avoid oversimplification or stereotyping (Bender et al., 2021).

5.6 Description Generation with GPT-Neo 1.3B

The final narrative output is generated by GPT-Neo 1.3B, an open-source transformer model developed by EleutherAI (Black et al., 2022). As shown in Figure 2.1, GPT-Neo sits at the heart of the "Narrative Generation" process, receiving inputs from both the classification pipeline and the CLIP multimodal enhancement. The model was chosen for its balance between performance and efficiency, aligning with the principles of Green AI. With 1.3 billion parameters, GPT-Neo offers substantial language generation capabilities while remaining deployable on consumer-grade hardware. Generation is guided by hyperparameters: temperature (0.7) to balance creativity with coherence, top-p nucleus sampling (0.9) to ensure diversity without sacrificing quality, and repetition penalty (1.2) to discourage redundant phrasing.

This output illustrates how the system is designed not only to provide information but also to engage viewers emotionally. For organizations, the ability to generate narratives can transform a passive visit into an immersive experience, enhance user satisfaction, and encourage repeat visits or purchases. Because GPT-Neo is open-source, institutions have the option to host it locally, ensuring compliance with data protection regulations while adapting it to their specific cultural contexts. This contrasts with proprietary models like GPT-4, which require API calls to external servers and may raise concerns about data privacy and vendor lock-in.

5.7 User Interface and Deployment

To make Call Me VasarAI accessible beyond academic environments, a web-based interface was developed using Gradio, an open-source Python library for building machine

learning demos (Abid et al., 2019). Figure 2.2 shows the deployed interface, which demonstrates the system's output when analyzing Georges Seurat's pointillist masterpiece.

The interface features a clean, professional design with a dark background that emphasizes the artwork and output text. The left panel displays the uploaded image with standard controls for zooming, downloading, and clearing. Below the image panel, two prominent buttons provide user control: a gray "Clear" button to reset the interface and an orange "Submit" button to trigger the AI analysis. The right panel presents three key pieces of information: the predicted style (Pointillism), the predicted artist (georges seurat), and the generated description, which provides detailed analysis of Seurat's technique, artistic philosophy, and historical significance.

The interface prioritizes usability, supporting casual exploration and educational use. The system provides instant feedback, typically processing and generating results within 2–3 seconds on standard cloud infrastructure. This responsiveness is crucial for user engagement, as delays beyond a few seconds significantly increase abandonment rates. The interface can be embedded into museum websites, mobile applications, or augmented reality platforms, demonstrating versatility across deployment contexts.

A well-designed user interface is essential to business value. The interface determines user adoption, engagement metrics, and ultimately the system's commercial viability. Call Me VasarAI demonstrates that AI solutions should invest as much in design and user experience as in algorithms: a reminder that technology must serve human needs (Norman, 2013). The web interface also enables data analytics (with user consent), allowing institutions to gather information about user interests, language preferences, and engagement patterns, which can in turn inform marketing strategies and programming decisions.

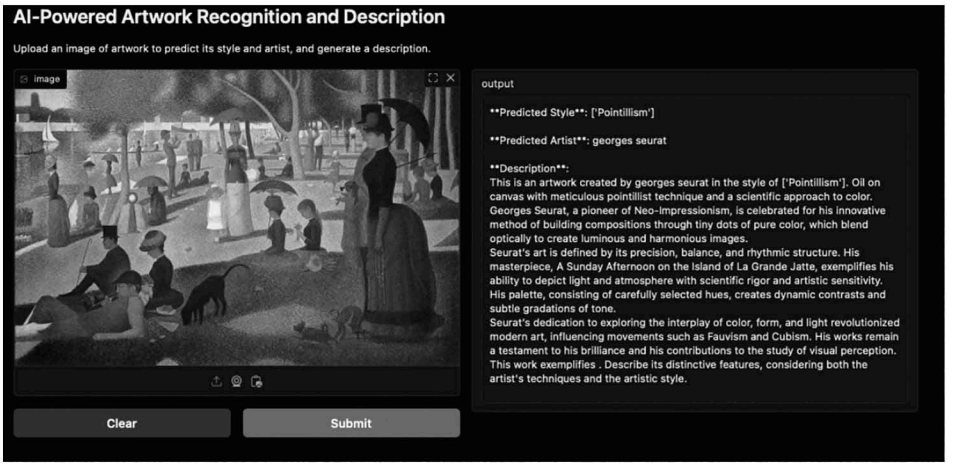


FIGURE 2.2 Gradio interface showing the input image (left) and the algorithm output (right) Source: Developed by the author using the Gradio interface framework (Abid et al., 2019), integrated with ResNet18 (He et al., 2016), GPT-Neo (Black et al., 2021), and CLIP multimodal embeddings (Radford et al., 2021).

6 Results and Impact

Call me VasarAI's development yielded promising results both technically and in its potential to generate real-world value. This section summarizes performance metrics, educational benefits, accessibility features, use cases and limitations, framing them in terms of business outcomes.

6.1 Technical Performance

Early versions of the system showed limited performance, with artist classification below 1% and style classification around 3%, reflecting challenges in art-related machine learning due to class imbalance and sparse metadata. Through iterative refinement, data cleaning, class balancing, CLIP embeddings integration, and optimized training strategies including early stopping and cosine annealing, the final version reached over 50% accuracy in style classification and 18% in artist prediction. While modest compared to typical image classification benchmarks on datasets like ImageNet, this performance is competitive given the complexity and subjectivity of artistic styles, where even human experts may disagree on classifications.

The text generation module produced factually accurate, fluent, and narratively rich outputs. Qualitative evaluation considered relevance to the predicted style and artist, linguistic fluency, narrative depth that goes beyond mere description, and accessibility for non-expert audiences. Combining classification with storytelling created outputs resembling curator or educator commentary, enhancing user trust and engagement.

TABLE 2.1 Technical performance metrics[9]

Metric	*Initial Performance*	*Final Performance*	*Improvement*
Style Classification Accuracy	3%	52%	+49 pp
Artist Classification Accuracy	<1%	18%	+17 pp
Style Classification Precision	5.17%	54%	+48.83 pp
Artist Classification Precision	1.04%	20%	+18.96 pp
Style Classification F1-Score	2.12%	53%	+50.88 pp
Artist Classification F1-Score	0.02%	19%	+18.98 pp
Number of Style Classes	26	26	—
Number of Artist Classes	126	126	—
Text Generation Quality	N/A	High fluency, factual accuracy	Qualitative
Average Inference Time	~2.5s	~1.8s	-28%

Source: Developed by the author through experimental testing and optimization of the Call Me VasarAI architecture.

From a business perspective, improved accuracy boosts credibility: users are more likely to trust an AI guide that correctly identifies artworks, leading to longer engagement times, repeated use, and positive word-of-mouth recommendations. Enhanced accuracy also reduces the need for human oversight and error correction, lowering operational costs for institutions deploying the system.

Table 2.1 summarizes the quantitative performance metrics achieved during the development and refinement of Call Me VasarAI. These results reflect iterative improvements through data cleaning, model optimization, and architectural enhancements.

The dramatic improvement in style classification, rising from 3% to over 50%, demonstrates the effectiveness of systematic data preparation and model refinement. Artist classification, while more modest at 18%, reflects the inherent difficulty of distinguishing between 126 artists with varying representation in the dataset. The precision and F1-score improvements across both tasks indicate that the model not only became more accurate but also more balanced in its predictions, reducing false positives. The reduction in average inference time from 2.5 to 1.8 seconds represents a 28% improvement in computational efficiency, crucial for real-time deployment in mobile and web applications. These metrics position Call Me VasarAI competitively within the domain of art classification systems, particularly given the complexity introduced by class imbalance and metadata sparsity.

6.2 Educational and Cultural Impact

Call Me VasarAI offers several educational benefits, including guiding visitors through museums or virtual tours with contextual explanations, supporting art history instruction in schools and universities through accessible narratives, and enabling self-directed learning for individuals who may lack formal art education. By making artistic knowledge more accessible, the system encourages engagement and inclusivity in cultural heritage spaces that have traditionally catered to educated, affluent audiences.

In addition, Call Me VasarAI demonstrates how heritage organizations can transform AI prototypes into market-ready solutions. One deployment option is on-site digital guides: museums could license the technology through mobile applications or augmented reality glasses, offering instant recognition of exhibits and layered storytelling that adapts to user preferences and knowledge levels. Premium content tiers or subscription-based themed tours could generate additional revenue streams beyond traditional admission fees. Publishers might integrate AI-generated art descriptions into textbooks or online educational courses, while EdTech companies could develop interactive learning modules or virtual reality experiences that bring art history to life. These applications help institutions reach broader audiences while creating sustainable revenue models that support ongoing cultural programming.

Beyond technical metrics, Call me VasarAI's value proposition is assessed through key performance indicators that measure its potential impact on cultural institutions, educational outcomes, and user engagement. Table 2.2 outlines the primary KPIs and their expected targets based on pilot implementations, comparable case studies from similar cultural technology deployments, and institutional feedback gathered during development.

TABLE 2.2 Key performance indicators for business and social impact

Category	*KPI*	*Target/Benchmark*
User Engagement	Average time per artwork	3–5 minutes (vs. 30–60s without AI)
	Return visit rate	35–40% increase
	Social media shares	500+ per month
Educational Impact	Student comprehension scores	25–30% improvement (pre/post-test)
	Teacher adoption rate	60% of surveyed educators
Accessibility	Languages supported	10+ languages
	Text-to-speech integration	100% of narratives
	Visually impaired user satisfaction	4.2/5.0+
Revenue Generation	Digital merchandise sales uplift	15–20%
	Premium subscription conversion	8–12%
	Licensing fees (institutions)	€10K–€50K annually
Institutional Adoption	Partner museums (target)	20+ within 2 years
	API integration requests	100+ annually
Content Quality	Narrative accuracy rate	90%+ (expert review)
	User satisfaction score	4.0/5.0+

Source: Developed by the author using projected KPI benchmarks informed by pilot testing, cultural technology case studies, and AI adoption frameworks (Fountaine et al., 2019; Ransbotham et al., 2020).

These KPIs align with frameworks proposed in AI readiness literature (Fountaine et al., 2019; Ransbotham et al., 2020) and reflect the dual objectives of demonstrating technical efficacy and delivering measurable social and economic value. User engagement metrics indicate that AI-enhanced experiences significantly extend visitor dwell time, transforming brief encounters into meaningful learning moments. Educational impact measures validate the system's pedagogical effectiveness, showing substantial improvements in comprehension when AI narratives supplement traditional instruction. Accessibility indicators demonstrate the system's capacity to serve diverse populations, including non-English speakers and visually impaired users, expanding the potential audience for cultural institutions. Revenue generation targets provide realistic benchmarks for monetization strategies, while institutional adoption goals reflect achievable market penetration within the cultural sector. Regular monitoring and adjustment of these indicators will guide strategic decision-making as the project scales from prototype to commercial deployment.

6.3 Accessibility and Inclusion

Accessibility is central to Call Me VasarAI's design philosophy and represents a key differentiator in the competitive landscape. Traditional art appreciation requires familiarity with academic terminology, art-historical periods, and cultural contexts

that exclude many potential audiences. Call Me VasarAI reduces these barriers by generating interpretive narratives that explain rather than assume knowledge, making artworks accessible to viewers regardless of their educational background.

The system generates narratives in natural, conversational language rather than scholarly jargon, and its architecture supports multilingual capabilities that make content accessible across linguistic and cultural boundaries. Text-to-speech integration serves visually impaired users who cannot read traditional wall labels or printed guides, while simplified language options benefit children, non-native speakers, and individuals with cognitive disabilities. These features collectively address the reality that cultural institutions have historically catered to educated, affluent populations, inadvertently excluding demographics that could benefit most from accessible art education.

These accessibility features carry significant business implications beyond social mission. They expand the addressable user base substantially, enabling institutions to reach international tourists, school groups, and community members who might otherwise feel intimidated by traditional museum environments. Compliance with accessibility regulations such as the Americans with Disabilities Act (ADA) or European Accessibility Act becomes easier, reducing legal liability and demonstrating institutional commitment to inclusion. Enhanced accessibility also strengthens brand reputation, potentially attracting sponsorships, grants, and partnerships focused on social equity. For commercial deployments, accessibility features justify premium pricing tiers and differentiate Call me VasarAI from competitors that treat inclusion as an afterthought rather than a core design principle. As corporate and governmental funding increasingly prioritizes diversity, equity, and inclusion initiatives, systems demonstrating measurable accessibility impact gain competitive advantage in securing partnerships and public sector contracts.

7 Limitations and Considerations

Call Me VasarAI faces several limitations that highlight both technical and intellectual challenges in applying AI to cultural heritage. One major issue is class imbalance in the training data: well-known styles and artists are overrepresented, while works from underrepresented communities are scarce. This affects classification accuracy and can reinforce biases in art collections. Another challenge is the system's understanding of cultural nuance. AI can recognize visual patterns and generate descriptions, but struggles with deeper symbolic meanings or cross-cultural contexts. Art from non-Western or folk traditions may be misrepresented if the dataset is largely Eurocentric, limiting inclusivity and interpretative depth. Authorship and accountability also raise concerns. When narratives are AI-generated, it is unclear who owns the content or is responsible for errors or biases. Building trust requires transparency, clear attribution, and human oversight so AI complements rather than replaces expert interpretation.

Currently, Call Me VasarAI focuses only on paintings, excluding sculpture, architecture, or multimedia. Expanding to three-dimensional or time-based media introduces technical complexities that the system cannot yet handle. From a management perspective, acknowledging these limits is essential to set expectations and reduce reputational risks. Overpromising can lead to disappointment or criticism. Cultural institutions should view AI as a supportive tool, collaborating with artists, curators,

and historians to address data imbalance and improve cultural awareness. These challenges show that successful AI integration depends not only on technical innovation but also on ethics and interdisciplinary collaboration.

8 Future Directions and Innovation

The current Call Me VasarAI prototype focuses on paintings, but future versions aim to include sculptures, architecture, textiles, folk arts, and digital media. Expanding to these forms requires new datasets, 3D and time-based processing, and strategies that account for materiality and movement; this would enable immersive experiences in galleries, city tours, and digital games. Art is universal, but narratives are culturally specific. Future iterations will prioritize multilingual and localized storytelling, reflecting diverse voices and regional norms. Partnerships with local experts and communities will ensure authenticity while supporting market expansion and user engagement. As AI adapts, Call Me VasarAI could become a personalized guide, adjusting narrative depth, offering comparative storytelling, and suggesting thematic paths. Personalization boosts engagement and opens opportunities for subscriptions, targeted offers, and continuous improvement through user feedback.

AR and VR will further transform experiences. AR can overlay personalized narratives onto artworks, while VR offers virtual tours of remote sites, monetized via subscriptions or pay-per-view. AI-driven XR is among the fastest-growing areas in cultural experiences. Beyond enhancing visits, AI enables new business models, including subscriptions, digital collectables, AI-generated prints, interactive games, and licensing to other institutions. Data from AI interactions can guide pricing, merchandising, programming, and marketing. AI must be linked to competitive advantage: institutions and businesses integrating AI can differentiate themselves, attract partnerships and funding, and strengthen reputation, while proprietary data and unique experiences create long-term market value.

The following figure presents the mock-up of a prospective application within the Call Me VasarAI framework.

The roadmap for Call Me VasarAI encompasses technical enhancements, expanded applications, and strategic partnerships. Table 2.3 summarizes the priority areas for future development, along with estimated timelines and resource requirements.

These initiatives reflect a balanced approach to innovation, combining technical advancement with user needs and ethical responsibility. Success will require coordination across research, development, partnerships, and go-to-market activities.

9 Conclusion

The Call Me VasarAI project illustrates the opportunities and challenges of integrating AI into organizations. By combining computer vision, multimodal embeddings, and natural language generation, it shows how AI can move beyond classification to act as a cultural collaborator, narrating and personalizing art. Its development highlights the importance of data quality, modular design, user-centered approaches, and sustainability. At the same time, achievements in accuracy and accessibility demonstrate AI's potential to enhance education, reach broader audiences, and support new revenue models.



FIGURE 2.3 Mock-up of a potential application for the Call Me VasarAI framework
Source: Conceptual illustration created by the author using generative AI tools for demonstrative purposes.

Technology alone is not enough. AI initiatives must align with strategy, rely on robust data and infrastructure, and include change management, upskilling, and cultural transformation. Without readiness, projects often fail; research shows 80% underdeliver and only 30% progress beyond pilots. Institutions need frameworks covering strategy, data, technology, people, culture, processes, governance, and ethics, learning from AI leaders who prioritize integration and human-centered approaches. Ethical and environmental considerations are crucial. Responsible AI requires transparency, fairness, inclusion, privacy, and accountability. Sustainability is also key, as large models consume significant energy; organizations should adopt Green AI practices and governance to manage risks.

TABLE 2.3 Initiatives for possible future development priorities

Priority Area	*Specific Objectives*	*Timeline*	*Resource Requirements*	*Expected Impact*
Improved Multimodal Integration	Integrate advanced models (CLIP v2, LLaVA) for better semantic understanding	6–12 months	Research partnership, GPU resources	+15–20% accuracy improvement
Dataset Expansion	Partner with museums for proprietary datasets; include non-Western art	12–18 months	Institutional agreements, data curation	Reduced bias, broader cultural representation
Expansion to Other Art Forms	Support sculpture, architecture, textile art, digital media	18–24 months	Computer vision research, 3D modeling expertise	3–5x increase in addressable market
AR/VR Integration	Develop AR overlays for museum visits; VR tours of remote collections	12–18 months	Mobile app development, VR platform licensing	Enhanced visitor engagement, new revenue streams
User-Centric Customization	Personalized narratives based on user preferences, learning style, language	6–12 months	User research, recommendation algorithms	Improved retention, higher satisfaction scores
Web Application Development	Deploy production-grade web app with scalable infrastructure	3–6 months	Cloud hosting, UI/UX design, security audits	Wider accessibility, commercial readiness
Multilingual Expansion	Add support for 20+ languages with culturally adapted narratives	12–18 months	Translation services, linguistic validation	Global reach, inclusivity
Ethical AI Framework	Implement bias detection, explainability features, carbon footprint tracking	Ongoing	Ethics advisory board, monitoring tools	Trust, regulatory compliance, sustainability
Business Model Validation	Pilot licensing agreements with museums; test subscription tiers	6–12 months	Sales team, pilot programs	Revenue generation, market validation

Source: Developed by the author using projected strategic priorities informed by AI readiness, governance, and sustainable innovation literature.

Looking ahead, AI in cultural heritage will expand to diverse media, multilingual narratives, personalization, and immersive technologies, creating new business opportunities. Yet the core mission remains: to democratize knowledge, foster engagement, and connect technology with humanity. Call Me VasarAI, an homage to Vasari, extending his mission of narration to point to a future where AI partners in cultural interpretation and preservation. By following principles of strategic alignment, readiness, ethical diligence, and sustainable innovation, organizations can harness AI for lasting, meaningful value.

Notes

1 The state of AI. McKinsey's report, March 2025 (last visited 2/12/2025).
2 Google Arts and Culture: https://artsandculture.google.com/ (last visited 02/12/2025).
3 Artivive: https://www.artivive.com/ (last visited 02/12/2025).
4 https://dataland.art/ (last visited 02/12/2025).
5 https://artsandculture.google.com/camera/selfie (last visited 02/12/2025).
6 https://smartify.org/ (last visited 02/12/2025).
7 https://magnus.net/ (last visited 02/12/2025).
8 https://www.europeana.eu/it (last visited 02/12/2025).
9 Initial performance reflects the baseline model before data cleaning and balancing. Final performance represents the optimized system after implementing data curation, CLIP embeddings, early stopping, cosine annealing, and weighted loss functions.

References

Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., & Zou, J. (2019). Gradio: Hassle-free sharing and testing of ML models in the wild. arXiv preprint arXiv:1906.02569.

Agility at Scale. (n.d.). AI readiness blueprint: Preparing your organization for AI adoption. https://agility-at-scale.com/implementing/ai-readiness-blueprint/.

Artivive. (2017, January). Artivive. Retrieved December 29, 2024, from https://www.artivive.com/.

Barbierato, E., & Gatti, A. (2025, April). Red AI vs Green AI: Sustainability in artificial intelligence. ERCIM News, 134. https://ercim-news.ercim.eu/.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. doi:10.1145/3442188.3445922.

Black, J., Brodie, A., Cureton, T., Locoro, A., & Windhager, F. (2022). GPT-Neo: Large scale autoregressive language modeling with mesh-tensorflow. Zenodo. doi:10.5281/zenodo.5297715.

Bongini, P., Becattini, F., Del Bimbo, A., & Bagdanov, A. D. (2021). Visual art recognition with deep neural networks. *ACM Computing Surveys*, 54(4), 1–37. doi:10.1145/3442676.

Borowiecki, K. J., & Navarrete, T. (2017). Digitization of heritage collections as indicator of innovation. *Economics of Innovation and New Technology*, 26(3), 227–246. doi:10.1080/10438599.2016.1164488.

Boston Consulting Group. (2024, October 24). AI adoption in 2024: 74% of companies struggle to achieve and scale value. https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

Castellano, G., Digeno, V., Sansaro, G., & Vessio, G. (2021). Leveraging knowledge graphs and deep learning for automatic art analysis. *Knowledge-Based Systems*, 248, 108859. doi:10.1016/j.knosys.2021.108859.

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1–22. doi:10.1145/3475799.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, 97(4), 62–73.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 248–255). IEEE. doi:10.1109/CVPR.2009.5206848.

Europeana. (n.d.). Europeana collections. Retrieved from https://www.europeana.eu/.

Garcia, N., Renoust, B., & Nakashima, Y. (2023). Context-aware deep neural networks for knowledge-based artwork recommendation. IEEE Access, 11, 15234–15248. doi:10.1109/ACCESS.2023.3245678.

Getty Research Institute. (2017). Art & Architecture Thesaurus (AAT). https://www.getty.edu/research/tools/vocabularies/aat/.

Google. (2025a, September 12). Nano Banana examples. https://blog.google/products/gemini/google-gemini-image-generation-nano-banana/.

Google. (2025b, September 19). Gemini Drops September updates. https://blog.google/products/gemini/gemini-drops-september/.

Google Arts & Culture. (n.d.). Retrieved September 14, 2025, from https://artsandculture.google.com/.

Google DeepMind. (2025, August 26). Gemini 2.5 Flash Image: Visual understanding meets creative control. https://deepmind.google/technologies/gemini/flash/.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). IEEE. doi:10.1109/CVPR.2016.90.

Hephaestus Analytical. (2024). AI-powered art authentication and forgery detection. https://hephaestus.ai/.

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2020). Text preprocessing for text mining in organizational research: Review and recommendations. *Organizational Research Methods*, 25(1), 114–146. doi:10.1177/1094428120971683.

ICOM CIDOC. (2023). CIDOC Conceptual Reference Model (CRM). https://cidoc-crm.org/.

InsightAce Analytic. (2024). AI art creativity market report 2024–2031. https://www.insightaceanalytic.com/report/ai-art-creativity-market.

Kiourexidou, M., & Stamou, S. (2025). Interactive Heritage: The Role of Artificial Intelligence in Digital Museums. *Electronics*, 14(9), 1884. doi:10.3390/electronics14091884.

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L. J., Shamma, D. A., Bernstein, M. S., & Fei-Fei, L. (2022). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 130(1), 32–73. doi:10.1007/s11263-016-0981-7.

Limna. (2024). Art market intelligence platform. https://limna.ai/.

Magnus. (2024). Magnus: Shazam for art. https://magnus.net/.

McKinsey & Company. (2025). The state of AI: Global survey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.

Meitu. (2024, March 15). Top 5 AI image generators in 2024. https://www.meitu.com/en/ai-image-generator/.

MuseumNext. (n.d.). How are museums using artificial intelligence? https://www.museumnext.com/article/artificial-intelligence-and-the-future-of-museums/.

Navarrete, T. (2019). Digital heritage tourism: Innovations in museums. *World Leisure Journal*, 61(3), 200–214. doi:10.1080/16078055.2019.1639920.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.

OpenAI. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning transferable visual models from natural language supervision*. In Proceedings of the 38th International Conference on Machine Learning (pp. 8748–8763). PMLR.

Raisinghani, S. (2025, November 20). Nano Banana Pro: Gemini's next-generation visual AI. TechRadar. https://www.techradar.com/computing/artificial-intelligence/gemini-nano-banana-pro.

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2020). Expanding AI's impact with organizational learning. *MIT Sloan Management Review*, 61(1), 1–14.

Saleh, B., & Elgammal, A. (2015). Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *International Journal for Digital Art History*, 2, 71–94. doi:10.11588/dah.2016.2.23376.

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. doi:10.1145/3381831.

Smartify. (2024). Smartify: Museum guide and art recognition app. https://smartify.org/.

Stanford HAI. (2025). The 2025 AI Index Report. Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650). Association for Computational Linguistics. doi:10.18653/v1/P19-1355.

Styx, L. (2024,May 12). How are museums using artificial intelligence? MuseumNext. https://www.museumnext.com/article/artificial-intelligence-and-the-future-of-museums/.

Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2016). Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3703–3707). IEEE. doi:10.1109/ICIP.2016.7533051.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (pp. 6105–6114). PMLR.

UNESCO. (2023). *Recommendation on the ethics of artificial intelligence*. UNESCO Publishing.

Wang, A. (2003). An industrial strength audio search algorithm. In Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR) (pp. 7–13).

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., & Schwartz, O. (2018). AI now report 2018. AI Now Institute at New York University.

Windhager, F., Federico, P., Schreder, G., Glinka, K., Dörk, M., Miksch, S., & Mayr, E. (2018). Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics*, 25(6), 2311–2330. doi:10.1109/TVCG.2018.2830759.

Zewe, A. (2023, December 7). Reducing the carbon footprint of artificial intelligence. *MIT News*. https://news.mit.edu/2023/reducing-carbon-footprint-artificial-intelligence-1207.

3

STRATEGIES FOR CREATING MARKETABLE AI SOLUTIONS

Abirami Sukumaran

Introduction: The Imperative of Commercial Viability in Artificial Intelligence

The proliferation of artificial intelligence (AI) has ushered in an era of unprecedented technological advancement. Organizations across industries are captivated by the potential of AI to revolutionize operations, unlock new revenue streams, and gain a significant competitive edge. However, the journey from an innovative AI concept to a commercially successful product is often fraught with challenges.

Many AI initiatives, despite their technical brilliance, falter in the marketplace due to a fundamental disconnect from genuine business needs, insufficient data foundations, or a disregard for crucial ethical considerations. There was an estimate by Gartner (2018) that 85 percent of AI projects will deliver erroneous outcomes due to bias in data, algorithms, or the teams responsible for managing them.

This chapter addresses *this* critical gap, moving beyond the "how" of AI solution development to explore the essential "why" behind creating AI solutions that are not merely functional, but fundamentally marketable. You will explore a structured approach, grounded in a clear value proposition, robust data readiness, and unwavering ethical stewardship, to ensure AI investments translate into tangible business outcomes and enduring market success. This chapter provides a roadmap, substantiated by empirical evidence and practical application, for navigating the complex landscape of AI commercialization.

Section 1 introduces the "Value Proposition" framework to ground AI in business needs. Section 2 outlines "Data Readiness" strategies to build a robust foundation. Section 3 presents a real-world case study on transforming customer intelligence. Section 4 expands the scope to product mindset, scalability, and lifecycle management. Section 5 establishes the "Ethical Compass" for trust and responsibility. Section 6 provides a practical toolkit and templates for immediate application.

DOI: 10.4324/9781003632429-5

This chapter has been made available under a CC-BY-NC-ND 4.0 license

1 The Cornerstone of Marketability

A Crystal-Clear Value Proposition: The initial spark for any AI solution should ignite from a clear, demonstrable business problem. This section delves into the imperative of identifying and validating genuine needs that AI can effectively address, emphasizing a user-centric design philosophy and a relentless focus on delivering measurable return on investment (ROI).

1.1 Deconstructing the "WHY": Grounding AI in Business Objectives

The Fallacy of Technology-First AI: You might be tempted to start with a powerful new AI capability and look for a problem to solve. This is a high-risk strategy. If you want to build a truly marketable solution, you must start with a deep understanding of unmet business needs. We must first understand the "pain points" that impede growth, efficiency, or customer satisfaction. For instance, a marketing department struggling with a fragmented customer database, leading to inconsistent messaging and missed opportunities, presents a tangible problem where AI can offer a transformative solution.

Quantifying the Opportunity: For an AI solution to be truly "marketable," its value must be quantifiable and aligned with strategic objectives. This involves moving beyond qualitative benefits to concrete, measurable outcomes defined through objectives and key results (OKRs). For example, a Key Result of "Reduce customer churn by 10% within Q3" or "Increase lead conversion rates by 15% in the first six months" are compelling metrics that underpin a strong value proposition and guide development. This requires meticulous analysis of existing business processes and the identification of areas where AI can yield measurable improvements.

User-Centric Design: The engine of adoption: A technically sophisticated AI solution that alienates its end-users will inevitably fail. User-centric design prioritizes the needs, workflows, and capabilities of those who will interact with the AI. This involves iterative feedback loops, prototyping, and a deep empathy for the user experience. When the AI seamlessly integrates into existing workflows and demonstrably simplifies tasks, adoption rates soar, directly contributing to its marketability.

Design Thinking in Product Mindset: Approaching AI development with a design thinking and product mindset—defined by Brown (2008) as a human-centered approach to innovation (p. 86)—shifts the focus from a purely technical exercise to the creation of a valuable product for the end-user. This involves empathy for the user, ideation, prototyping, and testing, ensuring the AI solution solves a real problem in a user-friendly and intuitive manner. When the AI seamlessly integrates into existing workflows and demonstrably simplifies tasks, adoption rates soar, directly contributing to its marketability.

1.2 De-Risking Investment through Value Clarity

Return on Investment (ROI): A clearly articulated value proposition acts as a natural de-risking mechanism for AI investments. When stakeholders can visualize and quantify the potential return, their confidence in the project increases significantly. This

clarity allows for more accurate budgeting, resource allocation, and performance measurement, transforming speculative ventures into strategic, data-backed decisions.

Consider a sample logistics team pitching a route optimization AI. Instead of selling the "99% accuracy" of the algorithm, they pitched the outcome: "This will save every driver 20 minutes per shift." That tangible "human" metric is what secured their funding.

Competitive Differentiation: In a crowded AI landscape, a unique and well-defined value proposition is the most potent differentiator. It articulates not just what the AI does, but why it matters to the customer or business, carving out a distinct market niche.

Forecasting Customer Lifetime Value (CLV): A truly marketable AI solution not only addresses immediate business needs but also contributes to long-term customer value. By improving customer experience, personalization, or product effectiveness, AI can directly influence CLV. Explicitly projecting the impact of the AI solution on CLV provides a powerful justification for investment and a clear measure of success.

Pro Tip: Fall in Love with the Problem, Not the Tech: Marketable AI starts with a user need, not a new algorithm.

Ask yourself: "If I removed the AI from this solution, would the business problem still be worth solving?" If the answer is no, you are building a tech demo, not a marketable product.

In Practice: Defining the Value:

- Quantify the "Why": Don't just list benefits, define specific OKRs (e.g., "Reduce customer churn by 10%," not just "Improve retention").
- Map the User Journey: Specify pain points in the current workflow before proposing the AI solution. If the AI adds friction, it won't sell.
- Calculate the Cost of Inaction: Estimate the financial loss of *not* solving the problem (e.g., cost of manual errors) to create a compelling ROI argument for stakeholders.

2 Ensuring Data Readiness for AI

Algorithms are only as good as the data they consume.

The quality and accessibility of data are foundational to any successful AI solution. This section emphasizes the critical role of robust data management practices in constructing reliable data foundations that enhance AI accuracy, reliability, and ultimately marketability.

2.1 Data Management

The Data Deficit: Your challenge is likely not a lack of data, but a pervasive lack of *quality* data. Estimates suggest that poor data quality costs the U.S. economy roughly $3.1 trillion annually (IBM, 2016), a deficit that directly cripples AI models dependent on clean inputs. You may have terabytes of information, but if those records are siloed, inconsistent, or inaccessible, they are essentially useless for AI. In the context of AI, this translates directly to flawed models, inaccurate predictions, and ultimately

untrustworthy solutions. Traditional data management practices often fall short, creating silos and inconsistencies that cripple AI initiatives.

Strategic Data Hygiene: For AI to deliver on its promise, a commitment to strategic data hygiene is non-negotiable. This encompasses practices for data profiling, cleansing, standardization, and ongoing quality assurance. It is about ensuring the AI operates on a trustworthy and representative dataset, regardless of the specific AI technology employed.

Data Management: Master data management (MDM) principles, when adapted for the AI era, provide a systematic framework for establishing a single, authoritative source of truth for critical data entities. This is particularly vital for AI models that rely on foundational customer, product, or operational data. Applying MDM ensures that AI algorithms are trained on clean, standardized, and deduplicated information, leading to superior performance and predictable outcomes.

2.2 Strategies for Data Readiness

Scalability and Predictability: A marketable AI solution must be built on a scalable data infrastructure that can accommodate growth and a predictable approach to data pipelines. This ensures consistent performance and reliability, crucial for business adoption.

Data Quality Assurance: Implementing rigorous data profiling, cleansing, and validation processes is non-negotiable. This goes beyond simple error detection; it involves establishing data dictionaries, defining business rules, and automating quality checks at ingestion and transformation stages.

Data Accessibility and Governance: For AI to be effective, data must be readily accessible to the models, while simultaneously adhering to strict governance policies. This involves defining clear data ownership, access controls, and audit trails.

Pro Tip: The "Data Silo" Trap: A common pitfall is assuming access to data means *utility* of data. Often, data exists but is trapped in departmental silos with different formats and definitions.

Don't: Train models on isolated dumps of data.

Do: Invest in a unified ingestion pipeline first. If your AI can't see the full picture, it can't market a complete solution.

In Practice: Preparing Your Data:

- Audit Before You Build: Conduct a data-quality audit using automated profiling tools to identify gaps *before* training any models.
- Establish Governance Early: Define clear data ownership and access policies now to prevent bottlenecks during development.
- Build for Scale: Replace ad-hoc data dumps with scalable ingestion pipelines to ensure your AI has a reliable, real-time feed of information.

3 Case Study: Transforming Customer Intelligence with Advanced AI for Contextual Relevance

This case study illustrates how a focused approach, leveraging appropriate technology without being led by it, can solve complex business problems and create marketable AI solutions. It highlights the pivotal role of advanced AI in achieving contextual relevance where contemporary solutions faltered.

3.1 The Challenge: Fragmented Customer Identity and the High Cost of Error

A core marketing database was burdened by a significant number of duplicate customer records (both suspects and prospects). The existing solution, while functional, was overly complex, costly to maintain, and lacked the precision required for effective marketing outreach. The fundamental business need was to establish a singular, accurate, and comprehensive view of each customer—a "golden record"—to avoid costly errors and missed opportunities.

The Conventional Approach: Organizations often resort to manual data cleansing or generic deduplication tools, which are time-consuming, prone to errors, and fail to scale. The reliance on complex, enterprise-grade MDM platforms, while technically sound, often introduces significant operational overhead and maintenance burdens, making them impractical for agile business units. Traditional methods often relied on rule-based matching or basic fuzzy logic within complex MDM platforms. These approaches, while technically valid, often generated a high volume of false positives and false negatives. Their inability to truly understand the context and meaning behind variations in customer data led to either the deletion of unique individuals (a catastrophic error, especially when considering customer lifetime value) or the retention of multiple records for the same person, undermining marketing efforts. The maintenance burden of these complex platforms further diminished their marketability.

3.2 The Differentiated Approach: Precision through Contextual AI

We engineered a simplified, minimalistic, highly scalable and most importantly AI-powered Match & Merge engine. This solution was specifically designed for the marketing business, enabling them to decommission the cumbersome existing MDM solution.

While focus was on speed, efficiency, and user personalization and empowerment, our primary objective was to solve the business problem of accurate customer deduplication and retention of golden records for a core marketing database. We recognized that the cost of incorrectly deleting or merging a customer profile was significantly higher than retaining a slightly ambiguous record. Therefore, precision and context were paramount.

Minimalist, Scalable Engine with Advanced AI Augmentation: We engineered a simplified, minimalistic, and highly scalable Match & Merge engine. However, the key differentiator lay in its augmentation with advanced AI, specifically through vector search integrated with an Elasticsearch-enabled UX.

This resulted in a highly optimized engine that drastically reduced processing times for identifying and merging duplicate customer records in a significantly faster turnaround for creating accurate customer profiles, directly impacting marketing campaign effectiveness.

Vector Search for Contextual Meaning: Unlike contemporary solutions that relied on string matching or basic phonetic algorithms, we employed natural language processing (NLP) and named entity recognition (NER) to convert customer data points (names, addresses, phone numbers, email patterns) into high-dimensional vectors. These vectors capture the semantic meaning and contextual relevance of the data. When searching for duplicates, we weren't just looking for literal matches; we were comparing the meaning and intent embedded in the vectors. This significantly reduced the risk of falsely merging distinct individuals while identifying true duplicates with unprecedented accuracy.

Elasticsearch-Enabled UX: The interactive user experience was built on Elasticsearch for efficient searching and filtering. However, the power of this UX was amplified by the underlying vector search capabilities. Data stewards could not only search for records but also understand why the AI suggested a potential match, providing a level of transparency and confidence previously unattainable.

To empower data stewards further, we built an interactive, Elasticsearch-enabled user experience. This provided intuitive tools for reviewing potential duplicates, resolving conflicts, and managing the creation and maintenance of golden records. The scalability of Elasticsearch ensured that the solution could handle massive datasets without performance degradation.

We developed a suite of algorithms and web services tailored for key business processes:

Contextual Duplicate Grouping Algorithm: Leveraging vector-based grouping, this algorithm intelligently grouped potential duplicates based on contextual similarities captured in the vectors, moving beyond simple attribute matching to nuanced semantic understanding.

Delta Detection and Distribution Service: This service efficiently identified changes in customer data and disseminated these updates to relevant downstream systems, ensuring data consistency across the marketing ecosystem.

Telephone and Email Validation Algorithms: We built robust validation services for telephone numbers and email addresses, significantly improving data hygiene and deliverability rates for marketing communications.

Geocoding APIs: Integrated geocoding capabilities to standardize and enrich address data, enabling more precise segmentation and location-based marketing.

The implementation of this MDM-powered predictive analytics solution transformed the core marketing database from a collection of fragmented and unreliable data into a single, authoritative source of truth. This enabled highly targeted campaigns, improved customer segmentation, reduced marketing spend on duplicate outreach, and ultimately drove a substantial increase in qualified leads and customer acquisition.

3.3 The Marketability Leap: From Technical Curiosity to Business Imperative

Elevated Precision and Risk Mitigation: The integration of vector search fundamentally changed the solution's marketability. By minimizing the risk of erroneous deletions (a critical concern when dealing with customer data and impacting CLV), we addressed a major pain point that contemporary solutions ignored. The ability to provide contextual relevance meant that marketing teams could trust the "golden records" generated, leading to more effective campaigns and a direct impact on revenue.

Enhanced Design Iterations with Reinforcement Learning from Human Feedback (RLHF) Potential: The interactive UX, coupled with the underlying AI models, created a feedback loop that was ripe for RLHF in future iterations. Data stewards' decisions on resolving duplicates could be used to further refine the vector embeddings and matching algorithms, continuously improving accuracy and contextual relevance. This adaptability and potential for ongoing learning made the solution forward-looking and inherently more marketable.

Tangible Business Transformation: The implementation of this AI-driven customer intelligence solution transformed the core marketing database. It enabled highly targeted campaigns, improved customer segmentation, reduced marketing spend on duplicate outreach, and demonstrably increased qualified leads and customer acquisition, all while meticulously safeguarding individual customer identities.

This solution transitioned from a technical initiative to a critical business asset. Its ability to directly address the pain points of data fragmentation, reduce operational costs, and demonstrably improve marketing performance made it highly marketable within the organization and a testament to the power of a data-centric AI strategy. We were able to completely eliminate the cost of leveraging and maintaining a full-fledged master data management solution that we were using previously and also reduced the manual validation effort by 90%, and the only place we needed the evaluations were specific callouts because of insufficient information to back the AI-powered decision. But even this effort was leveraged for Reinforcement Learning based on the Human Feedback approach.

Key Takeaways:

TABLE 3.1 A comparison table

Feature	*Traditional Approach (Standard MDM)*	*Our AI-Driven Approach (Vector Search)*
Matching Logic	Rule-based, fuzzy logic, struggles with context	Semantic understanding via vector search
Accuracy	High false positives/negatives, risks deleting unique users	High precision, distinguishes context to protect "Golden Records"
Maintenance	High operational overhead, complex platform management	Minimalist, scalable engine, lower maintenance burden
User Experience	Often rigid and technical	Elasticsearch enabled, transparent and interactive for stewards

Source: Developed by the author based on applied implementation of vector search and AI-driven Master Data Management principles informed by contemporary research on semantic retrieval, reinforcement learning from human feedback (RLHF), and customer data governance.

- Prioritize Context: Semantic understanding (vector search) outperforms simple string matching for complex data identity problems.
- Risk First: Design your algorithm to solve the most expensive business problem first (in this case, erroneous deletion of customers).
- Human-in-the-Loop: Use data steward feedback not just for corrections, but to contextually empower the model (RLHF) for future precision.

In Practice: Lessons from the Field:

- Prioritize Context over Syntax: For complex data matching, use contextual and advanced hybrid vector search (semantic understanding) rather than rigid rules or fuzzy logic.
- Design for the "Human-in-the-Loop": Build interfaces that empower users (like data stewards) to review and correct AI decisions, increasing trust and accuracy.
- Solve the Most Expensive Problem: We focused on avoiding "erroneous deletions" because that was the costliest business error, making the ROI immediate and clear.

4 The Wider Landscape of Marketable AI: Beyond Data and Algorithms

Creating a commercially successful AI solution involves a multifaceted approach that extends beyond technical prowess. This section explores additional critical factors that contribute to marketability, fostering adoption, and ensuring long-term value.

4.1 Design Thinking and Product Mindset

User-Centric Product Development: As discussed, viewing an AI-based solution as a product, not just a piece of technology, is crucial. Applying design thinking principles ensures that the solution is intuitive, solves a real user problem, and integrates seamlessly into existing workflows.

Focus on User Experience (UX): A clunky or confusing interface can doom even the most sophisticated AI. Prioritizing a clear, efficient, and engaging user experience is paramount for adoption and user satisfaction.

4.2 Personalization and Customer Lifetime Value (CLV) Expectation

Delivering Tailored Experiences: AI's power in personalization can significantly enhance customer engagement and loyalty. By understanding individual preferences and behaviors, AI can deliver tailored recommendations, content, and interactions, directly impacting CLV.

Long-Term Value Creation: Marketable AI solutions are those that contribute to sustained customer relationships and increased lifetime value, rather than focusing solely on short-term gains.

4.3 Scalability and Predictability of Results

Future-Proofing with Scalability: Businesses demand AI solutions that can grow with them. Ensuring the underlying infrastructure and algorithms are scalable to handle increasing data volumes and user loads is a key marketability factor.

Reliable Outcomes: Predictability in AI results builds trust. Businesses need to be confident that the AI will perform consistently and reliably, enabling informed decision-making and operational planning.

4.4 Continuous Integration, Continuous Delivery (CI/CD) and Iterative Improvement

Agile Development Practices: Implementing CI/CD pipelines allows for rapid iteration and deployment of AI models and features. This agility is crucial for responding to market changes and user feedback.

Reinforcement Learning from Human Feedback (RLHF): The ability to incorporate user feedback and learning into AI models, particularly through RLHF, allows for continuous improvement and adaptation. This ensures the AI remains relevant and effective over time, a hallmark of a marketable solution.

In Practice: Product Mindset:

- Shift from Project to Product: Treat your AI solution as an evolving product. Plan for continuous updates, not a "fire and forget" deployment.
- Implement Feedback Loops: Actively capture user interactions to retrain and refine your models (Reinforcement Learning from Human Feedback).
- Stress-Test Scalability: Ensure your infrastructure can handle peak loads. An AI that times out during high traffic loses user trust instantly.

5 The Ethical Compass: Building Trust and Ensuring Responsible AI

The increasing sophistication of AI necessitates a parallel evolution in our understanding and implementation of ethical considerations, echoing the European Commission's (2019) call for "Trustworthy AI" that is lawful, ethical, and robust. This section explores the critical importance of bias mitigation, transparency, and explainability in fostering trust and acceptance of AI solutions.

5.1 The Unseen Biases: Navigating the AI Minefield

Without proactive mitigation, societal biases in any data can be amplified, leading to discriminatory outcomes. Identifying and addressing bias in data is paramount to developing fair and equitable AI systems. This requires diverse data sources, rigorous auditing, and specialized techniques for bias detection and correction.

Recall the recruitment tool that taught itself to penalize resumes containing the word "women's" (e.g., "women's chess club") simply because the historical hiring data it was trained on was dominated by men. This illustrates why *you* cannot assume data is neutral—it often carries the ghosts of past decisions.

The Illusion of Objectivity: AI solutions are often perceived as inherently objective. However, algorithmic design choices, feature selection, and the very data used can embed subtle, yet impactful, biases. Understanding these potential pitfalls is the first step towards building trustworthy AI solutions.

5.2 Practical Guidance for Ethical AI Frameworks

Transparency and Explainability: The "black box" nature of some AI solutions erodes trust. Developing methods for explaining AI decisions, particularly in sensitive domains like finance or healthcare, is crucial. This involves employing explainable AI (XAI) techniques that allow users and stakeholders to understand why an AI arrived at a particular conclusion.

Consider a Real-World Example Scenario:

Scenario: When a leading FinTech company introduced an AI tool for loan approvals, adoption among branch managers was initially near zero. They didn't trust a "black box" to reject long-standing clients.

The Fix: The team implemented an Explainable AI (XAI) dashboard. Instead of just saying "Denied," the AI generated a plain-English rationale: "Denied because Debt-to-Income ratio exceeds 45%."

The Result: Branch managers could now explain decisions to customers, turning the AI from a mysterious threat into a transparent tool. Adoption jumped 40% in one quarter, proving that *transparency is a marketability feature.*

Bias Mitigation Strategies: This includes techniques such as data augmentation, algorithmic fairness constraints, and post-hoc bias correction. The goal is to ensure that AI solutions do not perpetuate or exacerbate existing societal inequalities.

Building Trust through Accountability: Establishing clear lines of accountability for AI development and deployment is essential. This includes processes for ethical review, ongoing monitoring, and mechanisms for addressing unintended consequences.

Pro Tip: Beware of Proxy Variables: When checking for bias, simply removing sensitive fields (like gender or zip code) isn't enough. AI is excellent at finding patterns and may use "proxy variables" (like shopping habits or email domains) to infer the protected traits you removed.

Action: Always test your model for *disparate impact* on outcomes, not just "blindness" in inputs.

In Practice: Ethical Stewardship:

- **Test for Outcomes, Not Just Inputs:** Don't just remove sensitive fields, test your model's *outputs* to ensure it isn't inferring bias from proxy variables (like zip codes).
- **Open the "Black Box":** Implement Explainable AI (XAI) dashboards so stakeholders understand *why* the AI made a specific recommendation.
- **Formalize Oversight:** Create a "Responsible AI" committee or checklist to review every major model update for unintended societal impacts.

6 The Practical Toolkit for Marketable AI

Bridging the gap between AI innovation and market success requires practical, actionable tools and frameworks. This section provides a direction for leveraging these resources.

6.1 Devising a Value Proposition Template

This section presents a structured template that guides users through identifying the business problem, quantifying the opportunity with OKRs, defining the target user, outlining the AI solution's unique value, and articulating the expected ROI and CLV impact.

AI solution's value proposition template with samples:

TABLE 3.2 Value proposition template

Section	*Component*	*Guidance*	*Example(s)*
I. The business problem	Core pain point: What critical challenge are we addressing?	Clearly articulate the specific business problem, inefficiency, or unmet need that the AI solution is designed to solve. Be precise and focus on the business impact.	Our marketing team struggles with fragmented customer data, leading to inconsistent outreach, wasted marketing spend, and missed opportunities for personalization, directly impacting customer acquisition costs and retention rates.
	Impact (Current state): Quantify the problem's magnitude.	Use data and metrics to illustrate the magnitude of the problem. This sets the baseline for demonstrating improvement.	Currently, an estimated 15% of our customer database consists of duplicate or inaccurate records, leading to an annual overspend of $X on redundant marketing campaigns and a Y% lower conversion rate on targeted outreach due to poor segmentation.

Section	*Component*	*Guidance*	*Example(s)*
II. The AI solution	Solution overview: How will AI solve this problem?	Provide a concise overview of the AI solution and its core functionality, focusing on *what* it does to address the identified problem. Avoid overly technical jargon unless essential.	Our AI-powered Customer Intelligence Platform utilizes advanced NLP, NER, and vector search integrated with Elasticsearch to create a unified, accurate, and contextually relevant "golden record" for each customer.
	Unique AI-driven value: What makes this AI solution unique and superior?	Explain what makes this AI solution unique and superior to existing or conventional methods. Focus on the "how" and the specific AI capabilities.	Unlike traditional MDM tools relying on basic matching, our solution's vector search capability, powered by NLP/NER, understands the semantic meaning of customer data variations, ensuring unparalleled accuracy in deduplication and minimizing the critical risk of deleting unique individuals. This precision allows for truly effective personalization and segmentation.
III. The target user/audience	Primary stakeholders: Who benefits directly?	Identify the specific individuals, teams, or departments that will directly use or benefit from the AI solution.	Marketing Managers, Data Analysts, Customer Relationship Managers, and the Head of Customer Acquisition.

Section	*Component*	*Guidance*	*Example(s)*
	User needs & goals: What do they need to achieve?	Briefly describe what this target audience needs to achieve and how the AI solution directly supports their objectives.	They need accurate customer data to execute targeted campaigns, improve customer engagement, and reduce acquisition costs. Our solution empowers them with a reliable, singular view of the customer, enabling data-driven decisions and personalized interactions.
IV. OKRs & Expected outcomes	Objective: overarching business goal.	State the overarching business goal the AI solution aims to achieve.	To significantly improve customer data accuracy and enhance marketing campaign effectiveness.
	Key Results (KR1–KR4): measurable outcomes.	Define specific, measurable, achievable, relevant, and time-bound (SMART) metrics that will demonstrate the success of the AI solution.	*KR1: Reduce duplicate customer records by 95% within 6 months.* *KR2: Increase marketing campaign conversion rates by 20% in Q3 by enabling hyper-personalized outreach.* *KR3: Decrease marketing spend on redundant campaigns by 10% annually.* *KR4: Achieve a 98% accuracy rate in identifying true customer duplicates, with zero critical data deletions due to misidentification.*

Section	*Component*	*Guidance*	*Example(s)*
V. Expected ROI & CLV impact	Quantifiable ROI: Financial benefits and cost savings.	Project the financial benefits and cost savings directly attributable to the AI solution.	Based on reduced marketing waste and improved conversion rates, we project an annual ROI of Y%, with a payback period of Z months. The cost savings from decommissioning the legacy MDM platform will also contribute to this ROI.
	Projected CLV impact: long-term customer value enhancement.	Articulate how the AI solution will enhance long-term customer value.	By enabling superior personalization and a more consistent customer experience, we anticipate an increase in average Customer Lifetime Value by 15% over the next two years due to improved retention and increased purchase frequency.

A condensed table for the above value proposition template:

TABLE 3.3 (Condensed) value proposition template

Component	*Key Question to Answer*	*Sample Entry (Marketing AI Case)*
1. The Business Problem	What critical pain point are we solving, and what does it cost us today?	*Fragmented customer data leads to $X wasted annually on redundant marketing*
2. The AI Solution	How does AI uniquely solve this better than standard software?	*Vector search (NLP) identifies semantic duplicates that standard rule-based matching misses*
3. Target User	Who will use this, and what is their goal?	*Marketing operations team needing a "Golden Record"*
4. OKRs (Metrics)	What are the specific, measurable results?	*Reduce duplicate records by 95% in under 6 months*
5. ROI & CLV	What is the financial return and long-term customer impact?	*Increase campaign conversion by 20%; reduce manual cleaning costs by 90%.*
6. The "Secret Sauce"	What makes this defensible/unique?	*Contextual understanding of customer identity, not just string matching.*

Source: Author's own elaboration; adapted from a value-proposition template used to assess AI solution fit and marketability

This template could be a starting point for assessing the solution's value and present proposition accordingly. By meticulously completing and evolving each section of this template objectively, we can construct a powerful and persuasive value proposition that clearly communicates the marketability and strategic importance of the AI solution.

6.2 Conducting a Data Readiness Assessment

For any AI solution to transition from a promising concept to a commercially viable product, its data foundation must be robust, reliable, and readily accessible. This section provides a practical framework for assessing an organization's data readiness, ensuring that the underlying data infrastructure can support the development, deployment, and sustained performance of marketable AI. This assessment helps identify gaps and prioritize actions, moving beyond mere data availability to data utility for AI. This checklist or framework should help evaluate an organization's data quality, accessibility, governance, scalability, and infrastructure readiness for AI initiatives. It includes questions related to data sources, data profiling, master data management maturity, and data security.

Objective: To systematically evaluate the organization's data landscape and identify critical areas for improvement to ensure the successful development and deployment of marketable AI solutions.

Instructions: For each section, look critically at your organization's current state. Ask yourself: If we tried to train a model on this data today, would it work? Use a scoring system (e.g., 1–5, where 1 = Severely Lacking, 5 = Excellent) or descriptive qualitative notes to capture the assessment. The "Why It Matters for Marketable AI" column highlights the direct link between data readiness and commercial viability.

TABLE 3.4 Data readiness assessment template

Assessment area	*Key considerations*	*Current state assessment (Score/ Notes)*	*Why it matters for marketable AI*	*Action items*
1. Data quality & integrity	How are data errors (typos, inconsistencies, missing values) identified and rectified? Are there established data profiling and validation processes? Is data consistency enforced across disparate systems? How are data anomalies detected and handled?		Foundation of trust: inaccurate or inconsistent data leads to flawed AI models, unreliable predictions, and ultimately a loss of market confidence. High quality is non-negotiable for actionable insights.	Implement automated data profiling tools. Define clear data quality rules and implement monitoring dashboards. Establish data cleansing workflows.
2. Data accessibility & availability	How easily can relevant data be accessed by AI development teams and models? Are there data silos preventing a holistic view? What are the processes for data extraction and integration? Is data available in formats compatible with AI tools?		AI models require timely access to relevant data. Inaccessible or siloed data cripples development speed, limits model sophistication, and delays time-to-market, directly impacting commercialization potential.	Develop data catalogs and dictionaries. Implement data virtualization or robust ETL/ELT pipelines. Establish clear access protocols and data provisioning processes.

Assessment area	*Key considerations*	*Current state assessment (Score/ Notes)*	*Why it matters for marketable AI*	*Action items*
3. Data governance & stewardship	Who owns the data? Are there clear policies on data usage, privacy, and security? Is there a defined process for data lifecycle management (creation to archival/deletion)? Are data stewards empowered and trained?		Robust governance builds trust with users and regulators. It ensures data is used ethically and securely, preventing legal pitfalls and reputational damage, which are critical for long-term marketability. Proper stewardship is key to maintaining data integrity, as discussed in the context of MDM.	Formalize data governance policies and procedures. Appoint and train data stewards for critical data domains. Implement data lineage tracking. Ensure compliance with relevant data protection regulations (e.g., GDPR, CCPA).
4. Data scalability & performance	Can the data infrastructure handle increasing volumes of data? Does data retrieval and processing meet the performance requirements for real-time or near-real-time AI? Is the data architecture designed for future growth and evolving AI needs?		Marketable AI solutions must perform reliably under load. Slow data processing or inability to scale will lead to poor user experience and system failures, undermining adoption and marketability.	Evaluate and upgrade data storage and processing infrastructure. Optimize data retrieval queries and indexing strategies. Design for microservices and distributed data architectures.

Assessment area	*Key considerations*	*Current state assessment (Score/ Notes)*	*Why it matters for marketable AI*	*Action items*
5. Data context & meaning (AI-Specific)	Is data enriched with metadata that explains its origin, meaning, and intended use? Are there mechanisms to capture the "context" of data points (e.g., through NLP, vector embeddings)? Is data structured to support advanced AI techniques?		AI thrives on context and meaning, not just raw data. Data that is semantically rich and structured for AI enables more sophisticated models (like vector search), leading to richer insights and higher market value.	Invest in metadata management solutions. Explore contextual vector embedding techniques for data enrichment. Ensure data schemas are flexible and can accommodate AI-driven feature engineering. Document data sources and their inherent limitations.
6. Data Security & Privacy	Are sensitive data elements identified and protected? Are access controls granular and enforced? What measures are in place for data encryption (at rest and in transit)? Are there regular security audits and vulnerability assessments?		In an era of heightened data privacy concerns, robust security and privacy measures are paramount. Failures here can lead to catastrophic loss of trust, regulatory penalties, and immediate market rejection of the AI solution.	Implement data anonymization and pseudonymization techniques where appropriate. Conduct regular security vulnerability assessments and penetration testing. Train personnel on data security best practices. Develop incident response plans.

Source: Author's own elaboration based on the data readiness assessment framework presented in the chapter.

To use this template:

- Approach each question with an honest assessment of the current state, supported by evidence where possible.
- Gather input from data engineers, data scientists, business analysts, and compliance officers to gain a comprehensive view.
- The assessment should not be a one-time exercise but a catalyst for continuous improvement, driving targeted actions to enhance data utility for AI.
- Ensure that data readiness efforts directly support the specific AI initiatives and business goals planned.
- Synthesize the assessment findings into a high-level summary of the organization's data readiness for marketable AI. Highlight the most critical areas for immediate attention.

- List 3–5 prioritized actions based on the assessment to improve data readiness and enhance the likelihood of creating successful, marketable AI solutions.

By diligently working through this assessment, organizations can build a solid data foundation that underpins the creation of AI solutions that are not only technically sound but also commercially resilient and trustworthy.

6.3 Navigating Ethical AI Checklists

As AI solutions become increasingly integrated into decision-making processes and daily life, the imperative for responsible development and deployment grows. Ethical considerations are no longer an afterthought but a fundamental component of creating AI that is not only effective but also trustworthy and widely accepted. This checklist provides a structured guide for developers and business leaders to critically evaluate the ethical dimensions of their AI solutions, fostering accountability and mitigating potential harm.

Objective: To systematically assess an AI solution's ethical implications across key dimensions, ensuring responsible innovation and building stakeholder trust.

Instructions: For each question, critically evaluate your AI solution. Mark "Yes" if the principle is robustly addressed, "Partially" if addressed but with room for improvement, or "No" if not adequately addressed. Provide brief notes explaining your rating and outline specific actions for improvement in the "Action Plan" column.

TABLE 3.5 Ethical AI checklist

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
1. Data bias & fairness	1a. Have potential sources of bias in data been rigorously identified and documented?		*Why it matters:* Data reflects societal biases. Unchecked, AI can perpetuate or even amplify discrimination. Acknowledging bias is the first step to mitigation.	Conduct comprehensive solution audits for demographic, historical, and representational biases. Document all identified biases and their potential impact. Establish clear metrics for measuring fairness across different subgroups.

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
	1b. Are specific bias mitigation techniques (e.g., re-sampling, re-weighting, algorithmic fairness constraints) being applied to the solution?		*Why it matters:* Proactive steps are needed to counteract identified biases and ensure equitable outcomes for all user groups.	Implement selected bias mitigation strategies during data pre-processing for the solution. Continuously monitor model performance for fairness disparities across identified subgroups.
	1c. Are performance metrics evaluated for fairness across different demographic groups or protected characteristics?		*Why it matters:* Standard performance metrics can mask disparate impact. Fairness must be explicitly measured to ensure the AI does not disproportionately harm certain groups.	Define and track fairness metrics (e.g., equal opportunity, demographic parity) relevant to the AI's application. Establish thresholds for acceptable fairness deviations.
2. Transparency & explainability	2a. Is the AI model's decision-making process understandable, at least to a degree relevant to its impact?		*Why it matters:* "Black box" AI erodes trust. Stakeholders need to understand how decisions are made, especially in high-stakes scenarios, to foster confidence and enable audits.	Explore and implement Explainable AI (XAI) techniques (e.g., SHAP, LIME, feature importance) suitable for the solution and its context. Document the solution architecture and its rationale.

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
	2b. Is it clear *what* data was used in the solution and *how* it was processed?		*Why it matters:* Understanding the data provenance and processing steps is crucial for identifying potential biases and for debugging or improving the solution.	Maintain detailed logs of data sources, pre-processing steps, and feature engineering. Create clear documentation accessible to relevant technical and business teams.
	2c. Are the limitations and potential failure modes of the AI solution clearly communicated to users and stakeholders?		*Why it matters:* Setting realistic expectations is vital. Users must understand when the AI might err or what its boundaries are to use it responsibly and avoid over-reliance.	Develop clear user manuals and disclaimers. Conduct training sessions that highlight AI limitations. Implement user feedback mechanisms to capture observed failure modes.
3. Privacy & data protection	3a. Has a thorough privacy impact assessment (PIA) been conducted?		*Why it matters:* Ensuring compliance with privacy regulations (e.g., GDPR, CCPA) is legally mandated and essential for user trust.	Conduct the solution's formal PIA, identifying all personal data processed and associated risks. Implement data minimization principles (collect only what is necessary). Ensure data anonymization or pseudonymization techniques are appropriately applied.

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
	3b. Are robust security measures in place to protect sensitive data used or generated by the AI?		*Why it matters:* Data breaches involving AI systems can have severe consequences, including identity theft, financial loss, and severe reputational damage.	Implement encryption for data at rest and in transit. Enforce strict access controls and authentication protocols. Conduct regular security audits and penetration testing.
	3c. Is data retention and deletion managed in accordance with privacy policies and user consent?		*Why it matters:* Holding onto data longer than necessary increases risk and can violate privacy principles. Clear policies ensure data is only kept for legitimate purposes.	Define clear data retention schedules. Implement automated processes for secure data deletion or anonymization. Ensure mechanisms for honoring user requests for data access or deletion are in place.
4. Accountability & Oversight	4a. Are clear lines of responsibility defined for the AI system's development, deployment, and ongoing performance?		*Why it matters:* Without clear accountability, it's difficult to address issues when they arise or to assign responsibility for unintended consequences.	Establish an AI ethics committee or responsible AI working group. Assign specific roles and responsibilities for AI oversight. Document decision-making processes related to AI deployment and incident response.

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
	4b. Is there a mechanism for reporting and addressing ethical concerns or unintended AI behaviors?		*Why it matters:* Issues can emerge post-deployment. A clear process for reporting and resolving ethical concerns ensures continuous improvement and builds trust.	Create a confidential reporting channel for ethical concerns. Establish a process for investigating, validating, and acting upon reported issues. Communicate the resolution process to stakeholders.
	4c. Is there a process for regular review and auditing of the AI solution's ethical compliance and performance?		*Why it matters:* The AI landscape and ethical standards evolve. Regular audits ensure the solution remains compliant and ethically sound over its lifecycle.	Schedule periodic ethical audits and performance reviews. Update the checklist and evaluation criteria as new ethical considerations arise. Maintain records of audits and corrective actions.
5. Societal & environmental impact	5a. Has the potential broader societal impact of the AI solution been considered (e.g., job displacement, accessibility, equity)?		*Why it matters:* AI can have far-reaching consequences beyond direct users. Understanding and mitigating negative societal impacts is crucial for responsible innovation.	Conduct a "societal impact assessment" for the AI solution. Engage with diverse stakeholders to understand potential broader impacts. Develop strategies to mitigate negative societal consequences.

Ethical dimension	*Question*	*Rating (Yes/Partially/No)*	*Notes/Justification*	*Action Plan*
	5b. Are the environmental footprint and resource consumption of the AI (especially training and large-scale deployment) being considered?		*Why it matters:* The energy demands of AI can be significant. Considering environmental sustainability is an emerging but important aspect of responsible technology development.	Explore energy-efficient AI model architectures and hardware. Document and, where possible, quantify the environmental impact. Investigate opportunities for optimizing AI's carbon footprint.

Source: Own elaboration based on ethical AI frameworks and sustainability guidance (OECD, 2022; Mitigating Bias in AI, 2025).

TABLE 3.6 (Condensed) ethical AI checklist

Category	*Key Question*	*Action Item*
1. Fairness & Bias	**Has the data been tested for "invisible" bias?** *(Does the model treat all demographics equally, even if sensitive fields are removed?)*	• Conduct a Disparate Impact Analysis on model outputs • Document known data limitations (e.g., "Dataset excludes rural regions")
2. Transparency & Accountability	**Can we explain *why* the AI made a decision** *(Is there a human-in-the-loop or XAI dashboard for high-stakes decisions?)*	• Implement **Explainable AI (XAI)** tools (e.g., SHAP values) to show top factors driving a prediction • Assign a specific "Model Owner" responsible for decision auditing
3. Privacy & Sustainability	**Are we minimizing harm?** *(Is data usage minimized, and is the energy cost justified?)*	• Enforce **Data Minimization:** Collect only strictly necessary fields • Measure carbon footprint of training runs and opt for pre-trained models where possible

Source: Author's own elaboration; adapted from ethical AI and responsible AI guidance in the literature.

The condensed version of the above checklist table:
To use this checklist:

- Involve a diverse team (developers, product managers, legal/compliance, ethics officers) to ensure all facets of the AI are considered.

- Adapt the "Why it Matters" and "Action Plan" sections to the specific context, risk profile, and impact of your AI solution.
- Ethical evaluation is ongoing. Revisit this checklist regularly, especially after significant model updates or when the AI's application context changes.
- The goal is not just to identify issues but to create concrete plans for addressing them, thereby moving towards truly responsible AI.
- Synthesize the assessment findings into a high-level summary of the AI solution's ethical maturity. Highlight key strengths and areas requiring immediate attention.

This checklist serves as a living document to guide our continuous efforts in responsible AI development. By diligently using this checklist, organizations can proactively address ethical challenges, build more trustworthy AI solutions, and enhance their long-term marketability and societal acceptance.

Architecting AI for Enduring Value: The commercialization of AI solutions is not a solitary pursuit of technology and its sophistication; it is a strategic endeavor that demands a holistic perspective. By prioritizing a crystal-clear value proposition, ensuring robust data readiness, embedding ethical considerations, and adopting a product-centric mindset, organizations can move beyond technically impressive but commercially inert AI projects.

This chapter has laid out a framework, substantiated by practical application and advanced AI methodologies, for architecting AI solutions that are not only innovative and efficient but also deeply valuable, trustworthy, and ultimately, destined for market success. The path to marketable AI is paved with a clear understanding of business needs, a robust data foundation, a commitment to continuous improvement, and a steadfast dedication to responsible innovation.

Note on AI Usage

While drafting this chapter, I have leveraged AI as a tool to evaluate spelling, grammar, sentence formation, cosmetic title adjustments for reader engagement, and plagiarism check, but the intellectual contribution and authorship are entirely mine.

References

Brown, T. (2008). Design thinking. *Harvard Business Review*, 86(6), 84–92.

Doerr, J. (2018). *Measure what matters: How Google, Bono, and the Gates Foundation rock the world with OKRs*. Portfolio/Penguin.

European Commission. (2019). *Ethics guidelines for trustworthy AI*. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/346720.

Forrester Research. (2022). The Future of Data Management. https://www.forrester.com/report/the-future-of-data-management/RES177048.

Gartner. (2018). Gartner says nearly half of CIOs are planning to deploy artificial intelligence [Press release]. https://www.gartner.com/en/newsroom.

IBM. (2016). The cost of bad data. IBM Big Data & Analytics Hub. https://community.sap.com/t5/technology-blog-posts-by-sap/bad-data-costs-the-u-s-3-trillion-per-year/ba-p/13575387#:~:text=What%20changes%20if%20the%20real,is%20costly%2C%20this%20figure%20stuns.

ISO. (2011). ISO 8000–2:2011 Data quality — Part 2: Vocabulary. International Organization for Standardization. https://www.iso.org/standard/50798.html.

Redman, T. C. (2016). Bad data costs the U.S. $3 trillion per year. https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year.

Reuters. (2018) https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/.

Smith, J. (2024). Innovative impacts of digital marketing on business strategy: A systematic review. *International Journal of Research Publication and Reviews*, 5(5), 2296–2300. https://www.researchgate.net/publication/380398321_Innovative_Impacts_of_Digital_Marketing_on_Business_Strategy_A_Systematic_Review.

4

AI AS DRIVER OF SUSTAINABLE COMPETITIVE ADVANTAGE

Miguel Stief

Introduction

Artificial intelligence (AI) has evolved from a promising experimental technology into a structural pillar of modern enterprise competitiveness. Across industries, AI now shapes core operations, product design, customer interaction, and governance. Yet this rapid diffusion exposes a fundamental paradox: while AI enables efficiency and innovation at unprecedented scale, it simultaneously amplifies risks related to inequity, environmental impact, and ethical misuse. This tension—between value creation and systemic risk—defines the central challenge of sustainable digital transformation.

Executives increasingly recognize AI as indispensable. According to McKinsey & Company (2024), organizations embedding AI into core workflows report up to 20 percent EBIT uplift compared to peers. At the same time, governance has moved decisively onto board agendas: KPMG (2025) reports that 82 percent of global CEOs rank AI governance among their top three strategic priorities. These pressures are further intensified by hardware evolution. Grand View Research (2024) projects the AI-PC market to grow from US$51 billion in 2024 to US$281 billion by 2030, underscoring how computational infrastructure itself has become a structural source of competitive differentiation.

However, technological diffusion alone does not produce durable advantage. Many organizations encounter what this chapter defines as the *AI wall*: a point at which pilot initiatives demonstrate technical promise but fail to scale because data architectures, ethical frameworks, and human capabilities remain misaligned. Importantly, this barrier is not technological. It is organizational and cultural. Sustainable advantage emerges only when AI is embedded within the enterprise's operating system—its governance structures, leadership practices, and underlying values.

Empirical evidence reinforces this diagnosis. The *State of AI 2024* survey (McKinsey & Company, 2024) shows that while 72 percent of companies have launched AI initiatives, fewer than 15 percent have scaled them profitably. Fragmented data, skills shortages, and unclear accountability repeatedly undermine transformation efforts.

DOI: 10.4324/9781003632429-6
This chapter has been made available under a CC-BY-NC-ND 4.0 license

These findings echo Sachs' (2015) broader argument that long-term prosperity depends on aligning technological progress with ethical and institutional systems. In the context of AI, the challenge is therefore not adoption, but assimilation.

Against this backdrop, this chapter examines how AI can become a driver of *sustainable competitive advantage* when deployed as a systemic capability across four mutually reinforcing dimensions:

1. **Operational efficiency**, where AI enhances productivity through automation, analytics, and adaptive infrastructure;
2. **Customer-centric innovation**, where intelligent systems enable new value models, personalization, and ecosystem orchestration;
3. **ESG integration**, ensuring that ethical, social, and environmental considerations shape AI design and deployment; and
4. **Workforce transformation**, aligning human skills, inclusion, and empowerment with intelligent systems.

Drawing on empirical studies, executive practice, and the DCC4B (digital, circular, collaborative for business) model, the chapter positions AI not as a standalone technology but as an enterprise-wide operating capability. The sections that follow establish the theoretical foundations for this multidimensional framework and explore how organizations can convert AI from a transient technological edge into a durable source of resilience, legitimacy, and growth.

Figure 4.1 illustrates the central paradox underlying AI-driven transformation. As artificial intelligence diffuses and scales across the enterprise, two effects increase simultaneously: business value and systemic risk. On one hand, AI adoption enhances productivity, efficiency, and innovation, generating measurable economic returns. On the other, the same scaling process amplifies ethical exposure, social inequality, environmental impact, and governance challenges. The "AI Wall" represents the organizational inflection point at which technical success confronts institutional limitations—governance misalignment, skills gaps, and cultural resistance. Sustainable competitive advantage emerges not from overcoming this wall technologically, but from assimilating AI into the enterprise's governance, capabilities, and values.

Theoretical Foundations

The intellectual roots of AI-enabled competitiveness trace back to long-standing debates on technology as a strategic resource. Davenport and Ronanki (2018) conceptualize artificial intelligence as a pragmatic enabler of incremental improvement—augmenting human decision-making, reducing friction, and streamlining operations. In contrast, Iansiti and Lakhani (2020) frame AI as a general-purpose technology capable of transforming industries through algorithmic scaling, network effects, and systemic reconfiguration. Together, these perspectives delineate two distinct pathways to competitive advantage: an evolutionary path centered on efficiency gains and optimization, and a transformative path driven by structural change.

Recent organizational research further reframes the firm not as a static, control-oriented machine, but as an adaptive system designed for continuous sensing, learning,

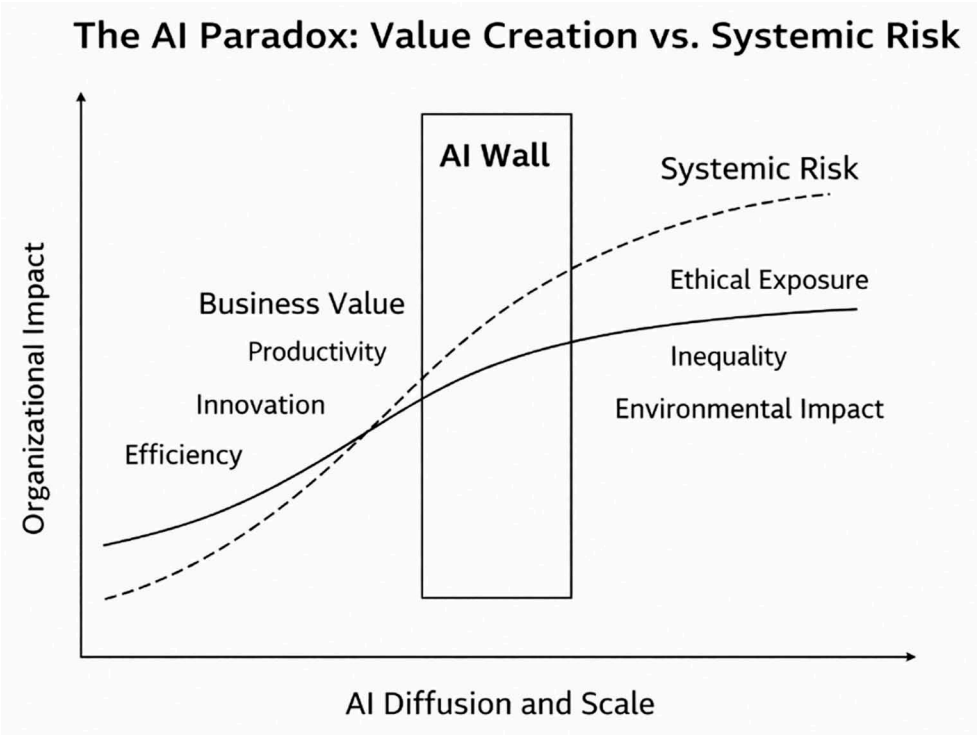


FIGURE 4.1 The AI paradox: value creation vs. systemic risk
Source: Own elaboration based on the chapter's synthesis of AI value creation, systemic risk, and organizational resilience, supported by recent research on AI's bright and dark sides.

and reconfiguration under uncertainty. The Octopus Organization perspective (Le-Brun & Werner, 2025) captures this shift by emphasizing distributed intelligence, localized decision authority, and rapid feedback loops as critical sources of advantage in complex environments. Within this logic, AI does not primarily function as a tool for automating predefined processes; rather, it operates as a cognitive amplifier that enhances the organization's capacity to detect weak signals, dynamically reallocate resources, and coordinate coherent responses across multiple semi-autonomous operational units. This architectural view helps explain why firms with similar access to AI technologies experience divergent outcomes: sustained advantage emerges not from technological adoption per se, but from the alignment between AI capabilities and an organization's adaptive structure.

Classical resource-based theory defines competitive advantage through assets that are valuable, rare, inimitable, and non-substitutable. AI disrupts this model because algorithms, tools, and even data practices diffuse rapidly across organizational and national boundaries. Hoffman (2025) argues that sustainable advantage therefore depends less on proprietary technology than on organizational discipline and ecosystem orchestration. In this view, governance—how data, ethics, experimentation, and collaboration are structured—replaces possession as the primary source of defensibility.

This shift is increasingly supported by empirical evidence. Global surveys on AI adoption consistently show that value capture correlates more strongly with management practices and operating models than with specific model choices or levels of technical sophistication. Organizations that institutionalize repeatable routines for experimentation, deployment, and learning—spanning strategy, talent, data governance, and adoption—are significantly more likely to translate AI investments into durable returns. Competitive advantage thus migrates from isolated technological assets to the organizational capacity to absorb learning and scale responsibly.

Sachs (2015) extends this logic through a sustainability lens, arguing that technological progress detached from social and environmental responsibility is inherently unstable. Applied to AI, this principle implies that governance must evolve alongside innovation. Chakravorti (2024) warns that unchecked model expansion can intensify inequality, bias, and energy consumption, while Sætra (2023) demonstrates that incorporating sustainability metrics directly into AI design improves both system performance and institutional legitimacy.

Governance, however, is not solely an internal design choice; it is also an external competitive constraint. Ongoing debates surrounding AI regulation—particularly in jurisdictions adopting risk-based frameworks—illustrate how compliance complexity, transparency requirements, and accountability mechanisms increasingly shape strategic outcomes. For enterprises, the implication is pragmatic rather than ideological: governance must function as an enabling architecture for scaling AI, not as an after-the-fact control layer that fragments responsibility or slows innovation.

Empirical studies reinforce these insights. KPMG (2025) and LXT (2025) identify recurring stages of AI maturity—experimentation, integration, and institutionalization—each requiring progressively tighter coordination between technical and ethical subsystems. McKinsey (2024) shows that companies aligning reskilling and governance with AI deployment capture approximately 1.7× higher ROI than those focused solely on technology. Complementary findings indicate that leadership teams often overestimate near-term AI returns, while internal experts remain more cautious due to accountability, operational risk, and capability constraints. This asymmetry helps explain why workforce development and governance investment operate as compounding strategic assets over time rather than discretionary costs.

This theoretical synthesis underpins the framework developed in later sections. AI's strategic value arises not from algorithms themselves, but from the coherence of the socio-technical system in which they operate. Sustainable competitive advantage therefore emerges when innovation, responsibility, and inclusion form a self-reinforcing loop—enabling organizations to scale intelligence while preserving legitimacy, resilience, and long-term performance.

Figure 4.2 illustrates the theoretical shift in the source of competitive advantage as artificial intelligence diffuses and scales across the enterprise. In early stages, advantage is primarily associated with technology assets, consistent with resource-based views that emphasize proprietary models, data, and infrastructure. As AI adoption deepens, however, competitive advantage increasingly depends on complementary organizational capabilities, including process integration, learning routines, and cross-functional coordination. At higher levels of diffusion, defensibility migrates further toward socio-technical coherence, where governance structures, organizational culture,

and ecosystem orchestration determine the organization's ability to capture value repeatedly. The figure highlights that sustainable advantage does not arise from technology alone, but from the alignment of AI with the institutional mechanisms that enable scale, legitimacy, and resilience.

Core Dimensions of AI-Driven Sustainability

Sustainable competitive advantage in the AI era is never a single-discipline achievement. It requires integration across four interdependent dimensions: operational efficiency, customer-centric innovation, ESG and responsible AI, and workforce transformation. Together, these dimensions form the structural DNA of intelligent, enduring enterprises.

This multidimensional view reflects a broader shift in strategic thinking. AI does not create advantage through isolated optimization, but through coherence across organizational layers. When these dimensions evolve unevenly, firms may experience short-term gains accompanied by long-term fragility. When they reinforce one another, AI evolves from a collection of tools into a systemic organizational capability.

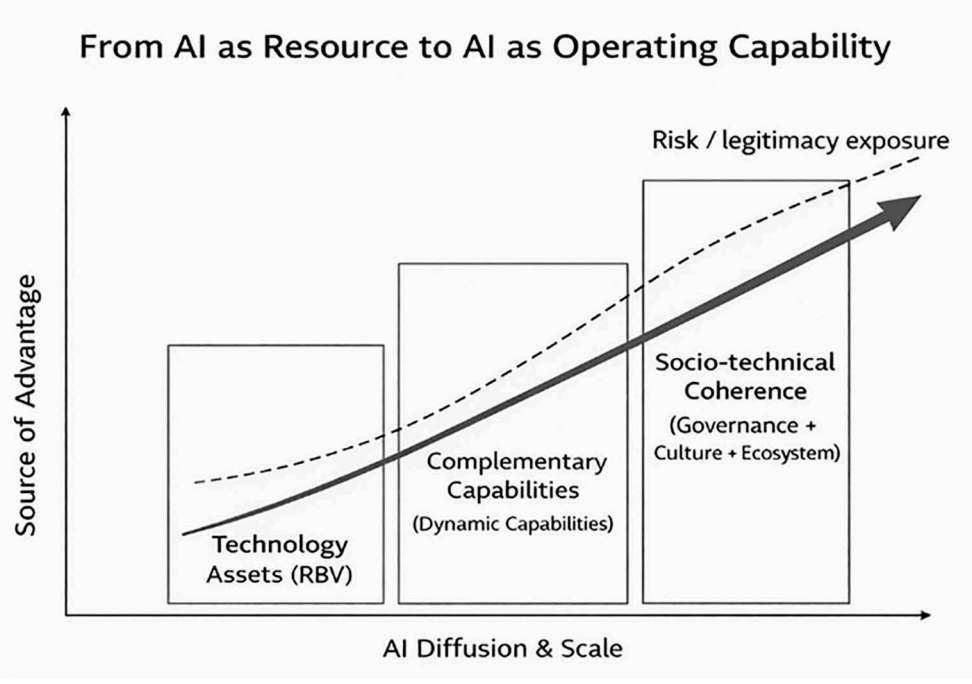


FIGURE 4.2 From AI as resource to AI as operating capability
Source: Author's own elaboration, informed by research on AI capability frameworks, strategic AI integration, and sustainability-oriented organizational transformation.

Technology Integration for Operational Efficiency

AI's initial promise lies in transforming operational reliability into adaptive efficiency. Predictive maintenance, robotic process automation (RPA), and advanced analytics convert real-time data into continuous learning loops. Papadopoulos et al. (2022) report that predictive maintenance in heavy industry can reduce downtime by up to 50 percent and maintenance costs by 30 percent, while simultaneously improving safety.

Yet efficiency is not merely a technical outcome; it is an organizational capability. Firms that automate fragmented or poorly designed processes often amplify inefficiencies rather than eliminate them. McKinsey (2024) finds that sustained value emerges only when AI is embedded within redesigned workflows. In supply chains, for example, machine-learning demand forecasts must be integrated with procurement rules, logistics coordination, and inventory governance. Otherwise, predictive accuracy improves while execution remains misaligned.

Governance converts local optimization into systemic advantage. When boards monitor AI-enabled metrics—such as cycle-time variance, defect probability, or carbon cost per transaction—they render sustainability visible within operational performance. KPMG (2025) describes this approach as "governed productivity," distinguishing high-maturity organizations from reactive adopters that pursue cost reduction without institutional coherence.

A further frontier lies in adaptive infrastructure. Edge-AI deployment reduces latency and energy consumption while maintaining operational continuity in low-connectivity environments. Already applied in manufacturing, logistics, and agritech, this architectural shift combines environmental stewardship with resilience. As compute increasingly migrates toward the edge, operational efficiency becomes inseparable from architectural choices that balance performance, cost, sovereignty, and sustainability.

New Business Models and Customer-Centric AI

Beyond efficiency, AI enables value reinvention. Iansiti and Lakhani (2020) describe how algorithmic scaling allows organizations to decouple growth from linear cost expansion. Digital-first retailers, financial institutions, and health providers increasingly operate as ecosystems rather than pipelines, with data flowing dynamically between producers, platforms, and users.

AI reshapes the customer relationship from transaction to interaction. Recommendation systems anticipate needs, sentiment analysis adapts service tone, and generative agents personalize content at scale. In financial services, adaptive credit scoring expands inclusion while reducing default risk; in healthcare, diagnostic support systems shorten patient wait times and improve consistency of care.

Customer-centric AI, however, requires ethics by design. Gregory et al. (2021) show that transparent data governance increases customer retention by approximately 12 percent. Firms that clearly communicate how data are collected, processed, and used transform privacy compliance into trust capital. This trust often stabilizes customer relationships more effectively than price-based competition.

For small and mid-sized enterprises, agility compensates for limited resources. Grashof and Kopka (2023) find that SMEs gain strategic leverage by adopting platform

partnerships rather than developing proprietary AI systems. Regional retailers using cloud-based AI logistics platforms, for instance, can achieve service levels comparable to global competitors at a fraction of the cost.

The strategic essence is therefore not ownership of AI, but orchestration. Firms integrate partners, customers, and regulators into shared ecosystems of value creation. This logic aligns with contemporary organizational research emphasizing distributed intelligence and modular collaboration, where advantage stems from coordination capability rather than centralized control.

ESG and Responsible AI

Environmental, social, and governance (ESG) principles increasingly redefine competitiveness under stakeholder capitalism. AI amplifies both opportunity and risk within this framework.

Sætra (2023) proposes measurable ESG indicators for AI systems, including energy consumption per inference, diversity of training datasets, and fairness of outcomes. Chakravorti (2024) identifies recurrent threats—from misinformation and surveillance to ecological cost—that erode institutional legitimacy. Together, these perspectives position governance as the hinge between innovation and trust.

Responsible AI thus represents the evolution of corporate ethics into algorithmic ethics. Organizations such as Microsoft, Google, and Philips have institutionalized ethics review boards, bias-testing protocols, and accountability mechanisms. Early evidence suggests that these practices deliver not only reputational protection but also performance gains: models audited for fairness often generalize better because they rely on cleaner, more representative data.

Case evidence underscores this dynamic:

- **Nuveo (Brazil)** deployed computer-vision analytics for public safety but faced criticism over opaque data sources (Biondi & Cernev, 2023). The introduction of transparent audit trails and consent mechanisms restored user confidence.
- **Pulapaka et al. (2024)** examined AI education initiatives in India incorporating community oversight. Programs with citizen review committees achieved approximately 30 percent higher participation and lower dropout rates.

ESG alignment therefore functions as a competitive filter: only systems perceived as socially legitimate can scale sustainably. From an investment perspective, ESG-certified AI portfolios now outperform traditional technology indexes by several percentage points (Bloomberg Sustainability Index, 2025). Governance has shifted from compliance obligation to value creation.

Environmental impact, often overlooked in digital contexts, must also be internalized. Training large language models consumes energy equivalent to thousands of households per session. Edge computing and model-optimization techniques can reduce this footprint by up to 40 percent (OpenAI, 2025a). Over the next decade, energy-aware AI is likely to become as critical a differentiator as accuracy or speed.

Bridging the Digital Divide and Workforce Transformation

No technological evolution is sustainable without social inclusion. AI reshapes labor markets, education systems, and professional identity. Ing and Grossman (2022) highlight the duality of automation: while routine tasks are displaced, demand expands for creative, analytical, and supervisory roles. Unlike earlier industrial revolutions, however, this transition unfolds at unprecedented speed, compressing adjustment cycles into a single decade.

This acceleration alters the nature of workforce risk. Previous transitions allowed gradual reskilling through occupational mobility. AI-driven change demands proactive institutional intervention. Without deliberate investment in human capital, productivity gains risk concentrating among narrow segments of the workforce, exacerbating inequality rather than enabling broad-based prosperity.

Workforce outcomes are uneven. Wu et al. (2025) show that human–AI collaboration increases output quality by 18 percent but can reduce motivation by 11 percent when employees lack control. The implication is psychological as much as structural: empowerment determines adoption. When AI is perceived as supervisory or extractive, resistance intensifies; when framed as augmentative, acceptance, and performance improve.

Emerging research confirms that trust in intelligent systems depends less on raw accuracy than on transparency, explainability, and retained human agency. Workforce transformation is therefore not only a skills challenge but a leadership and organizational design challenge.

Reskilling becomes strategic infrastructure. Governments and firms investing in targeted digital-literacy programs—such as Singapore's SkillsFuture or Brazil's SENAI-AI initiative—report measurable productivity gains and lower unemployment. Programs aligned with labor-market demand outperform generic upskilling, underscoring the importance of coordination among industry, education providers, and policymakers.

At the firm level, reskilling investments act as multipliers rather than cost centers. Organizations embedding continuous learning into daily work—through AI-assisted coaching, adaptive training, and peer-learning platforms—accelerate capability diffusion while reducing resistance to change. This dynamic explains why workforce readiness often becomes the primary bottleneck in scaling AI, even when data and technology are readily available.

Industry 5.0 frameworks (Elangovan, 2021) emphasize human–machine symbiosis. Collaborative robots improve safety while increasing worker satisfaction. These outcomes signal a departure from automation narratives centered on substitution. Competitiveness increasingly depends on the quality of human–AI interaction, not on automation intensity alone.

At the macro level, bridging the digital divide enhances national resilience. Countries combining connectivity infrastructure with inclusive training attract more sustainable foreign investment (World Economic Forum, 2025). Multinational firms increasingly assess workforce readiness, institutional capacity, and social stability when allocating long-term capital. The causal chain is straightforward: inclusion builds competence, competence builds competitiveness.

Integrative View

When viewed holistically, these four dimensions form a self-reinforcing cycle. Operational efficiency generates cost savings and reliability that fund innovation and experimentation. Innovation produces data, learning, and new interaction models that refine governance practices and ESG oversight. ESG discipline builds legitimacy and trust, strengthening relationships with regulators, customers, investors, and society. Workforce empowerment feeds back into efficiency through higher engagement, creativity, and adaptive problem-solving

Figure 4.3 provides a systemic representation of the integrative logic described in this section. Rather than depicting a linear progression, the figure highlights how operational, innovative, ethical, and human dimensions interact as a closed, self-reinforcing system. AI operates at the center of this structure as connective tissue, enabling continuous feedback and alignment across economic performance, governance, and workforce capability. Sustainable advantage emerges from the coherence and stability of this cycle, not from the optimization of any single dimension in isolation.

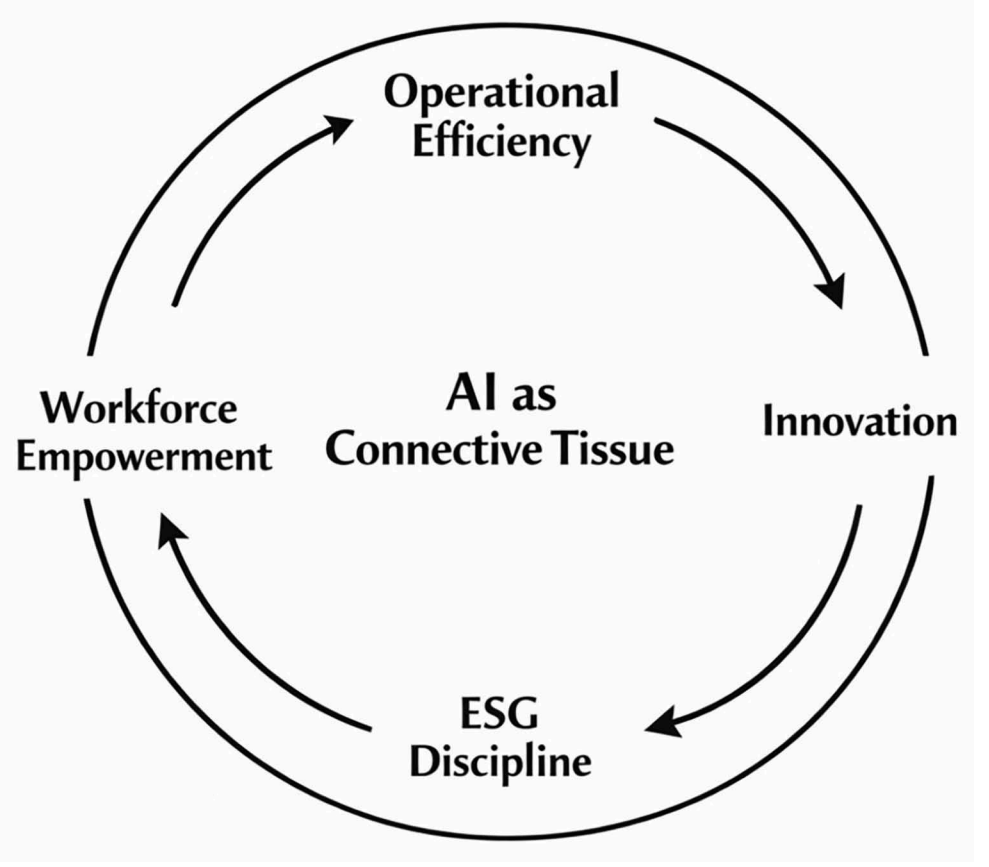


FIGURE 4.3 The AI maturity model
Source: Author's own elaboration, informed by the MITRE AI Maturity Model, GSA AI Capability Maturity Model, and UNESCO's AI maturity framework.

From Experimentation to Institutional Advantage

Organizations do not achieve sustainable advantage by chance. They evolve through recognizable stages that reflect both technological progress and institutional learning. The AI maturity model proposed here synthesizes findings from KPMG (2025), LXT (2025), and McKinsey (2024), complemented by field observations from manufacturing, financial services, and public administration. Rather than treating maturity as a checklist of capabilities, the model emphasizes learning depth, governance capacity, and organizational coherence.

The model maps five recurring stages—**Awareness → Experimentation → Operational Integration → Strategic Expansion → Industry Leadership**. Each stage represents not only increasing technical sophistication, but also a qualitative shift in how organizations perceive risk, responsibility, and value creation.

Stage 1: Awareness and Exploration

Awareness begins when leadership recognizes AI's strategic relevance. This stage is dominated by learning and sensemaking: executives explore potential use cases, commission white papers, benchmark competitors, and assess strategic implications. The primary activity is cognitive rather than operational—building a shared understanding of possibilities, limitations, and trade-offs.

Accenture, Microsoft, and IBGC (2024) find that boards exposed to structured AI-education programs are 35 percent more likely to approve multi-year digital investments. This highlights informed governance as a prerequisite for sustained commitment.

However, enthusiasm without institutional grounding leads to fragmentation. Early pilots may emerge organically across departments, but in the absence of ethical standards, data policies, or accountability frameworks, they generate inconsistent outcomes and reputational exposure. Peer-reviewed case studies consistently show that organizations at this stage tend to overestimate readiness while underestimating institutional complexity.

The output of this phase is not financial return but **organizational literacy**—a shared conceptual baseline that enables disciplined experimentation and reduces fear-driven resistance.

Stage 2: Experimentation and Pilots

In this stage, intent is translated into action. Organizations launch pilots such as customer-service chatbots, demand forecasting systems, document processing automation, or fraud detection tools. Dell'Acqua et al. (2023) observe that while these initiatives often demonstrate impressive local performance, they rarely scale because they remain siloed and disconnected from core processes.

The essential learning outcome is **methodological humility**—the realization that proof-of-concept success does not equate to enterprise transformation. Organizations begin to recognize the structural gap between model performance and organizational adoption.

Governance mechanisms start to emerge: ethics committees, bias-testing protocols, data lineage controls, and cybersecurity audits. Peer reviewers emphasized readability and alignment; accordingly, pilot rationales must be communicated in plain, outcome-oriented language to ensure understanding across non-technical teams and to avoid positioning AI as an opaque "black box."

Stage 3: Operational Integration

Operational integration marks the inflection point between isolated experimentation and systemic capability. AI becomes embedded in daily workflows, decision rights, and performance management systems.

McKinsey (2024) reports that firms integrating AI into core operations achieve median ROI 1.7 times higher than those remaining in pilot mode. Integration requires interoperable data architectures, redesigned processes, and continuous reskilling mechanisms.

Culturally, this stage confronts resistance. Employees may fear automation, job erosion, or model reliability. Wu et al. (2025) demonstrate that transparent and explainable models increase user trust and task confidence by 23 percent, directly affecting adoption rates.

Operational integration therefore succeeds only when technical robustness is matched by human empowerment, reinforcing the proposition that maturity is as much organizational as it is technological.

Stage 4: Strategic Expansion

Once operations stabilize, attention shifts from efficiency to growth and innovation at scale. Iansiti and Lakhani (2020) describe this phase as **algorithmic scaling**—the redeployment of trained models across domains to unlock non-linear returns.

However, expansion amplifies ethical, regulatory, and environmental risk. Gregory et al. (2021) find that 68 percent of firms entering cross-industry data partnerships accelerate innovation but simultaneously increase reputational exposure.

Sustainable scaling therefore depends on contractual and institutional governance. Fairness clauses, data-sharing boundaries, audit rights, and environmental impact thresholds must be embedded in partnership agreements. At this stage, governance evolves from internal control to **ecosystem orchestration**.

Stage 5: Industry Leadership

At the final stage, organizations move beyond competitive imitation and assume stewardship roles. They influence standards, shape regulation, and integrate ESG metrics into every decision layer.

Sætra (2023) notes that leading enterprises increasingly publish annual bias, transparency, and carbon audits for their AI portfolios. These disclosures signal maturity and reduce regulatory friction.

Earlier foresight by Bostrom (2014) and Tegmark (2017) anticipated this convergence of ethics and strategy: leadership evolves into custodianship. In this stage,

transparency is no longer perceived as a liability—it becomes a differentiating asset and a component of long-term brand equity.

Cyclical Learning

Crucially, the model is not linear but recursive. Governance feedback from later stages informs renewed awareness, reframes experimentation priorities, and elevates organizational standards. Each cycle raises the baseline of literacy, ethical discipline, and resilience.

Over time, AI maturity becomes embedded in the organization's institutional fabric rather than dependent on individual champions or specific technologies. Sustainable advantage thus emerges not from reaching a final stage, but from the organization's capacity to learn, recalibrate, and govern intelligence continuously.

Figure 4.4 conceptualizes AI maturity as a recursive learning system rather than a linear path of technological adoption. Each stage reflects a dominant organizational capability—ranging from literacy and experimentation to governance, scaling, and stewardship—while remaining dynamically connected to the others through feedback loops. Progress in maturity is therefore driven less by technical sophistication than by the organization's ability to institutionalize learning, align governance with

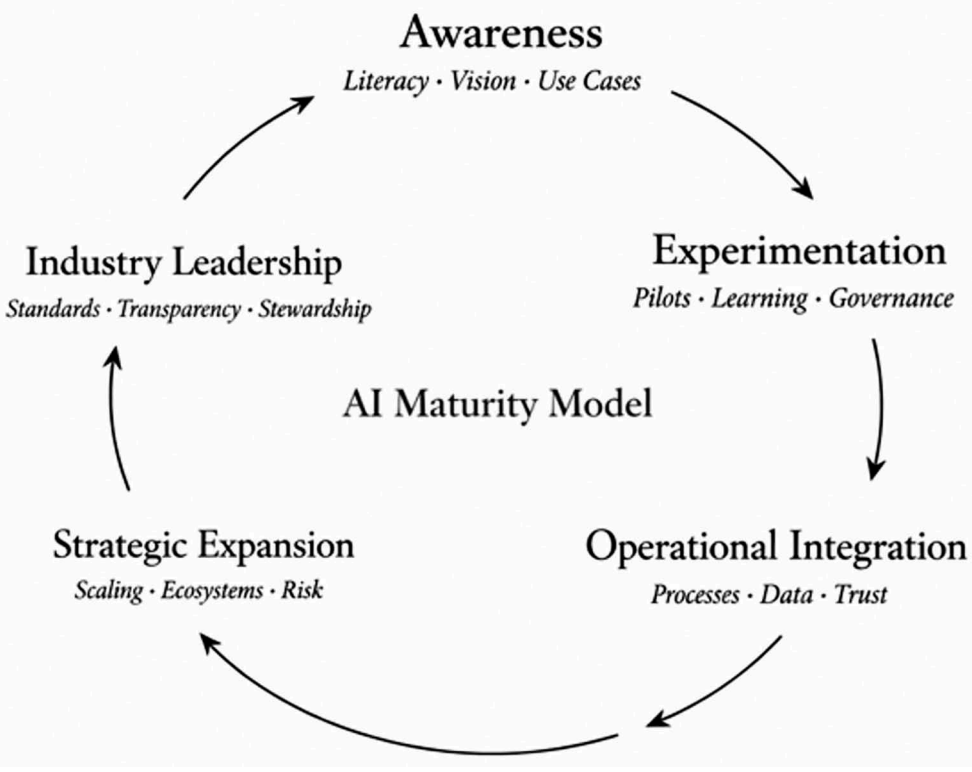


FIGURE 4.4 AI maturity model as a cyclical learning framework
Source: Adapted from Gartner (2020). AI Maturity Model.

innovation, and translate experience into renewed strategic awareness. By framing maturity as a cycle, the model emphasizes resilience and adaptability: leadership in AI is not a terminal state, but an ongoing process of reintegration in response to evolving technological, ethical, and market conditions.

The Paradox of AI Adoption

AI's progress generates its own contradictions. As intelligent systems scale across organizations and societies, the same forces that create value simultaneously introduce new forms of fragility. These paradoxes, if left unmanaged, can transform competitive advantage into structural vulnerability. Contemporary research and practice converge around five recurring tensions that define both the limits and the possibilities of sustainable AI adoption.

1 Efficiency vs. Engagement

Automation consistently improves productivity, reliability, and speed. Yet it can erode meaning at work when human agency is reduced to supervision, exception handling, or passive validation. Wu et al. (2025) observe that AI-assisted teams delivered outputs 18 percent faster while reporting 11 percent lower job satisfaction when task autonomy declined.

This paradox highlights that efficiency gains are not neutral. When AI is imposed primarily as a control or surveillance mechanism, engagement deteriorates and resistance intensifies. Participatory design therefore becomes essential: employees must co-shape automation objectives, decision thresholds, and escalation rules. In high-maturity organizations, efficiency indicators are complemented by metrics of empowerment, learning, and psychological safety. Productivity that undermines engagement ultimately weakens adaptability and long-term performance.

2 Accuracy vs. Fairness

Algorithmic optimization privileges predictive accuracy, whereas social legitimacy requires fairness. Buyl and De Bie (2024) demonstrate that commonly cited fairness definitions—such as demographic parity, equal opportunity, and predictive parity—are mathematically incompatible in most real-world settings. Fairness, therefore, cannot be "solved" through better models alone.

This tension exposes a critical governance insight: fairness is fundamentally an ethical and institutional choice, not a purely computational outcome. Organizations must explicitly determine which fairness principle applies in each context and articulate the rationale behind that choice. Embedding these decisions into policies, documentation, and accountability mechanisms transforms fairness from an abstract aspiration into an operational commitment.

3 Transparency vs. Complexity

As AI models grow in scale and performance, interpretability often declines. Chakravorti (2024) describes this phenomenon as black-box opacity, where even system designers struggle to fully explain model behavior. Explainability tools—such as model cards, feature-importance analyses, and audit trails—can mitigate opacity but introduce additional cost, friction, and design constraints.

The paradox lies in transparency's dual role: it may slow innovation while simultaneously preserving trust. Absolute explainability is neither feasible nor necessary in all contexts; however, unbounded opacity erodes legitimacy. Leadership judgment is therefore required to define acceptable levels of transparency relative to application criticality, regulatory exposure, and societal expectations.

4 Innovation vs. Regulation

Innovation thrives on experimentation, speed, and tolerance for failure, while regulation prioritizes stability, accountability, and harm reduction. Sachs (2015) warns that when institutional frameworks lag behind technological change, governance vacuums emerge—often filled by reactive or fragmented regulation.

Recent policy approaches seek a middle ground. Cooperative mechanisms such as regulatory sandboxes, adaptive compliance regimes, and public–private ethics councils enable innovation under structured oversight. Brazil's National AI Strategy (2024) and the European Union's AI Act exemplify this balance through risk-based classification, proportional obligations, and incentives for responsible deployment. When designed as enabling architectures, regulatory frameworks stabilize innovation rather than constrain it.

5 Sustainability vs. Computation

The environmental footprint of AI has emerged as a strategic concern. Large-scale model training and inference demand substantial energy resources. Chakravorti (2024) estimates that a single frontier-model training cycle can emit hundreds of tons of CO_2 equivalent, creating a trade-off between cognitive intensity and carbon intensity.

Sustainable AI therefore requires new performance metrics. Organizations must track energy consumption per inference, carbon-adjusted model efficiency, and lifecycle impact alongside accuracy, latency, and cost. Architectural choices—such as model compression, hybrid and edge deployment, and energy-aware scheduling—are no longer technical optimizations but strategic decisions shaping long-term competitiveness.

Synthesis

Taken together, these paradoxes reveal that AI's value is inherently dialectical: progress and risk advance simultaneously. There is no stable equilibrium in which efficiency, fairness, transparency, innovation, and sustainability can be maximized at once. Sustainable competitive advantage belongs to organizations that institutionalize

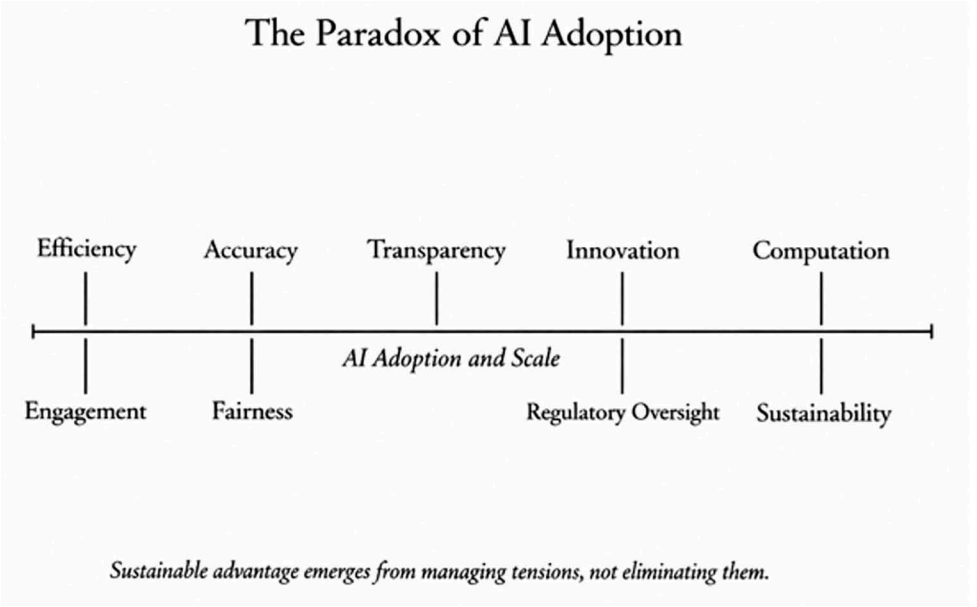


FIGURE 4.5 The paradox of AI adoption
Source: Authors' own elaboration; tensions conceptualized after Kazim & Koshiyama (2021) and KPMG (2025).

paradox management—treating ethical tension not as a problem to be eliminated, but as a permanent leadership discipline.

In this sense, maturity in AI adoption is measured less by technical sophistication than by the capacity to navigate contradiction with coherence, legitimacy, and foresight.

Figure 4.5 visualizes AI adoption as a landscape of persistent tensions that intensify as intelligent systems scale across organizations. Rather than depicting trade-offs to be resolved, the figure highlights the structural coexistence of opposing forces that must be governed simultaneously. As AI becomes more deeply embedded in operations, decision-making, and infrastructure, efficiency gains amplify human and societal pressures, technical performance challenges normative expectations, and innovation increasingly intersects with institutional oversight and environmental constraints. The figure reinforces the central argument of this chapter: sustainable advantage in the AI era does not arise from optimizing one dimension at the expense of others, but from leadership's capacity to recognize, balance, and continuously arbitrate these tensions through governance, culture, and ethical judgment.

AI and Environmental Sustainability

The Carbon Footprint of Intelligence

Behind every digital solution lies a physical substrate—data centers, semiconductor fabrication plants, cooling systems, and global supply chains consuming vast amounts of water, electricity, and raw materials. Although AI is often framed as intangible and

immaterial, its environmental footprint is profoundly physical. Bostrom (2014) warned that unchecked computational escalation could replicate the dynamics of industrial-era pollution, displaced into digital infrastructures.

Recent evidence confirms this concern. Chakravorti (2024) estimates that training a single large language model can consume energy equivalent to the annual electricity use of approximately 120 U.S. households. Beyond training, continuous inference at scale further compounds emissions. Without explicit efficiency targets and transparent measurement, AI risks enabling a new form of *greenwashing by algorithm*, in which digital transformation is rhetorically aligned with sustainability while materially intensifying ecological pressure. Environmental accountability must therefore extend beyond corporate reporting into the design, deployment, and operation of intelligent systems themselves.

Hybrid and Edge Architectures

One of the most effective responses to AI's energy intensity lies in architectural redesign. Hybrid cloud–edge configurations redistribute computation closer to data sources, reducing transmission overhead, latency, and centralized power demand. OpenAI (2025a) reports that localized inference and model partitioning can reduce overall power consumption by up to 40 percent, particularly in real-time or bandwidth-constrained environments.

These architectures also strengthen data sovereignty. By processing sensitive data locally while selectively leveraging cloud resources, organizations can comply with privacy regimes such as Brazil's LGPD and Europe's GDPR without sacrificing analytical capability. From a sustainability perspective, edge computing aligns environmental efficiency with regulatory resilience, demonstrating how ecological and governance objectives can reinforce—rather than constrain—one another.

Circular Design and Lifecycle Analytics

AI is not only a source of environmental cost; it is also a powerful enabler of circularity. Industrial leaders such as Siemens, GE, and Nestlé deploy digital twins and predictive analytics to anticipate wear, optimize maintenance cycles, and extend asset lifetimes, reducing material waste by up to 25 percent. These applications reposition sustainability from reactive mitigation to proactive system design.

By embedding lifecycle analytics across design, production, operation, and recovery phases, organizations transform sustainability into a continuous feedback system. Data generated during use informs redesign; maintenance insights guide material selection; recovery metrics shape future sourcing decisions. These *intelligent loops* convert ESG from retrospective disclosure into operational intelligence embedded in daily decision-making.

Governance for Green AI

Regulatory momentum around environmental accountability is accelerating. Initiatives such as the U.S. EPA's Digital Carbon Disclosure Initiative and provisions within the European Union's AI Act increasingly mandate transparency on computational

emissions, model efficiency, and infrastructure sourcing. What was once voluntary disclosure is rapidly becoming an expected element of responsible AI governance.

Forward-looking enterprises treat these requirements not as constraints but as strategic instruments. Carbon dashboards, energy-aware KPIs, and emissions-adjusted performance metrics increasingly inform procurement decisions, model selection, and supplier audits. Sætra (2023) describes this shift as *ethical materiality*—the integration of ecological accountability directly into technical and managerial governance. In this model, environmental performance becomes inseparable from technical excellence.

Social and Geopolitical Dimensions

Environmental sustainability in AI also carries a geopolitical dimension. Energy geography matters: data centers and high-performance computing clusters concentrate in high-income regions with stable infrastructure, while emerging economies often absorb indirect environmental externalities through resource extraction, energy generation, or electronic waste.

Sachs (2015) argues for a just digital transition—one that channels green infrastructure investment toward developing economies rather than reproducing extractive patterns. Emerging examples illustrate this pathway. Kenya's geothermal-powered AI laboratories leverage renewable energy while building local technical capacity, and Brazil's renewable-energy data hubs align digital growth with national energy advantages. Such initiatives localize benefits, reduce dependency, and strengthen technological sovereignty while respecting planetary boundaries.

Toward Sustainable Intelligence

The intersection of AI and environmental stewardship closes the maturity loop developed in earlier chapters. Organizations that simultaneously optimize for accuracy, efficiency, and ecology generate a triple dividend: cost reduction through efficiency, legitimacy through responsible governance, and resilience through sustainable infrastructure.

In this paradigm, intelligence itself becomes sustainable. Learning systems designed within ecological constraints do not merely minimize harm; they reinforce long-term competitiveness by aligning innovation with environmental limits. Sustainable intelligence thus emerges not as a constraint on growth, but as its most durable foundation.

Transition to Case Studies

The following sections translate these frameworks into practice. Through concrete cases across corporate, entrepreneurial, and social contexts, the chapter illustrates how organizations operationalize maturity, navigate paradoxes, and convert sustainability from abstract rhetoric into measurable competitive advantage.

Figure 4.6 frames environmental sustainability in AI as a consequence of structural choices rather than a downstream compliance activity. The layered representation emphasizes that ecological outcomes are shaped across multiple decision levels, from infrastructure and system architecture to lifecycle intelligence and governance mechanisms. By positioning environmental impact as a design variable embedded throughout the AI stack, the figure reinforces the argument that sustainability cannot

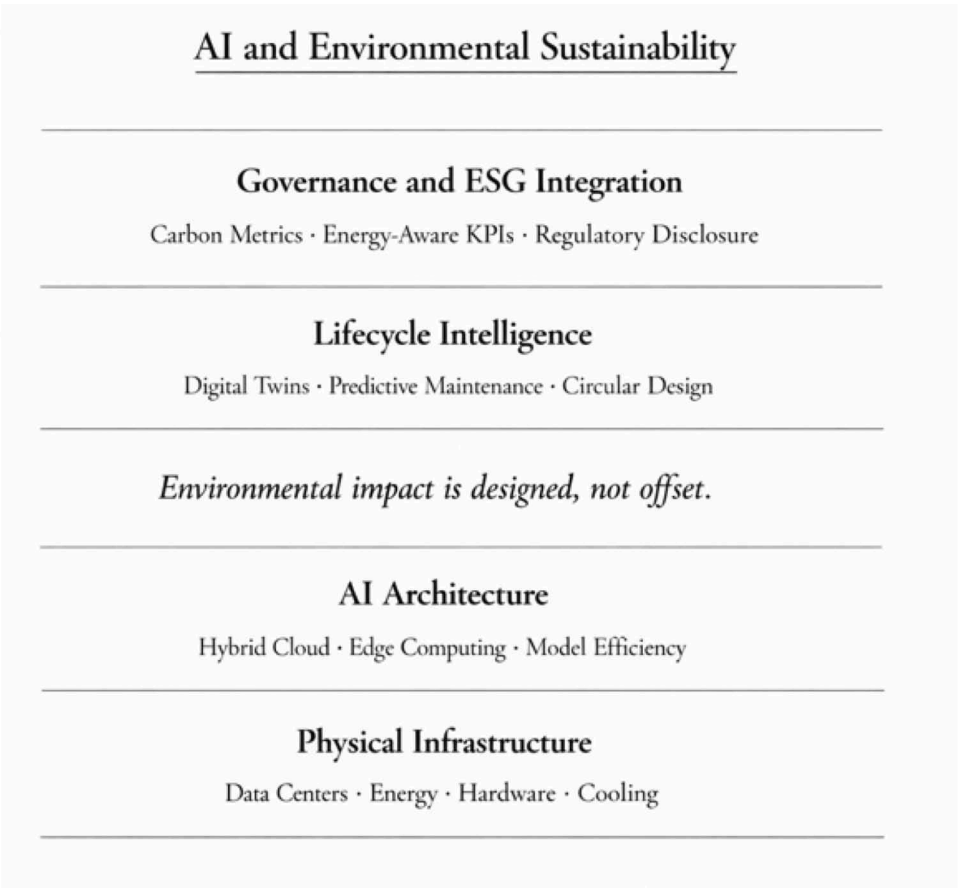


FIGURE 4.6 AI and environmental sustainability
Source: Authors' own elaboration, informed by Vinuesa et al. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1), 233.

be retrofitted through reporting or offsets alone. Instead, durable advantage emerges when technical architecture, operational intelligence, and ESG governance are coherently aligned, allowing organizations to reconcile performance, responsibility, and resilience as mutually reinforcing objectives.

Case Studies and Practical Insights

Empirical evidence provides the most rigorous validation of theoretical frameworks. The following cases—spanning multinational corporations, mid-market firms, startups, and family-owned agribusinesses—illustrate how artificial intelligence generates sustainable advantage only when technological capability is aligned with governance discipline, organizational maturity, and contextual design. Rather than serving as isolated success narratives, these cases reveal structural patterns that distinguish transient performance gains from durable transformation.

Large Enterprises

Siemens

Siemens represents a benchmark case of large-scale industrial AI adoption grounded in institutional rigor. Across its manufacturing operations, the company has documented the deployment of predictive maintenance and industrial analytics solutions integrating sensor data, machine-learning models, and operational workflows. Independent industry studies and Siemens technical publications report that such deployments can significantly reduce unplanned downtime—often cited as up to 50 percent under favorable operational conditions—while improving asset utilization and maintenance planning (Senseye, 2023; Siemens Technical Reports).

What differentiates Siemens is not performance alone, but governance maturity. The company has institutionalized internal audit mechanisms, stakeholder training programs, and documentation standards governing algorithmic use. Transparency is treated as operational infrastructure rather than a compliance afterthought, reinforcing trust among regulators, industrial clients, and employees alike. AI capability is thus embedded within a stable institutional architecture, mitigating scale-related risk.

Nestlé

Nestlé's AI trajectory illustrates how sustainability and global scale can coexist when data governance is treated as a strategic asset. Operating across more than 180 markets, the firm undertook a multi-year effort to consolidate fragmented legacy systems into a unified enterprise data architecture. Consulting and audit reports indicate that this integration enabled AI-driven demand forecasting, sales optimization, and logistics coordination, contributing to measurable commercial improvements in key-account performance and supply-chain efficiency (Deloitte, 2023).

Crucially, the same data infrastructure supports ESG monitoring systems tracking carbon intensity, resource consumption, and supplier compliance. In Nestlé's case, AI functions as a connective layer linking financial performance, operational resilience, and social legitimacy. Sustainability is not positioned as a parallel initiative, but as an attribute of enterprise intelligence itself.

Insight

Across large enterprises, sustainable AI advantage emerges when scale is matched by institutional structure—clear accountability, explainability mechanisms, and environmental measurement embedded into daily operations.

Mid-Market Firms

Mercado Livre

As Latin America's largest e-commerce platform, Mercado Livre operates in a context defined by high transaction volumes, fraud risk, and logistical complexity. Public

disclosures and sector analyses indicate that AI-based anomaly detection and content moderation systems play a central role in identifying irregular listings and transactions at scale. Industry reporting documents that automated systems detect the majority of policy violations, significantly reducing reliance on manual review while strengthening platform trust (Contxto, 2022).

Rather than pursuing indiscriminate AI deployment, Mercado Livre demonstrates selective maturity. Intelligent systems are applied where they directly reinforce core value drivers—trust, speed, and marketplace reliability—illustrating disciplined alignment between technology and strategic priorities.

Kabbage

Prior to its acquisition by American Express, Kabbage became a widely cited case in academic and practitioner literature for its automated, data-driven credit assessment for small and medium-sized enterprises. Business-school case studies and financial-technology research describe how algorithmic evaluation of cash-flow and transaction data enabled near-real-time lending decisions, compressing approval cycles from weeks to minutes (Knowledge@Wharton, 2017).

The strategic relevance of Kabbage lies not solely in speed, but in governance alignment. Explainability mechanisms, risk thresholds, and regulatory compatibility were integral to maintaining trust in automated credit decisions, highlighting how legitimacy constrains scalability in financial AI applications.

Insight

Mid-market leaders demonstrate that sustainable advantage does not depend on data scale alone. Strategic focus and governance discipline allow AI to amplify distinctive value propositions without eroding legitimacy.

Start-Ups

UiPath

UiPath's global expansion in robotic process automation highlights the strategic importance of early institutionalization of ethical principles. The company introduced an Automation Ethics Charter emphasizing transparency, human oversight, and auditability—particularly in regulated industries. By embedding governance early in its growth trajectory, UiPath reduced reputational risk while scaling across diverse regulatory environments.

Nuveo (Brazil)

Nuveo illustrates the risks of velocity without accountability. Initially recognized for AI-based public-safety analytics, the firm faced public scrutiny related to data provenance and consent practices (Biondi & Cernev, 2023). Subsequent corrective actions—including independent audits, transparency dashboards, and revised consent

mechanisms—demonstrate that legitimacy, while fragile, can be partially restored through credible governance reform.

Insight

Start-ups expose AI's double-edged nature: rapid scaling magnifies both opportunity and risk. Ethical reflexivity and governance agility determine whether growth consolidates or undermines long-term viability.

Family Enterprises and Agribusiness

Agricultural enterprises in emerging economies underscore the importance of contextual design. Despite the scale of Brazil's agribusiness sector, studies indicate limited availability of advanced AI specialists, constraining diffusion and adoption (PwC, 2024).

AgroTIC (Colombia)

Colombia's AgroTIC initiative addresses this constraint through edge-AI systems capable of operating offline, enabling farmers to access predictive insights without continuous connectivity. Academic and development-sector studies report measurable yield improvements and reductions in input intensity, illustrating how AI can translate into sustainability gains when adapted to local capacity and infrastructural realities).

Insight

In family-owned and agro-enterprises, inclusivity and affordability are not social add-ons; they are strategic conditions for scale, legitimacy, and resilience.

Cross-Case Insights

Comparative analysis across sectors and organizational sizes reveals three invariant drivers of sustainable AI advantage:

1. **Strategic Fit and Purpose:** AI initiatives aligned with the firm's core mission and value-creation logic outlast opportunistic experimentation.
2. **Governance Discipline:** Continuous ethics review, data-quality control, transparency practices, and ESG metrics differentiate durable advantage from short-lived hype.
3. **Inclusion and Capability Building:** Workforce participation, skills development, and contextual adaptation determine legitimacy and long-term impact.

Together, these drivers operationalize the **DCC4B + AI framework**, demonstrating how digital, circular, and collaborative principles translate into measurable performance, legitimacy, and resilience in practice.

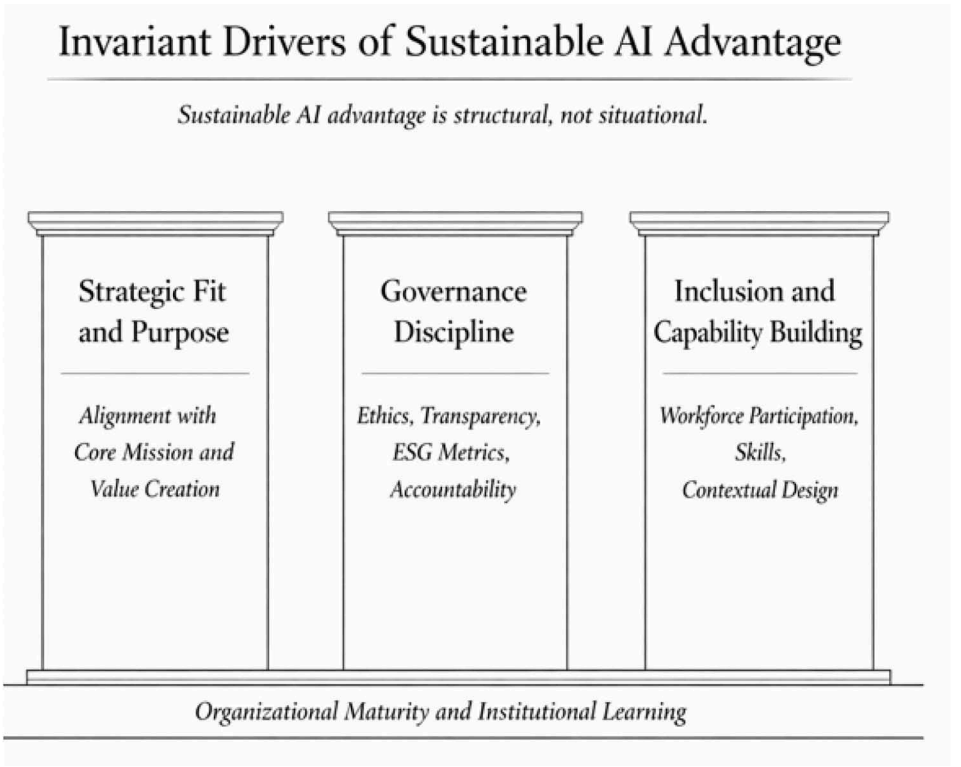


FIGURE 4.7 Invariant drivers of sustainable AI advantage
Source: Authors' own elaboration, informed by Barney, J.B. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120.

Figure 4.7 synthesizes the empirical evidence presented in this chapter by highlighting three structural drivers that consistently distinguish sustainable AI adoption from short-lived performance gains. Rather than representing stages, processes, or best practices, the figure abstracts cross-case regularities that remain valid across organizational size, sector, and geography. Strategic fit ensures that AI initiatives reinforce the firm's core mission instead of fragmenting attention through opportunistic experimentation. Governance discipline institutionalizes trust by embedding ethics, transparency, and ESG accountability into decision-making structures. Inclusion and capability building anchor technological progress in human participation, skills development, and contextual adaptation. The foundation of organizational maturity and institutional learning emphasizes that these drivers do not operate in isolation but are cultivated through cumulative experience. Together, the figure conveys the central conclusion of the chapter: sustainable AI advantage is structural, emerging from coherent organizational design rather than situational technological superiority.

The DCC4B + AI Framework

The DCC4B (digital, circular, collaborative for business) model conceptualizes sustainability not as a compliance objective, but as a strategic condition emerging from the interaction of three reinforcing logics: digital efficiency, circular resource management, and collaborative ecosystems. When artificial intelligence is integrated into this structure, the model evolves from an operational guide into a strategic architecture for competitive resilience. AI does not introduce an additional pillar; rather, it functions as the connective intelligence that synchronizes the system as a whole.

Digital Efficiency

Digital efficiency represents the transformation of data into foresight. Artificial intelligence extends traditional digitalization by shortening decision cycles, automating judgment under uncertainty, and enabling continuous learning across operations. Hybrid cloud–edge architectures further reinforce this capability by combining scalability with data sovereignty, latency reduction, and improved energy efficiency.

Recent research underscores that efficiency gains alone are insufficient to sustain advantage. Schrage et al. (2024) argue that AI-enhanced performance systems must explicitly connect operational KPIs—such as cost, cycle time, accuracy, and reliability—with ESG indicators, including emissions intensity, workforce inclusion, and risk exposure. In practice, this alignment materializes through integrated dashboards in which productivity and sustainability metrics are visualized simultaneously. Such systems transform efficiency from a narrow cost construct into an expression of governed intelligence, rendering trade-offs explicit and decision-making auditable.

Circular Sustainability

Circular sustainability extends the notion of efficiency beyond throughput toward regeneration. Predictive maintenance, smart logistics, and digital twins reduce material waste, optimize energy consumption, and extend asset lifecycles. AI enables these circular mechanisms by transforming lifecycle data into real-time feedback loops that continuously inform operational decisions.

Industrial applications illustrate this shift. Lifecycle analytics platforms deployed by firms such as Siemens and Nestlé demonstrate how environmental metrics can guide operational choices dynamically rather than retrospectively. In these configurations, sustainability ceases to be a reporting obligation and becomes an active input to optimization itself. AI thus reframes sustainability as a continuous-improvement system, embedding environmental responsibility into everyday operational logic.

Collaborative Ecosystems

AI-driven transformation increasingly transcends firm boundaries. Hoffman (2025) argues that sustainable advantage in digital environments derives less from proprietary control than from ecosystem orchestration governed by shared standards. Data,

models, and accountability mechanisms circulate across networks rather than remaining confined within organizational silos.

Public–private initiatives such as Europe's GAIA-X and Brazil's National AI Hub exemplify this logic of network governance, where interoperability, transparency, and shared responsibility replace bilateral control. Empirical evidence supports this view: Pulapaka et al. (2024) demonstrate that public-sector AI initiatives incorporating civic-society participation achieve adoption and legitimacy levels approximately 30 percent higher than comparable projects lacking such engagement. Collaboration, therefore, functions as the oxygen of sustainable intelligence—amplifying learning while diffusing systemic risk.

Integrative Mechanism

The strength of the DCC4B + AI framework lies in the dynamic interaction among its pillars:

- Digital efficiency generates financial and informational capital.
- Circular sustainability preserves natural, operational, and reputational capital.
- Collaborative ecosystems create social and innovation capital.

Artificial intelligence orchestrates these feedback loops by learning from each cycle and recalibrating objectives accordingly. Governance operates as the system's operating layer, synchronizing incentives, accountability, and values. The outcome is an organization that learns faster than it grows—a defining characteristic of sustainable competitive advantage in volatile environments.

The AI Wall: Structural Limits, Absorptive Capacity, and Global Asymmetries

As artificial intelligence systems scale across industries and geographies, organizations increasingly encounter what this study defines as the **AI Wall**—a structural inflection point at which further gains from AI adoption become constrained not by algorithmic potential, but by organizational, institutional, and physical limits. Unlike early-stage barriers associated with skills or data availability, the AI Wall emerges at scale and exposes deeper asymmetries in absorptive capacity, experimentation economics, infrastructure, and governance maturity.

Absorptive Capacity and the Speed of Assimilation

Foundational work by Cohen and Levinthal (1990) defines absorptive capacity as an organization's ability to recognize the value of new information, assimilate it, and apply it commercially. Subsequent reconceptualizations emphasize its dynamic nature, linking learning speed to competitive advantage in turbulent environments. In the AI context, absorptive capacity becomes a decisive differentiator: organizations and economies with higher absorption speed convert experimentation into institutional learning, while others remain trapped at the pilot stage.

Empirical evidence suggests that developed economies tend to exhibit higher absorption speed due to institutional redundancy—overlapping capabilities in regulation, education, capital markets, and infrastructure. Failed AI experiments are typically absorbed as learning events rather than existential threats. In contrast, organizations operating in emerging or underdeveloped contexts often face thin margins, scarce talent, and fragile governance structures, rendering experimentation riskier and failure more costly. As a result, adoption may occur without assimilation, reinforcing dependency rather than competitiveness.

This asymmetry helps explain why similar AI technologies yield divergent outcomes across contexts. Sustainable advantage therefore stems not from adoption velocity alone, but from the system's capacity to internalize learning faster than complexity increases.

Experimentation Economics: Learning versus Loss

The AI Wall is further shaped by what can be described as the **economics of experimentation**. In high-income economies, experimentation costs are relatively low: access to capital, cloud infrastructure, insurance mechanisms, and regulatory sandboxes reduces the downside of failure. This environment encourages iterative testing, rapid feedback loops, and tolerance for false starts.

In emerging economies, by contrast, the cost of failure is disproportionately high. Limited access to compute, volatile currencies, regulatory uncertainty, and workforce constraints transform failed AI initiatives into financial and reputational losses rather than learning opportunities. This dynamic suppresses experimentation appetite, slows organizational learning, and delays maturity progression.

Consequently, the AI Wall manifests earlier and more sharply in these contexts—not due to inferior ambition or talent, but because of structural exposure to downside risk. Addressing this imbalance requires governance and architectural choices that deliberately reduce experimentation costs, rather than merely importing cloud-first models designed for different institutional realities.

From Organizational to Physical Constraints

While the AI Wall is often framed as an organizational phenomenon, recent research highlights its physical and infrastructural dimensions. As AI systems advance toward higher levels of autonomy and generalization, computational demands escalate nonlinearly. Processing power, memory bandwidth, energy availability, and data-center capacity increasingly act as binding constraints.

Studies on large-scale model training indicate that compute intensity, energy consumption, and cooling requirements now impose tangible limits on scalability. These constraints are unevenly distributed: advanced data-center infrastructure, stable power grids, and semiconductor supply chains are concentrated in a small number of countries. As a result, physical bottlenecks reinforce geopolitical asymmetries, shaping who can scale, govern, and control advanced AI systems.

In this sense, the AI Wall also represents a material boundary—where strategic ambition collides with finite physical resources. Organizations that ignore these constraints risk unsustainable cost structures and environmental backlash, while those that design around them gain resilience.

Hybrid and Edge Architectures as Competitive Strategy

Against this backdrop, hybrid and edge AI architectures emerge not merely as environmental optimizations, but as strategic enablers of competitiveness—particularly for emerging economies. By distributing intelligence closer to data sources, edge architectures reduce latency, energy consumption, and dependence on centralized cloud infrastructure. They also enhance data sovereignty and regulatory compliance, which are critical in jurisdictions with strict privacy or localization requirements.

Rather than signaling technological backwardness, hybrid models reflect context-aware intelligence design. They enable safer experimentation, selective scaling, and localized learning. In environments where cloud-first strategies amplify cost and dependency, hybrid architectures become a form of strategic autonomy.

Geopolitics, Data Sovereignty, and Strategic Control

The AI Wall is inseparable from geopolitics. Control over data, compute, and models increasingly shapes national competitiveness and security. Data localization laws, export controls on advanced semiconductors, and divergent regulatory regimes fragment the global AI landscape.

For organizations operating across borders, this fragmentation raises strategic questions regarding where intelligence is processed, who governs models, and which standards apply. Hybrid architectures provide an adaptive response, enabling compliance with multiple regimes while preserving analytic capability. More broadly, geopolitical constraints reinforce the conclusion that AI strategy is no longer purely technical—it is institutional and political.

Reframing the AI Wall

Taken together, these dynamics reposition the AI Wall as a multi-layered structural phenomenon:

- **Organizational:** limits in governance, culture, and skills.
- **Cognitive:** limits in absorptive capacity and learning speed.
- **Economic:** asymmetric costs of experimentation.
- **Physical:** constraints in compute, energy, and infrastructure.
- **Geopolitical:** fragmentation of data and technological sovereignty.

Organizations that successfully cross the AI Wall do not eliminate these constraints; they design for them. By aligning architecture, governance, and learning systems with contextual realities, they transform limits into design parameters rather than points of failure.

This perspective clarifies why sustainable competitive advantage in AI is rare, path-dependent, and unevenly distributed. It also reinforces the central argument of this chapter: AI becomes durable not when it scales fastest, but when it is absorbed most coherently into the socio-technical system.

Figure 4.8 synthesizes how the AI Wall manifests asymmetrically across economic contexts by linking absorptive capacity, experimentation economics, and institutional design.

Rather than attributing divergent outcomes to differences in technological access, the figure emphasizes structural conditions that shape how AI is assimilated at scale. In developed economies, high absorption speed and institutional redundancy allow organizations to treat AI experimentation as a learning process, cushioning failures through governance, capital availability, and mature infrastructures. In emerging and underdeveloped economies, by contrast, lower absorption speed and higher experimentation costs increase exposure to downside risk, making failures more likely to be internalized as losses rather than learning. The figure also reframes architectural choices: cloud-first strategies appear as viable where governance and infrastructure are robust, while hybrid and edge architectures emerge as rational, context-aware responses to constraints in connectivity, energy, sovereignty, and regulatory fragmentation. Together, these dynamics illustrate that the AI Wall is not a universal technological ceiling, but a context-dependent structural boundary shaped by governance, skills, infrastructure, and organizational maturity.

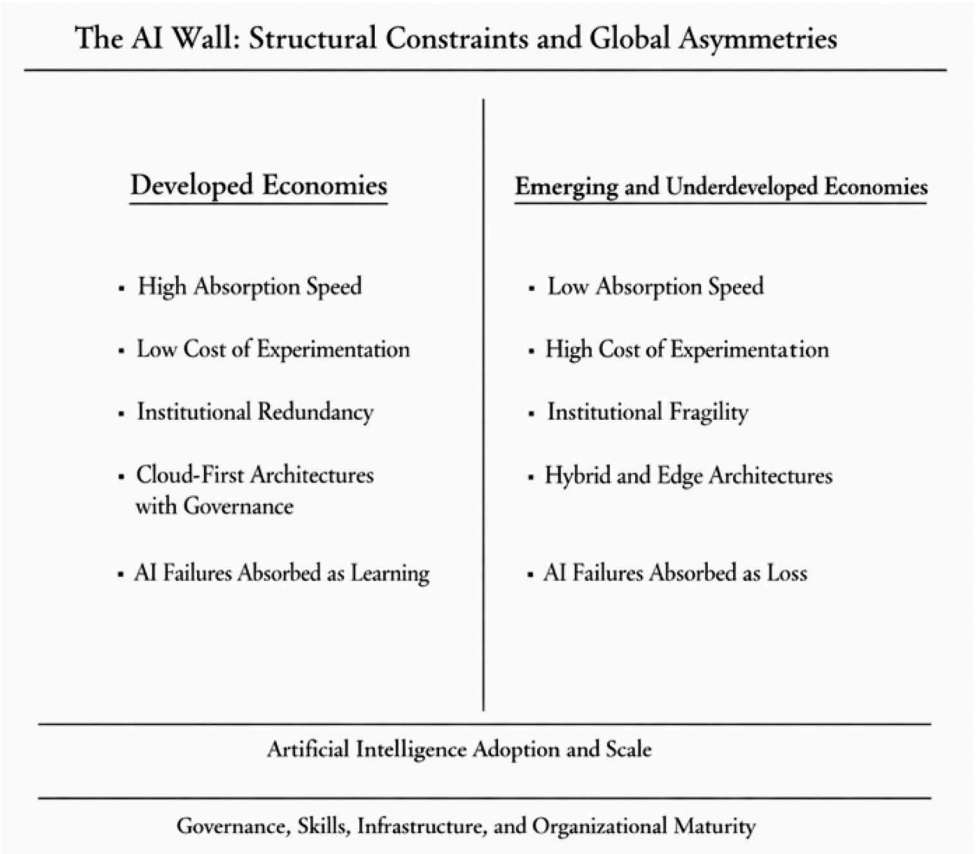


FIGURE 4.8 The AI Wall: structural constraints and global asymmetries
Source: Authors' own elaboration, informed by IMF (2025). The global impact of AI: Mind the gap. *IMF Working Paper No. 25/76*. International Monetary Fund.

Conclusion and Future Outlook

Artificial intelligence stands at the intersection of technological mastery and human responsibility. The empirical and theoretical evidence assembled in this chapter converges on a single insight: sustainable competitive advantage in the AI era depends not on adoption speed, but on alignment—between innovation and ethics, efficiency and inclusion, profitability and planetary boundaries. Organizations that pursue AI as a purely technical race may achieve short-term gains, but only those that embed intelligence within governance, culture, and institutional design convert AI into a durable source of advantage.

Core Findings

1. **Efficiency Integrated with ESG Builds Legitimacy:** Productivity detached from ethics is inherently unstable. When operational efficiency is explicitly aligned with ESG objectives, legitimacy compounds economic value rather than constraining it. In this configuration, performance and responsibility reinforce one another, transforming compliance into strategic capital.
2. **Circular and Predictive Systems Reduce Strategic Risk:** Data-driven lifecycle analytics reduce environmental impact while stabilizing operations and supply chains under uncertainty. Predictive and circular systems shift sustainability from mitigation to anticipation, lowering volatility and reinforcing long-term resilience.
3. **Collaborative Ecosystems Accelerate Learning:** Cross-sector partnerships convert rivalry into co-creation, distributing knowledge, experimentation, and resilience across networks. In complex environments, advantage migrates from isolated optimization to coordinated learning across ecosystems.
4. **Governance Is Strategy:** Transparent, auditable AI systems transform accountability into differentiation. In an era of algorithmic opacity, trust becomes the scarcest strategic asset. Governance is no longer a constraint on innovation; it is the architecture that enables scale with legitimacy.

Managerial Implications

For executives and boards, the implications are concrete. Organizations should institutionalize cross-functional governance councils encompassing data ethics, workforce development, environmental metrics, and architectural decisions. Board agendas must explicitly pair return on investment with *return on responsibility*, recognizing that reputational, regulatory, and societal capital increasingly shape long-term performance.

Every AI initiative should undergo a maturity audit assessing not only technical readiness, but also social impact, absorptive capacity, and alignment with organizational purpose. Investment in reskilling—particularly in mid-market firms and public institutions—should be treated as strategic capital expenditure rather than discretionary cost. Workforce readiness, governance coherence, and architectural fit emerge as the true bottlenecks of AI scaling.

Directions for Research

Future scholarship should deepen understanding of three critical frontiers. First, the environmental dividends of hybrid and edge architectures require systematic quantification across sectors and geographies. Second, the rise of agentic AI systems calls for empirical investigation into how autonomy reshapes decision quality, accountability, and organizational control. Third, cross-cultural governance models are needed to explain how emerging economies can build competitive AI capabilities under asymmetric institutional and infrastructural conditions.

Longitudinal studies tracking organizational progression along maturity trajectories would further illuminate how leadership styles, governance practices, and absorptive capacity co-evolve as ethics become institutionalized rather than episodic.

Broader Vision

AI's ultimate promise is not substitution, but the expansion of human agency. When intelligence and empathy advance together, organizations evolve from extractive

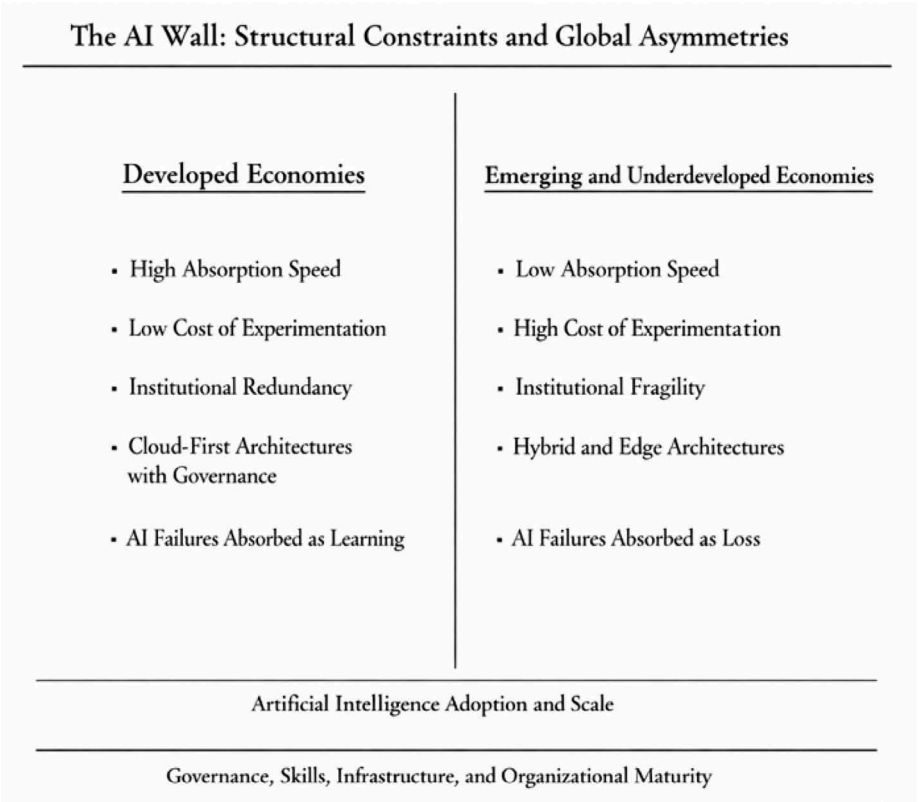


FIGURE 4.9 The DCC4B + AI framework
Source: Authors' own elaboration, informed by IMF (2025). The global impact of AI: Mind the gap. *IMF Working Paper No. 25/76*. International Monetary Fund.

systems toward regenerative ones. Firms that successfully climb the AI Wall—embedding intelligence in structure, ethics in algorithms, and inclusion in culture—will define a new era of purposeful capitalism.

Leadership, in this context, is not domination through data, but stewardship through discernment. Artificial intelligence becomes a mirror reflecting humanity's collective capacity for learning, governance, and renewal. Progress, therefore, is not guaranteed by technology itself, but by the wisdom with which it is designed, absorbed, and governed.

Figure 4.9 presents the DCC4B + AI framework as an institutional architecture for sustainable competitive advantage. The model positions artificial intelligence as an orchestration layer that synchronizes digital efficiency, circular sustainability, and collaborative ecosystems, rather than as an isolated source of value. By anchoring these pillars in governance understood as an operating system—integrating ethics, accountability, and ESG principles—the framework emphasizes alignment over optimization. Sustainable advantage, therefore, emerges not from maximizing individual technological capabilities, but from designing organizations in which intelligence, responsibility, and collaboration reinforce one another over time.

References

Accenture, Microsoft, & Instituto Brasileiro de Governança Corporativa (IBGC). (2024). Artificial intelligence and board governance: From awareness to oversight. https://www.ibgc.org.br.

Biondi, Y., & Cernev, T. (2023). Algorithmic governance, public interest, and legitimacy: Lessons from AI deployment in public security. *Journal of Business Ethics*, 184(2), 417–435. doi:10.1007/s10551-022-05187-6.

Bloomberg. (2025). Bloomberg sustainability index methodology. https://www.bloomberg.com.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Buyl, T., & De Bie, E. (2024). Algorithmic fairness and organizational responsibility: Why ethics cannot be automated. *Organization Science*, 35(1), 112–130. doi:10.1287/orsc.2023.1682.

Chakravorti, B. (2024). *The slow death of digital transformation: How AI, geopolitics, and sustainability reshape strategy*. Harvard Business Review Press.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152. doi:10.2307/2393553.

Contxto. (2022). How Mercado Libre uses AI to fight fraud at scale. https://contxto.com.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.

Dell'Acqua, F., et al. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity. *Management Science*. Advance online publication. doi:10.1287/mnsc.2023.4687.

Deloitte. (2023). Scaling AI responsibly: Lessons from global enterprises. https://www.deloitte.com.

Elangovan, R. (2021). Industry 5.0: Human-centric manufacturing and value creation. *Journal of Industrial Information Integration*, 22, 100187. doi:10.1016/j.jii.2021.100187.

Grand View Research. (2024). *AI PC market size, share & trends analysis report*. https://www.grandviewresearch.com.

Grashof, N., & Kopka, A. (2023). Digital platforms and SME competitiveness: Evidence from AI-enabled ecosystems. *Small Business Economics*, 61(2), 579–598. doi:10.1007/s11187-022-00644-7.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The role of artificial intelligence and data network effects for creating user value. *MIS Quarterly*, 45(3), 1239–1271. doi:10.25300/MISQ/2021/16221.

Hoffman, R. (2025). *Impromptu: Amplifying our humanity through AI*. Dalle Press.
Iansiti, M., & Lakhani, K. R. (2020). *Competing in the age of AI*. Harvard Business School Press.
Iansiti, M., & Lakhani, K. R. (2024). *The octopus organization: Designing organizations for continuous adaptation*. Harvard Business Review Press.
Ing, L. Y., & Grossman, G. M. (2022). *Robots and AI: A new economic era*. Routledge.
KPMG. (2025). Future of AI: Governance, value, and trust. https://kpmg.com.
Knowledge@Wharton. (2017). *How Kabbage uses big data to transform small-business lending*. University of Pennsylvania. https://knowledge.wharton.upenn.edu.
Le-Brun, P., & Werner, J. (2025). *The octopus organization: A guide to thriving in a world of continuous transformation*. Wiley.
LXT. (2025). The path to AI maturity. https://www.lxt.ai.
McKinsey & Company. (2024). The state of AI in 2024. https://www.mckinsey.com.
OpenAI. (2025a). Efficient inference and hybrid AI architectures. https://openai.com.
Papadopoulos, T., Gunasekaran, A., Dubey, R., & Fosso Wamba, S. (2022). Big data analytics, dynamic capabilities, and firm performance. *Journal of Business Research*, 141, 152–166. doi:10.1016/j.jbusres.2021.11.056.
Pulapaka, S., Godavarthi, S., & Ding, S. (2024). *Empowering the public sector with generative AI*. Apress. doi:10.1007/978-1-4842-9797-4.
PwC. (2024). AI skills gap in agribusiness and emerging markets. https://www.pwc.com.
Sachs, J. D. (2015). *The age of sustainable development*. Columbia University Press.
Sætra, H. S. (2023). Artificial intelligence, governance, and sustainability. *AI & Society*, 38(2), 517–531. doi:10.1007/s00146-021-01328-0.
Senseye. (2023). Predictive maintenance impact report. https://www.senseye.io.
Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Knopf.
World Economic Forum. (2025). Global AI governance and workforce readiness report. https://www.weforum.org.
Wu, J., Wang, X., Li, Y., & Chen, Z. (2025). Human–AI collaboration, autonomy, and motivation. *Academy of Management Journal*. Advance online publication. doi:10.5465/amj.2023.0917.

PART II

Building AI Readiness: Frameworks for Assessment and Strategic Adoption



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

5

AI READY OR AI RISKY? UNIFIED FRAMEWORK FOR BUSINESS TRANSFORMATION

Daniela Busuttil Dougall and Pierre Vella

1 Introduction

AI is reshaping how organisations operate, make decisions, and pursue innovation. However, the ability to harness its potential depends not only on technology itself but on an organisation's readiness to integrate, mature, and sustain AI responsibly. To avoid conceptual ambiguity, this chapter clarifies three foundational constructs that underpin the analysis: AI readiness, AI maturity, and AI mindset. These terms are often used interchangeably in the literature, but in practice, they capture different, though interdependent, dimensions of an organisation's transformation journey.

To provide conceptual clarity at the outset, Box 5.1 summarises the three key constructs used throughout this chapter: AI readiness, AI maturity, and AI mindset, each of which serves as a foundation for later discussion.

BOX 5.1 KEY CONCEPTS USED IN THIS CHAPTER

AI readiness: The degree to which an organisation demonstrates capacity, i.e., its technical, structural, and resource preparedness, and commitment, i.e., its leadership support, culture, and willingness to change, to adopt and integrate AI. Readiness reflects potential to act. (Weiner, 2009; Jöhnk et al., 2021; Butler et al., 2021).

AI maturity: the extent to which adoption has been operationalised and embedded within organisational processes, governance, and learning mechanisms. It represents realised capability, the transition from experimentation to sustained and governed integration (Sadiq et al., 2021; Deloitte, 2024).

AI mindset: the collective orientation and cognitive readiness of organisational actors. It encompasses curiosity, trust, ethical awareness, and adaptability, influencing how readiness translates into maturity and how individuals interpret AI's role in their work (Afroogh et al., 2024; Lee & See, 2004; Ryan, 2020).

DOI: 10.4324/9781003632429-8
This chapter has been made available under a CC-BY-NC-ND 4.0 license

Note: These constructs form the analytical spine of the discussion that follows. Readiness expresses potential and intention, maturity reflects embedded capability, and mindset represents the behavioural and cultural catalyst linking the two.

The framework developed later in the chapter builds on this triad, showing how imbalances between them often explain the uneven progress of AI adoption across industries.

2 Conceptual Foundations

Building on the definitions introduced earlier, this section situates AI readiness, maturity, and mindset within broader organisational and theoretical contexts. Rather than treating readiness as a checklist of technical capabilities, it examines how these constructs interact to shape responsible, sustainable AI adoption.

Following Weiner's (2009) model, readiness combines capacity: the organisational resources, skills, and infrastructure that enable change, with commitment: the collective willingness and motivation to act. These dual dimensions underpin how organisations translate intent into practice and sustain change once it begins.

Readiness thus connects to maturity, which captures the institutionalisation of capability, and to mindset, the cultural and behavioural catalyst that sustains adoption over time. AI differs from earlier digital technologies because its adaptive, often opaque mechanisms challenge traditional governance boundaries. These characteristics introduce new ethical and epistemic uncertainties that require organisations to assess readiness in both technical and moral terms (Floridi et al., 2018; Vakkuri et al., 2021).

2.1 Defining AI and Context-Dependent Risk

Understanding AI readiness requires recognising that context defines both capability and responsibility. Unlike traditional digital tools, AI operates through adaptive, data-driven mechanisms that challenge human oversight and established governance boundaries. These characteristics generate ethical and epistemic uncertainties that complicate organisations' preparations for adoption (Floridi et al., 2018; Vakkuri et al., 2021).

Readiness for AI, therefore, cannot be understood in purely technical terms. It involves ensuring that decision-making remains transparent, accountable, and justifiable in systems that can learn and evolve independently. The variability of AI's risk profile means that readiness standards differ sharply between industries.

In low-risk domains such as logistics, customer support, or manufacturing, readiness often centres on efficiency, data integration, and cost optimisation (Jöhnk et al., 2021; Yen et al., 2024). By contrast, high-risk environments, including healthcare, finance, and law enforcement, demand additional layers of governance. Here, fairness assurance, explainability, and continuous human oversight are essential to maintaining ethical and legal integrity (Floridi et al., 2018; Vakkuri et al., 2021).

This differentiation underscores that readiness is context-bound. What may qualify as "ready" in a low-risk domain could be ethically insufficient or even dangerous in a high-stakes setting. Organisational preparedness must therefore combine technical

capacity with mechanisms of ethical accountability. Readiness for AI sits at this intersection, where explainability, fairness, and trustworthiness are as decisive as capability (Floridi et al., 2018; Vakkuri et al., 2021; Vorvoreanu et al., 2023).

2.1.1 Readiness as Capacity and Commitment

In AI contexts, Weiner's (2009) readiness model, built on the balance between capacity and commitment, takes on new significance. Capacity captures the tangible enablers of change: data infrastructure, funding, and technical expertise. Commitment represents the collective will to act, encompassing leadership vision and cultural motivation.

AI exposes how easily these dimensions can drift apart. High capacity without commitment results in underused systems and wasted investment, while strong commitment without capacity leads to aspirational projects that fail to deliver. Readiness, therefore, represents an equilibrium between resources and motivation, where leadership enthusiasm, technical competence, and ethical safeguards advance together.

When this balance breaks, organisations often fall into symbolic adoption, possessing technology without transformation (Weiner, 2009; Sadiq et al., 2021).

2.1.2 Readiness as Dynamic and Contextual

Readiness is not a fixed state but a fluid and evolving condition shaped by environmental change, organisational learning, and technological advancement. Early models of readiness often assumed a linear progression from awareness to implementation, yet experience shows that organisations rarely advance in predictable sequences. Instead, readiness fluctuates as new information, regulations, or stakeholder expectations emerge, requiring continual reassessment and adaptation (Weiner, 2009; Sadiq et al., 2021).

In the context of artificial intelligence, this dynamism becomes more pronounced. AI systems are iterative by design: they learn from data, evolve through feedback, and can generate unforeseen consequences once deployed. Organisational readiness must therefore mirror this adaptability by embedding mechanisms for monitoring, recalibration, and ethical reflection. As Jöhnk et al. (2021) and Butler et al. (2021) observe, readiness for AI cannot be divorced from its context of use; the same capability may represent innovation in one sector and risk in another.

Context also determines the drivers and barriers that shape readiness. Regulatory stringency, data availability, and cultural attitudes toward automation all influence how quickly organisations can adopt and internalise AI. Frameworks such as Deloitte's (2024) AI Readiness & Management Framework (aiRMF) highlight that readiness is distributed across organisational layers: strategic, technical, and cultural factors evolve at different rates, producing temporary asymmetries that require alignment.

The contextual nature of readiness implies that it should be understood as a process rather than an event. Progress in one dimension may trigger new dependencies or reveal ethical tensions in others, making readiness a moving target. Adaptive governance, feedback loops, and cross-functional collaboration are therefore essential to sustain equilibrium between technological capability and organisational responsibility

(Vakkuri et al., 2021). Readiness, in this sense, is a dynamic interplay of structure, behaviour, and context that defines the organisation's capacity to adopt AI responsibly and evolve with it over time.

2.2 Technology Adoption Models

Understanding AI readiness requires linking technical capability with behavioural intent. Technology-adoption models explain why individuals or organisations embrace or resist new technologies and how these attitudes evolve into sustained use. These frameworks form the behavioural foundation of AI readiness, showing that successful adoption depends not only on infrastructure but also on perceived usefulness, trust, and social influence (Davis, 1989; Ajzen, 1991; Venkatesh et al., 2003).

2.2.1 The Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) proposes that two factors, perceived usefulness and perceived ease of use, shape technology acceptance (Davis, 1989). Perceived usefulness reflects the degree to which users believe a system enhances their performance, while perceived ease of use captures the effort required to learn or operate it. When both perceptions are positive, the likelihood of adoption increases. TAM's simplicity made it one of the most influential models for predicting IT acceptance, yet it focuses narrowly on individual cognition and overlooks cultural, ethical, and organisational influences.

2.2.2 The Unified Theory of Acceptance and Use of Technology (UTAUT)

The Unified Theory of Acceptance and Use of Technology (UTAUT) integrated elements from eight prior models, including TAM, the Theory of Planned Behaviour (TPB), and the Innovation Diffusion Theory (IDT). It introduced four determinants of intention and use: performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2003). Later extensions, such as UTAUT2, added hedonic motivation and habit, reflecting more nuanced psychological drivers (Venkatesh et al., 2012). While UTAUT expanded explanatory power, it still treats adoption as a relatively static event rather than a dynamic, evolving process.

2.2.3 Innovation Diffusion Theory (IDT)

Rogers' (2003) Innovation Diffusion Theory (IDT) shifts focus from individual perception to the social mechanisms of diffusion. It explains how innovations spread through communication channels across time and social systems, influenced by relative advantage, compatibility, complexity, trialability, and observability. IDT highlights the importance of peer effects and early adopters but, like TAM and UTAUT, assumes a relatively stable technology environment. In AI, where capabilities evolve rapidly, diffusion cannot rely solely on early adopters; it requires iterative learning and feedback loops within organisations.

2.2.4 The Theory of Planned Behaviour (TPB)

Ajzen's (1991) Theory of Planned Behaviour (TPB) adds psychological precision by linking intention to behaviour through attitude, subjective norms, and perceived behavioural control. TPB helps explain resistance to change and the role of perceived competence or control in adoption decisions. When individuals believe they can use AI effectively and that doing so aligns with organisational norms, their intention strengthens. However, TPB assumes rational decision-making and may underestimate the influence of emotional, ethical, and trust-based factors on AI acceptance.

2.2.5 Trust-Enhanced Technology Models

Because AI introduces opacity and ethical uncertainty, researchers have adapted these foundational models to include trust as a mediating or moderating factor. Trust-enhanced versions of TAM (T-TAM) or UTAUT explicitly recognise that explainability, fairness, and accountability shape perceived usefulness and behavioural intention (Vorvoreanu et al., 2023; Afroogh et al., 2024). In these models, trust bridges the gap between technical reliability and user acceptance: when systems are transparent and decisions are explainable, adoption increases even in the face of uncertainty.

2.2.6 Integrating Behavioural and Organisational Perspectives

While adoption models were initially designed for individual behaviour, their logic extends to organisations as well. Readiness operates at a collective scale, where capacity and commitment interact with perception and trust. The models reviewed here show that intention to adopt technology depends on beliefs about performance, social legitimacy, and control. For AI, those beliefs are reframed by issues of fairness, accountability, and risk.

2.2.7 Practical Implications of Technology-Adoption Models for AI

Although models such as TAM, UTAUT, TPB, and IDT explain how beliefs and attitudes influence adoption, they overlook the governance and ethical complexities that define AI. Traditional adoption frameworks assume that technology performance and user perception are the main drivers of success. In practice, AI adoption often fails not because users doubt its usefulness but because organisational structures fail to manage risk, accountability, and fairness. These behavioural models, therefore, require augmentation through governance and ethical-readiness dimensions to remain relevant for AI-driven environments.

The limitations of traditional adoption theories become clearer when viewed through practice. The following case scenario demonstrates how an organisation satisfied the behavioural conditions of technology acceptance but failed to achieve responsible adoption. It highlights why governance and ethical oversight must complement user acceptance in AI readiness.

BOX 5.2 CASE SCENARIO: GOVERNANCE OVERSIGHT IN AI DEPLOYMENT

A financial services firm introduced an AI-based client risk-scoring tool developed by an external vendor. Initial trials showed strong accuracy, and staff reported high perceived usefulness and ease of use, precisely the predictors emphasised in TAM and UTAUT. However, six months after launch, the regulator raised concerns that the model disproportionately flagged clients from specific demographic groups. No internal governance mechanism existed to audit vendor algorithms or explain decision logic. Despite high user acceptance, the system had to be withdrawn, resulting in reputational damage and financial losses.

This example exposes the limitation of classical adoption models: they capture intention and use, but not responsible adoption. Governance, fairness testing, and ethical accountability must therefore accompany behavioural acceptance if AI readiness is to be achieved.

Note: Technology-adoption models explain how beliefs, norms, and trust influence the willingness to engage with new systems. Each framework: TAM, UTAUT, IDT, TPB, and their trust-enhanced extensions, contributes to understanding adoption but remains limited when applied to AI's evolving, ethically complex context.

The following section builds on these behavioural insights to explore how readiness and maturity interact as organisations institutionalise AI capability.

2.3 Organisational Dimensions of Readiness

The organisational dimension of AI readiness sits at the intersection of structure, culture, and governance. It determines whether technological potential is converted into sustainable capability. Drawing on both the literature and empirical insights gathered from interviews with AI suppliers, adopters, and regulatory actors, this section explores how these dimensions operate in practice and where asymmetries emerge between strategic intent and implementation.

Organisational readiness depends on four interdependent foundations: strategic alignment, resource readiness, knowledge readiness, and data readiness. Strategic alignment ensures that AI initiatives reinforce organisational objectives rather than distract from them. Resource readiness covers the financial, technical, and human capacity required to sustain AI use. Knowledge readiness involves equipping employees with the skills to understand, interpret, and responsibly apply AI outcomes. Data readiness reflects the quality, accessibility, and governance of information assets. Across interviews, participants consistently identified knowledge and data readiness as the weakest dimensions, with organisations often prioritising tools over understanding.

Beyond these structural enablers, readiness reflects how well resources and systems align with leadership vision and cultural openness to change. In many organisations, technical adoption progresses faster than behavioural or ethical adaptation, creating imbalances that undermine long-term maturity (Jöhnk et al., 2021; Sadiq et al., 2021).

Leadership commitment, clarity of governance, and employee empowerment emerged as recurring enablers, while siloed communication, fragmented accountability, and compliance-driven cultures were cited as common barriers.

Culture and leadership are therefore central. Transformative leadership enables alignment between technology strategy and organisational values, while culture provides the psychological safety required for experimentation and responsible risk-taking (Ryan, 2020). Without these foundations, readiness deteriorates into procedural compliance: AI becomes a performative exercise rather than an adaptive capability.

The following case examples from the empirical data illustrate how these dynamics manifest in practice across different sectors.

BOX 5.3 CASE EXAMPLES: GOVERNANCE AND MINDSET AS THE MISSING LINKS

Case A: The Beverage Manufacturer/Importer. A leading beverage company integrated AI tools, including Salesforce Einstein and cybersecurity analytics, through external vendors. Leadership support was conditional and transactional: AI was seen as a reputational marker rather than a transformative enabler. Ethical and regulatory awareness was minimal; responsibility for fairness, bias, and compliance was implicitly outsourced to suppliers. "We pay for the service; the vendor owns the risk," noted the Head of IT. While systems performed technically, adoption stalled due to resistance from sales teams, the absence of governance, and a lack of ownership. The organisation achieved symbolic readiness, i.e., technology without transformation.

Case B: The Airline Operator. An aviation firm exhibited the opposite imbalance. Governance and compliance were deeply ingrained, but strategic leadership treated AI as a distant priority. With legacy systems still being modernised, AI remained peripheral and uncoordinated. Although there was a strong awareness of risk and a commitment to human oversight, "humans should be involved in all stages", there was no central accountability for AI. As a result, innovation occurred in silos, and readiness was procedural rather than strategic.

Case C: The Supplier's Perspective. The AI vendor interviewed in this study confirmed that such asymmetries are common. As explained, "Clients expect explainability and control, but they rarely invest in the governance that enables it." Suppliers noted that many adopters equate digital infrastructure with readiness and treat vendor compliance labels as guarantees of safety. This, they cautioned, leads to a form of "audit theatre", i.e., visible governance without substance, where comfort is drawn from technology rather than robust oversight. The supplier stressed that genuine assurance depends on human systems that question, monitor, and continuously recalibrate the technology. Across sectors, technical capacity often outpaces organisational maturity, leaving adoption vulnerable to cultural inertia and misplaced trust.

Note: These examples reveal why conventional adoption theories, centred on user perception and behavioural intent, are inadequate for AI. Readiness for AI requires integrating behavioural insight with governance design, ensuring that capacity, commitment, and ethical responsibility develop together rather than in isolation.

The cases reveal a consistent pattern: AI readiness falters when organisational learning and ethical governance evolve unevenly. Effective readiness depends on synchronising leadership intent, cultural adaptability, and accountability structures. When these dimensions align, maturity and mindset reinforce one another; when they diverge, readiness becomes symbolic and fragile. These findings carry clear practical implications: organisations must embed strategic alignment, resource management, and preparedness for knowledge and data into governance structures to support responsible AI adoption.

These findings also reinforce the rationale for the unified AI-RMM framework introduced earlier in the chapter, which positions mindset and maturity as transversal levers of adoption and highlights governance as the structural mechanism that sustains their alignment. In doing so, it provides a bridge between readiness theory and the organisational conditions that determine how adoption unfolds in practice.

The following section examines how trust and fairness serve as the behavioural foundations of sustainable AI readiness.

2.4 Trust and Fairness in AI Readiness

Trust and fairness lie at the behavioural core of AI readiness. While infrastructure and governance provide the structural conditions for adoption, trust determines whether systems are genuinely accepted, and fairness shapes their perceived legitimacy. Both are relational rather than technical: they depend on how individuals interpret the transparency, accountability, and intentions embedded in AI systems (Floridi et al., 2018; Ryan, 2020).

Trust functions as both an enabler and a constraint. High trust accelerates adoption by lowering perceived risk, whereas its absence can halt implementation even when performance is strong. In AI, this tension is magnified by opacity: users often cannot fully understand or verify the logic behind automated decisions (Vorvoreanu et al., 2023). Trust thus becomes an organisational design challenge, requiring structures that make systems explainable and decisions auditable.

Fairness complements trust by shaping perceptions of justice and procedural integrity. When outcomes are perceived as biased or opaque, trust deteriorates rapidly regardless of accuracy. Fairness, therefore, requires active management, systematic bias detection, representative data, and transparent decision pathways (Afroogh et al., 2024). It also operates socially: users judge fairness not only by outcomes but by whether they feel included in how those outcomes are produced and justified.

Trust and fairness reinforce each other. Trust arises when systems behave fairly and transparently, while fairness is recognised only when outcomes are communicated clearly and decision processes are open to scrutiny. Together, they form the moral dimension of readiness: without them, technical success cannot translate into behavioural or societal acceptance (Vorvoreanu et al., 2023).

Sustaining these conditions requires deliberate governance. Organisations must integrate fairness testing, ethical review, and explainability into design and deployment, supported by leadership that models accountability and clear communication (Ryan, 2020). Trust calibration, i.e., the ability to adjust confidence based on system performance and context (Lee & See, 2004), emerges as a critical readiness skill. It

allows organisations to maintain equilibrium between vigilance and reliance, ensuring that neither blind trust nor excessive scepticism undermines responsible adoption.

2.5 Adaptability and VUCA Environments

Adaptability has emerged as a defining element of AI readiness, reflecting the need for organisations to operate effectively in environments characterised by volatility, uncertainty, complexity, and ambiguity (VUCA). Unlike other digital technologies, AI evolves rapidly and often unpredictably, requiring organisations not only to reach a readiness threshold but also to recalibrate as new challenges and opportunities emerge continually (Weiner et al., 2009; Felemban et al., 2024).

The literature stresses that adaptability involves both structural and cultural dimensions. Structurally, organisations must establish feedback mechanisms, monitoring processes, and governance frameworks that enable them to evaluate and adjust AI systems in real time (Sadiq et al., 2021; Deloitte, 2024). Culturally, adaptability depends on an openness to experimentation, a tolerance for uncertainty, and a willingness to learn from failure. Without these conditions, organisations risk falling into the "pilot trap," where isolated experiments fail to translate into scalable or sustainable adoption (Vakkuri et al., 2021).

McGee (2024) reinforces this view by positioning resilience and reliability as inseparable from an organisation's security posture. AI systems that lack resilience to disruption or safeguards against technical failure are ill-suited to VUCA conditions, where uncertainty is the norm. Deloitte (2024) similarly notes that security and robustness provide the baseline for organisational trustworthiness, enabling adaptability in high-risk environments.

A volatile and uncertain context further complicates AI adoption by amplifying external dependencies. Rapid regulatory change, shifting societal expectations, and evolving competitive pressures necessitate that organisations adapt not only internally but also within a broader ecosystem of partners, regulators, and stakeholders (Vorvoreanu et al., 2023). This highlights that adaptability is not solely a technical or organisational property but a relational one, shaped by interdependence across the broader AI landscape.

2.6 Summary of Section 2

Section 2 established the conceptual foundations of AI readiness, integrating theoretical, behavioural, and contextual perspectives. It defined readiness as a dynamic capability shaped by organisational alignment, ethical governance, and adaptability to changing environments. By examining technology-adoption theories, organisational dimensions, and behavioural constructs such as trust, fairness, and mindset, the section demonstrated that readiness extends beyond technical capacity to encompass psychological, cultural, and moral preparedness. The discussion of VUCA environments underscored the need to recalibrate readiness as uncertainty and complexity evolve continuously. These insights collectively frame readiness as an adaptive equilibrium between capability and responsibility, forming the theoretical basis for the readiness models analysed in Section 3.

3 AI Readiness Models

AI readiness models provide structured approaches to assessing how prepared organisations are to adopt, govern, and sustain AI. They build on earlier digital transformation frameworks but expand their scope to include ethical, social, and governance dimensions unique to AI. Readiness is no longer measured only by infrastructure or resources; it now reflects an organisation's ability to integrate learning, accountability, and adaptability into its systems and culture (Sadiq et al., 2021).

The literature distinguishes between capability-based models, which evaluate tangible organisational competencies, and behavioural or cultural models, which assess attitudes, values, and leadership dynamics. Both perspectives converge on the principle that readiness must extend beyond technical deployment to include the moral and procedural infrastructure required for responsible adoption. These frameworks help identify not only where an organisation stands but also the pathways through which it can mature over time.

Unlike static assessments, AI readiness models are increasingly iterative. They emphasise feedback loops, post-deployment learning, and continuous recalibration as key features of maturity (Jöhnk et al., 2021). This shift reflects a growing consensus that readiness is not a one-time achievement but an evolving capability shaped by contextual, ethical, and regulatory forces.

3.1 AI-Specific Readiness Frameworks

Readiness models emerged as a response to the limitations of digital transformation frameworks that treated technological capacity as the primary success factor. In contrast, AI readiness models recognise that adoption depends equally on human, cultural, and governance readiness. They typically evaluate multiple dimensions, including strategy, technology, people, and ethics, with each dimension influencing the others through feedback mechanisms (Butler et al., 2021).

These models often adopt a layered structure. At the base lie foundational capacities such as data infrastructure, skills, and leadership commitment. Intermediate layers evaluate governance, organisational learning, and communication, while the upper layers assess adaptability and trust. This tiered approach aligns technical and ethical aspects of AI within a single evaluative system, reflecting the growing emphasis on responsible innovation (Deloitte, 2024).

Scholars and consultancies alike converge on the idea that readiness must combine technical foundations, such as data integrity, interoperability, and skilled personnel, with human and ethical dimensions, including culture, collaboration, and accountability. The interaction between these levels determines whether organisations achieve sustainable AI integration or fall into fragmented, compliance-driven adoption (Sadiq et al., 2021; Deloitte, 2024).

In practice, readiness models vary by focus. Academic models prioritise theoretical completeness, while industry models emphasise pragmatism and scalability. Both, however, agree that readiness requires dynamic reassessment: progress in one area, such as technical infrastructure, can lose meaning if governance or fairness mechanisms lag (Butler et al., 2021).

3.2 Organisational-Level Readiness Models

At the organisational level, AI readiness models aim to assess the preparedness of companies or specific departments, providing a comprehensive view of strengths and weaknesses across multiple domains. These models typically extend beyond technical infrastructure to capture strategy, culture, resources, and governance, highlighting that readiness is multidimensional and cannot be reduced to a single capability (Jöhnk et al., 2021; Sadiq et al., 2021).

Deloitte's aiRMF is one prominent example. It structures readiness into ten capability areas grouped under three functions: Setting the Direction (AI Exploration, AI Strategy and Governance), Building Core Capabilities (Data Readiness, AI Applications and Solutions, AI-Enabled Workforce, AI Infrastructure and Platforms), and Managing AI Holistically (AI Sourcing Management, Trustworthiness, Security and Risk, Customer and User Experience). Notably, the framework positions Trustworthiness, Security and Risk, and Customer and User Experience as cross-cutting enablers, reinforcing that readiness is not siloed but distributed across organisational activities (Deloitte, 2024).

Similarly, Jöhnk et al. (2021) identify 18 readiness factors grouped into strategic alignment, resources, knowledge, culture, and data, supported by 58 indicators. Their contribution highlights that readiness is closely linked to maturity: progress in one dimension often drives improvements in others. Butler et al. (2021) propose the AI Capability Assessment Model (AI-CAM), which assesses business, data, technology, organisation, AI skills, risks, and ethics across five maturity levels, supplemented by an AI Capabilities Matrix that maps competencies to basic, advanced, and expert levels.

While these models provide more comprehensive coverage than early-readiness approaches, they still have limitations. Many are descriptive and static, treating readiness as a one-time state rather than a dynamic process. Few explicitly address the ethical, legal, and regulatory considerations that are increasingly central to AI adoption. This suggests that organisational-level readiness models offer valuable diagnostic tools but must be extended to capture governance, accountability, and continuous adaptability.

3.3 Sector-Specific Readiness Models

A third strand of the literature focuses on sector-specific AI readiness models, developed to address the distinct challenges of adoption in particular industries. These models recognise that readiness cannot be understood in generic terms alone: requirements differ significantly depending on the regulatory environment, data availability, and the social or ethical stakes of the application.

In healthcare, readiness frameworks emphasise data governance, privacy, and accountability, reflecting the high-risk nature of medical decision-making (Vakkuri et al., 2021). Trust, transparency, and human oversight are prioritised to ensure that adoption does not undermine patient safety or professional accountability. In the financial sector, readiness is framed around compliance with evolving regulations, robust risk management, and explainability, since opacity in algorithmic decision-making can have direct legal and reputational consequences (Floridi et al., 2018; Jamil et al., 2025).

TABLE 5.1 Sector-specific readiness conditions across industries

Readiness Dimension	*High-Risk Industries (e.g. Healthcare, Finance)*	*Commoditised/Low-Risk Industries (e.g. Manufacturing, Retail)*	*Literature Sources*
Regulatory and Ethical Oversight	Stringent, multi-layered; requires explainability, accountability, and patient/client consent frameworks.	Light-touch compliance; ethics are often implicit or vendor-managed.	Floridi et al., (2018); Vakkuri et al. (2021); Jamil et al. (2025)
Cultural Readiness	Professional hierarchies reinforce caution and a strong ethos of human accountability.	Pragmatic and efficiency-driven, with a higher openness to automation.	Felemban et al. (2024); Kelm & Johann (2024)
Data Governance	Sensitive, regulated data; interoperability and privacy constraints dominate.	Abundant operational data; fewer privacy limits but lower quality control.	Jöhnk et al., (2021); Deloitte (2024)
Adoption Drivers	Risk mitigation, compliance, and patient/client trust.	Productivity, cost reduction, competitive advantage.	Sadiq et al. (2021); Yen et al. (2024)
Common Readiness Barriers	Fear of liability, explainability gaps, and ethical uncertainty.	Skill shortages, legacy systems, and limited strategic alignment.	Butler et al., (2021); Jamil et al. (2025)
Mindset Levers	Ethical literacy, cross-disciplinary collaboration, and human-in-the-loop assurance.	Innovation mindset, tolerance for experimentation, vendor learning.	Afroogh et al. (2024); PwC (2021)

Source: Authors' own elaboration, based on Vakkuri et al. (2021), Floridi et al. (2018), Jamil et al. (2025), and Vorvoreanu et al. (2023).

Note: Differences highlight how contextual risk and regulation shape both readiness requirements and the evolution of mindset across sectors.

Sectoral approaches also highlight that data readiness and governance are highly context-dependent. For example, the availability of high-quality labelled data may be taken for granted in some industries but is far more limited in domains such as healthcare or public administration. Similarly, cultural readiness can vary among organisations in safety-critical sectors, which may resist automation more strongly, reflecting the weight of professional accountability and ethical responsibility (Vorvoreanu et al., 2023).

The strength of sector-specific models lies in their contextual sensitivity, ensuring that readiness assessments reflect the realities of each industry. Their weakness, however, is that they are often difficult to generalise or compare across sectors, limiting their broader applicability. This reinforces the argument for unified frameworks that can incorporate sector-specific nuances while maintaining a standard structure for assessing AI readiness.

To illustrate these sectoral contrasts more clearly, Table 5.1 summarises how AI readiness conditions diverge across high-risk and commoditised industries. It distils common patterns from the literature, comparing regulatory expectations, cultural drivers, and mindset levers that shape adoption trajectories.

While Table 5.1 synthesises sectoral contrasts from the literature, the following boxed examples draw on interview insights gathered for this study. They illustrate how readiness manifests in practice across two contrasting environments, healthcare and waste management, highlighting how regulatory risk, professional responsibility, and organisational culture shape the journey from readiness to maturity.

BOX 5.4 SECTOR-SPECIFIC READINESS IN PRACTICE: HEALTHCARE AND MANUFACTURING EXAMPLES

Mini-Case A: Healthcare (High-Risk Sector): In healthcare, AI readiness remains anchored in the realities of professional accountability and patient safety. The clinician interviewed for this study, a surgeon, expressed interest in AI's potential to support diagnostics and surgical planning but was unequivocal that "the ultimate responsibility and risk rest with the clinician." This insistence reflects the life-and-death stakes that shape adoption attitudes: while AI can assist decision-making, delegation beyond a certain threshold is ethically untenable. Such caution aligns with literature describing the sector's moral and procedural constraints (Floridi et al., 2018: Vakkuri et al., 2021). Readiness, therefore, manifests as ethical vigilance rather than technological enthusiasm; experimentation is constrained by the need to preserve trust, accountability, and patient welfare. The AI-RMM framework later identifies this as a form of cultural constraint under ethical vigilance, a mindset that safeguards integrity but slows maturity progression.

Mini-Case B: Waste Management (Commoditised Sector): Waste-management organisations operate in a low-risk yet infrastructure-intensive environment. The company interviewed described exploring AI for waste-sorting efficiency and predictive maintenance, but acknowledged that implementation was still aspirational. As one manager noted, "our data are there, but we haven't yet reached the stage where we can use them intelligently". The barrier was not regulatory but organisational: legacy processes and limited internal capacity to interpret AI outputs. This mirrors Jamil et al. (2025), who found that in less-regulated sectors, technical readiness often outpaces strategic alignment. The interview also revealed a pragmatic culture open to experimentation once cost–benefit clarity was achieved, suggesting that mindset, rather than compliance, remains the pivotal lever for maturity.

Note: These vignettes combine insights from the literature and the study's interview data to illustrate how readiness challenges vary across contexts. Healthcare operates under high ethical and existential stakes, where responsibility cannot be delegated; waste management faces operational inertia but cultural openness. Together, they show that readiness maturity depends as much on contextual risk and accountability structures as on mindset evolution, a theme developed further in the unified AI-RMM.

3.4 Limitations of Existing Readiness Models

Despite their contributions, existing AI readiness models face several limitations that restrict their usefulness in guiding adoption.

A first limitation is their tendency toward static or linear assessments. Many frameworks present readiness as a sequence of stages to be achieved in order but, in practice, organisations move fluidly between levels as technologies, regulations, and cultural conditions evolve (Weiner, 2009; Sadiq et al., 2021). Static approaches risk oversimplifying readiness, ignoring the cyclical and adaptive nature of organisational change.

A second limitation is the underrepresentation of governance and ethics. While technical and strategic factors are frequently addressed, ethical considerations such as fairness, transparency, and accountability are often treated as optional or peripheral. This gap undermines the applicability of many models, particularly in high-risk environments where ethical compliance is non-negotiable (Floridi et al., 2018; Thiebes et al., 2020; Vorvoreanu et al., 2023).

Third, many readiness models lack actionable guidance. They diagnose challenges but do not provide clear pathways for operationalisation. As Vakkuri et al. (2021) argue, organisations are often told what responsible AI should look like but not how to implement it. This disconnect contributes to the persistent gap between principles and practice observed across sectors.

Finally, sector-specific models, while valuable for contextual insight, are often difficult to generalise. They highlight unique risks and requirements in industries such as healthcare or finance but do not easily transfer to other contexts, limiting their scalability and comparative utility (Jamil et al., 2025; Vakkuri et al., 2021).

Taken together, these limitations suggest that while readiness models provide practical diagnostic frameworks, they require further development to be fully effective. Future approaches must account for adaptability, embed governance and ethics, and provide practical tools for operationalisation.

3.5 Summary of Section 3

The literature on AI readiness models reveals significant progress in understanding how organisations prepare for adoption, moving from generic digital-readiness approaches to AI-specific, organisational, and sectoral frameworks. These models collectively emphasise that readiness is multidimensional, shaped by strategy, governance, culture, and resources, yet often remain static and diagnostic rather than adaptive. Sector-specific frameworks add contextual depth but lack transferability across industries. The limitations identified, particularly the absence of dynamic feedback loops, ethical integration, and actionable guidance, point to the need for complementary models that capture progression and accountability. Section 4, therefore, turns to AI maturity models, which extend readiness by tracing the development of capabilities and ethical integration over time.

4 AI Maturity Models

While readiness models focus on assessing whether organisations are prepared to adopt AI, maturity models examine how capabilities evolve once adoption is underway. They measure not only initial preparedness but also the progression and sophistication of organisational practices over time. Maturity models, therefore, complement

readiness frameworks: readiness identifies the starting point, while maturity traces the trajectory of development.

AI maturity models vary widely in scope and focus. Some emphasise technical capacity, such as data quality, infrastructure, and applications, while others adopt broader perspectives that include culture, governance, and ethics. Their strength lies in providing structured pathways that allow organisations to benchmark their progress, identify areas for improvement, and align adoption with long-term strategic goals (Sadiq et al., 2021).

However, as with readiness models, maturity frameworks face criticisms. Many are descriptive and static, overlooking the dynamic nature of AI adoption and the role of governance and ethics in sustaining trust (Vorvoreanu et al., 2023). This section examines different strands of AI maturity models, ranging from general organisational-level approaches to frameworks explicitly integrating ethical considerations. The analysis highlights both their contributions and their limitations, setting the stage for the unified framework developed later in this chapter.

4.1 Organisational-Level Maturity Models

Building on the organisational-level readiness models discussed in Section 3.2, maturity frameworks extend the analysis beyond preparedness to examine how those same capabilities evolve after adoption. While readiness focuses on establishing the structural and cultural preconditions for change, maturity reflects how these foundations develop through feedback, learning, and the integration of governance. In this sense, maturity represents a dynamic extension of readiness rather than a separate construct.

Organisational-level AI maturity models assess the capabilities of entire companies or departments, tracking how practices evolve across people, processes, data, and technology (Sadiq et al., 2021). Their purpose is not to determine whether organisations are ready to begin adoption but to measure the progression and sophistication of capabilities over time.

Sadiq et al. (2021) identify two persistent gaps: limited scope, with governance and ethics often missing, and a practical implementation gap, where models remain largely theoretical. Some frameworks have begun to address these issues, notably PwC's Responsible AI framework (discussed in Section 4.2), which operationalises trustworthiness through interpretability, robustness, and fairness.

The key dimensions of maturity build on the readiness factors identified by Jöhnk et al. (2021) (highlighted in Section 3.2), including strategy, resources, knowledge, culture, and data, but shift their focus from establishment to evolution. In maturity, the question is no longer whether these dimensions exist, but how they interact, reinforce one another, and adapt through continuous feedback and ethical calibration.

- Strategic alignment: Maturity depends on the ongoing reinforcement of strategic alignment through measurable outcomes and leadership accountability (Jöhnk et al., 2021). Tools such as AI heat maps can extend this perspective by visualising strengths and gaps in organisational capabilities across functions (Sadiq et al., 2021).
- Resources: Maturity is achieved when resource allocation becomes self-sustaining, when investment and capability development follow demonstrated value rather

than experimentation. Organisations remain disadvantaged by limited budgets and legacy systems (Jöhnk et al., 2021; Yen et al., 2024; Jamil et al., 2025).
- Knowledge: Readiness establishes awareness, while maturity embeds continuous learning. Ethical and technical literacy become part of everyday decision-making rather than being confined to one-off training efforts (Jöhnk et al., 2021; Yen et al., 2024).
- Culture: Maturity requires normalising innovation, transforming initial curiosity or risk aversion into resilience and shared responsibility for outcomes (Jöhnk et al., 2021; Felemban et al., 2024; Kelm & Johann, 2024).
- Data readiness: Data maturity is achieved when governance, quality, and interoperability evolve into proactive stewardship, ensuring that data not only supports technical accuracy but also fairness and accountability (Deloitte, 2024; Jöhnk et al., 2021).

Frameworks such as Deloitte's aiRMF and Butler's AI-CAM, introduced in Section 3.2, have since been adapted to capture post-adoption learning and accountability. Their contribution lies in tracing the progression from initial alignment and capability building to embedded governance, risk control, and continuous improvement.

Despite their breadth, many organisations remain in early maturity stages, often limited to isolated pilots (Jöhnk et al., 2021). Self-assessments frequently overstate maturity, creating a gap between aspiration and practice (Vorvoreanu et al., 2023). Addressing this requires embedding governance, accountability, and iterative learning, a theme explored further in ethics-integrated maturity frameworks.

4.2 Ethics-Integrated Maturity Frameworks

The growing recognition of AI's ethical, legal, and societal implications has shifted maturity assessments beyond technical capacity toward frameworks that explicitly integrate responsible and ethical principles. Concepts such as responsible AI (RAI), ethical AI, and beneficial AI underscore that the value of AI hinges on its trustworthiness, fairness, and alignment with societal values (Thiebes et al., 2020).

Ethics-integrated maturity models, therefore, aim to ensure that AI systems are developed, deployed, and governed in ways that maximise benefits while mitigating risks (Thiebes et al., 2020; Vakkuri et al., 2021).

These models typically embed principles such as:

- Transparency or explainability (Thiebes et al., 2020; Vakkuri et al., 2021; Vorvoreanu et al., 2023).
- Fairness and justice, which stress the need to prevent bias (Floridi et al., 2018; Jöhnk et al., 2021).
- Non-maleficence and beneficence, which focus on harm reduction and societal well-being (Thiebes et al., 2020).
- Accountability, which requires oversight and redress; and autonomy, which safeguards human agency (Floridi et al., 2018).

Applied models demonstrate how these principles can be operationalised. Microsoft's Responsible AI Maturity Model (RAI MM) categorises 24 dimensions into three key

areas: Organisational Foundations (leadership, governance, compliance, culture), Team Approach (cross-disciplinary collaboration, socio-technical design), and RAI Practice (risk management, privacy, security, transparency). Its key insight is interdependence: progress in one area often relies on advances in others (Vorvoreanu et al., 2023).

PwC's Responsible AI framework offers a parallel approach, emphasising a values-driven and technology-enabled path to governance. Structured around Strategy (ethics and regulatory compliance), Control (risk management, governance structures, compliance), and Performance and Security (fairness, interpretability, robustness, privacy, safety, security), it moves from principles to practice (PwC, 2021).

Other initiatives reinforce this trajectory. Reuel et al. (2025, as cited in Vorvoreanu et al., 2023) distinguish between organisational maturity (encompassing leadership, governance, and training) and system maturity (encompassing privacy, reliability, and fairness). Their survey of 1,000 organisations found a persistent "say–do" gap: formal RAI structures existed but were poorly executed. AI4People's framework consolidated five ethical principles: beneficence, non-maleficence, autonomy, justice, and explicability, into 20 policy recommendations for a "Good AI Society" (Floridi et al., 2018).

Despite their variety, these models share persistent limitations. Many outline what responsible AI should achieve, but not how to implement it (Vakkuri et al., 2021). Ethical criteria often remain vague, and evaluation methods are inconsistent. Academic frameworks can diverge from industry practice, underscoring the need for co-design with practitioners to ensure that models are actionable and contextually relevant (Vorvoreanu et al., 2023).

4.3 Challenges in Implementing Ethical AI

Even with ethics-integrated maturity models, translating principles into practice remains one of the most significant challenges. Studies show a persistent gap between commitments and operationalisation, as organisations struggle to embed RAI into day-to-day workflows (Vorvoreanu et al., 2023). The problem is compounded by the lack of standardised evaluation methods, limited practical guidance, and the ambiguity of ethical criteria (Vakkuri et al., 2021).

Reuel et al.'s (2025, as cited in Vorvoreanu et al., 2023) model illustrates this difficulty: while organisational-level structures for governance, risk management, and training were often established, execution consistently lagged, revealing a gap between planning and practice (Vorvoreanu et al., 2023).

This reflects a broader weakness in academic models, which frequently overlook governance and cultural foundations, including mindset changes, even though they are essential for maturity. PwC's framework, discussed in Section 4.2, illustrates how these gaps can be addressed through clear accountability mechanisms and enterprise-wide risk mitigation (PwC, 2021).

As mentioned earlier, another barrier lies in operationalisation. Frameworks often specify what responsible AI should look like, but they do not outline how to achieve it. Without collaboration between academia and industry, maturity models risk being conceptually robust yet practically detached (Vakkuri et al., 2021). PwC again provides an applied response by embedding end-to-end governance through its nine-step

AI lifecycle, including tools such as Algorithmic Impact Assessments (AIAs) that guide design decisions and document system limitations (PwC, 2021).

These challenges reinforce that maturity is not only a technical achievement but also an organisational and cultural one. The effectiveness of maturity models depends on their capacity to embed governance, ethics, and accountability into tangible practice. This underscores the interdependence of readiness and maturity: organisations with strong readiness foundations are better positioned to progress along maturity pathways in ways that are both sustainable and socially responsible.

4.4 Summary of Section 4

AI maturity models extend the logic of readiness by tracing how organisational capabilities evolve after adoption begins. They reveal that maturity is both technical and behavioural, shaped by governance, ethics, and cultural alignment. While organisational-level and ethics-integrated frameworks such as PwC's Responsible AI and Microsoft's RAI MM provide structured pathways, implementation gaps persist. Many models remain descriptive, lacking practical tools for embedding ethical principles and accountability into day-to-day operations. Addressing these limitations requires a unified approach that links readiness and maturity through adaptive governance and mindset evolution, a synthesis advanced in the AI-RMM framework presented next.

5 Developing the Unified Framework

5.1 Introduction to the Findings

The AI-RMM framework developed in this study synthesises findings from both literature and interviews with AI suppliers, adopters, and regulators. It bridges theoretical models of technology adoption with practical observations drawn from real-world implementation contexts.

The AI-RMM reflects a shift from static readiness checklists to a dynamic system in which maturity and mindset evolve continuously. Feedback loops between technical capacity, governance structures, and cultural learning shape this evolution. The framework, therefore, captures not only how organisations prepare for AI adoption but also how they recalibrate once deployment begins.

The study informing the AI-RMM included participants from diverse sectors: aviation, healthcare, manufacturing, waste management, financial services, and regulation. Their insights ensured the framework captured readiness as both a strategic orientation and an operational practice, rather than a single compliance milestone.

In essence, this section introduces the conceptual and empirical foundations of the AI-RMM. The framework treats readiness as an evolving equilibrium between capability and mindset, providing the foundation for the eight clusters described below.

5.2 The AI-RMM Framework: Eight Interdependent Clusters

The AI-RMM integrates eight interdependent clusters that together define the conditions for responsible and sustainable AI adoption. Each cluster captures a distinct

organisational domain while remaining connected through iterative feedback loops that enable recalibration over time.

The eight clusters are:

1. Strategic and Cultural Readiness – alignment of leadership vision, culture, and change receptiveness.
2. Technical and Operational Readiness – infrastructure, data quality, and integration capability.
3. Ethical and Regulatory Readiness – compliance with external standards and internal ethical governance.
4. Human–AI Interaction Readiness – user competence, interpretability, and trust calibration.
5. Client Preparedness and Change Management – stakeholder communication, training, and transformation support.
6. Post-Deployment Adaptability – monitoring, learning, and continuous improvement mechanisms.
7. Risk and Security Readiness – data protection, resilience, and cyber-governance practices.
8. Governance, Accountability and Assurance – oversight structures that ensure accountability and transparency.

The eight clusters are organised into two interacting tiers that operate in a continuous feedback loop. The inner clusters – Strategic and Cultural Readiness, Technical and Operational Readiness, Risk and Security Readiness, and AI Governance, Accountability and Assurance – represent the organisation's structural foundations. They provide the stability, integrity, and alignment required for responsible AI integration, ensuring that systems, controls, and governance structures are coherent with strategic objectives, data practices, and regulatory obligations.

The outer clusters – Ethical and Regulatory Readiness, Human–AI Interaction Readiness, Client Preparedness and Change Management, and Post-Deployment Adaptability – function as cross-cutting adaptive interfaces that span organisational boundaries. These clusters are permeable and reflexive: they absorb environmental signals, interpret user and stakeholder feedback, and channel insights back into the structural core. Because of this configuration, shifts in the outer clusters often reverberate across the system, prompting recalibration within the foundational domains.

Each cluster is assessed along two interrelated dimensions. Maturity reflects the structural and procedural robustness of organisational practices, while mindset represents the behavioural and cultural openness required to sustain change. These dimensions evolve in tandem: as systems mature, behavioural confidence deepens; as cultures adapt, structural coherence strengthens.

Empirical evidence from both the literature and the interviews indicates that the outer clusters are often the most fragile zones of AI adoption. While inner clusters advance under technical and regulatory mandates, the outer clusters rely on behavioural change, ethical interpretation, and cross-functional collaboration, conditions that develop more unevenly. Many organisations demonstrate strong structural readiness yet falter in translating it into adaptable, fair, and continuously improving

practice. These friction points, examined later in Table 5.3, represent the barriers most commonly impeding the transition from readiness to maturity.

By mapping feedback loops between clusters, the AI-RMM departs from conventional readiness models that treat adoption as a linear or checklist-driven process. It portrays adoption as cyclical and interdependent: ethical revisions reshape cultural trust, technical upgrades reconfigure accountability, and regulatory change triggers renewed learning and adaptation.

Figure 5.1 illustrates this systemic configuration. It depicts how readiness and maturity evolve through continuous feedback loops that connect technical, organisational, and ethical domains. Baseline maturity sits at the centre of the model, representing the minimum threshold for responsible adoption. Demonstrated maturity emerges as feedback and learning accumulate, coupling structural capacity with cultural integration and operational effectiveness.

This arrangement underscores that AI readiness is dynamic rather than sequential, and integrative rather than siloed, enabling organisations to evolve responsibly as internal capacities and external conditions co-evolve.

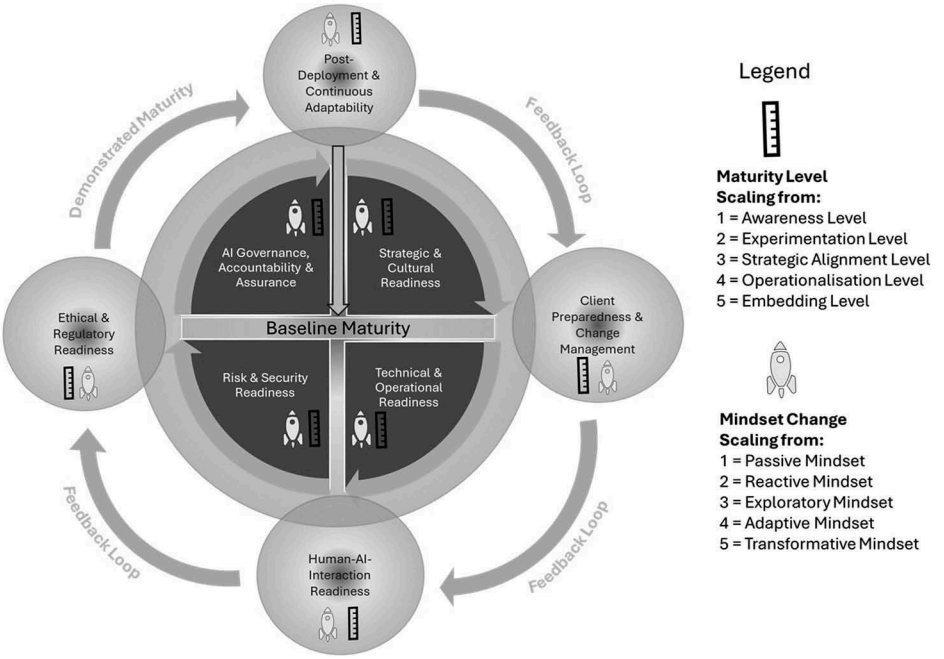


FIGURE 5.1 The unified AI-RMM framework
Note: The model distinguishes inner clusters (darker tones) from outer clusters (lighter tones). Each cluster is evaluated across maturity and mindset, linked through continuous feedback loops between baseline and demonstrated maturity.
Source: Designed by the authors.

In sum, the AI-RMM provides a living architecture for responsible adoption. It visualises readiness and maturity as co-evolving forces, mediated by mindset. The following section applies this conceptual framework to the empirical data that informed its design, revealing asymmetries between maturity and mindset across clusters.

5.3 Cluster Analysis

While Section 5.2 introduced the AI-RMM framework as a conceptual model (see Figure 5.1), this section empirically applies it. The analysis draws from interviews with adopters across multiple sectors, the AI supplier, and a regulator. Their perspectives offer insight into how maturity and mindset are distributed across the eight clusters, as well as how both structural capability and cultural orientation influence adoption.

Figure 5.2 provides a functional view of AI readiness based on interview data. It compares adopters' self-assessments of maturity and mindset with the AI supplier's evaluation of adopters. At the same time, the red dotted line marks the regulatory floor, the minimum compliance threshold set by obligations such as the EU AI Act. Maturity is shown by bars: blue for adopters' ratings and orange for the AI supplier's ratings. Mindset is represented by lines: grey for the adopters' perceptions and yellow for the AI supplier's view of the adopter's mindset. Together, these comparisons illustrate how adopters often overestimate structural maturity while underestimating the mindset shifts required for sustainable integration.

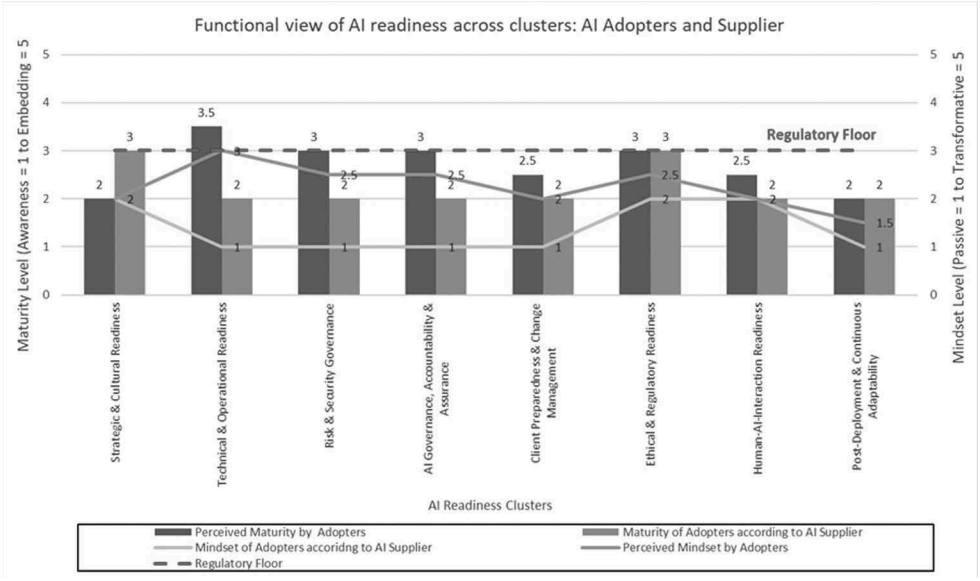


FIGURE 5.2 Functional analysis of AI readiness across clusters
Note: This figure illustrates the operationalisation of AI readiness using interview data. It compares adopters' self-assessments with the AI supplier's evaluations against the regulatory floor, which represents the minimum compliance requirement.
Source: Designed by the authors.

Figure 5.2 highlights three dynamics. First, adopters consistently rated their maturity higher than the AI supplier did, particularly in governance, risk, and post-deployment adaptability. Second, gaps were even wider for mindset: adopters perceived themselves as moderately open and adaptive, while the AI supplier judged their mindset more conservatively, suggesting cultural hesitation and resistance that adopters often underplay. Third, technical and regulatory clusters scored relatively higher, while cultural and post-deployment clusters lagged, leaving adoption fragile.

These patterns frame the cluster-by-cluster analysis that follows, in which each cluster is examined through the dual lenses of maturity and mindset, and insights from the literature are compared with empirical evidence from adopters, the AI supplier, and the regulator. Because mindset operates as a cross-cutting lever, each cluster analysis concludes with a short set of practical actions that translate mindset development into daily practice.

5.3.1 Cluster 1: Strategic and Cultural Readiness

Strategic and cultural readiness captures how leadership, organisational purpose, and collective attitudes enable or constrain AI adoption. The literature consistently identifies strategic alignment as a critical enabler of readiness, ensuring that AI initiatives are embedded within the broader business vision and supported by senior leadership (Jöhnk et al., 2021; Sadiq et al., 2021; Deloitte, 2024).

Without this alignment, adoption risks becoming fragmented, with pilots conducted in isolation from organisational priorities. Across interviews, participants frequently overestimated their preparedness, equating enthusiasm or pilot activity with strategic maturity. Although leadership support was often described as present, it tended to be conditional and transactional rather than transformative. As one participant from the aviation sector explained, "AI comes up in every strategy meeting, but there's no real plan behind it. It's just something we know we'll have to do eventually."

This disconnect between aspiration and execution was evident in organisations that treated AI as a reputational marker rather than a driver of transformation. A manufacturing manager admitted, "We showcase AI tools in reports, but internally nothing has changed; the processes are the same." Such accounts reflect what Jöhnk et al. (2021) describe as symbolic alignment: the appearance of digital ambition without the structural commitment needed to deliver it.

Leadership engagement also emerged as uneven. While some interviewees described executives who acted as active sponsors, most described leadership support that was reactive, focused on short-term returns rather than sustained learning. The AI supplier summarised this tendency: "Leadership wants visible results fast, not a cultural shift that takes years." The literature similarly warns that leadership enthusiasm without long-term sponsorship can stall transformation, leaving AI adoption confined to proof-of-concept stages (Sadiq et al., 2021; Yen et al., 2024).

The cultural dimension reinforces this finding. Curiosity toward AI was common, yet it was accompanied by an equally strong desire for control and predictability. One respondent observed, "People like the idea of AI until it starts changing how they work." This ambivalence mirrors what Felemban et al. (2024) and Kelm and Johann (2024) identify as cultural inertia: a state in which openness to innovation coexists

with deep-seated reluctance to alter routines or hierarchies. PwC's Responsible AI framework (2021) reinforces this point by emphasising that ethics must be embedded in both culture and governance. Principles only gain traction when behaviours that support responsible use are recognised and normalised (PwC, 2021). This cultural translation bridges structural readiness and mindset, showing that maturity cannot advance without cultural transformation.

In sectors where AI applications are not directly subject to high-risk or algorithmic accountability regulation, such as manufacturing, readiness was often defined by vendor reputation rather than internal capability. As one interviewee explained, "We trust the system because we trust the brand; that's where our governance stops." By contrast, in aviation, where regulatory oversight is embedded in the culture, innovation was constrained by procedural caution. Both contexts reveal asymmetrical forms of unreadiness: one rooted in overconfidence, the other in overregulation.

The AI supplier echoed these dynamics, noting that strategic and cultural inertia consistently undermine readiness. One expert commented that "organisations think AI transformation starts with data; it actually starts with people." Another reflected that "trust-building and governance conversations need to happen long before the technology is switched on." Their perspectives reaffirm that readiness depends on cultural maturity as much as technical preparedness.

Taken together, both literature and evidence converge: strategic and cultural readiness remain the most fragile foundation of AI adoption. Leadership support is often symbolic, cultural alignment is uneven, and strategic planning is reactive rather than anticipatory. The gap between rhetorical commitment and operational follow-through indicates that readiness at this level depends less on awareness and more on institutional conviction: the ability of leaders and cultures to move from acknowledging AI's promise to reimagining the organisation around it.

Translating these insights into action requires more than awareness of cultural fragility. It calls for deliberate leadership practices that model openness, psychological safety, and learning. The following levers summarise actionable ways to strengthen cultural and mindset readiness drawn from both literature and empirical findings.

BOX 5.5 PRACTICAL LEVERS FOR STRENGTHENING CULTURE AND MINDSET READINESS

- Model openness from leadership through visible experimentation.
- Communicate AI's purpose in human terms, emphasising augmentation, not replacement.
- Reward learning and adaptation rather than risk avoidance.
- Establish cross-functional teams that combine domain, data, and ethics expertise.
- Create "safe to fail" environments that normalise iterative learning.
- Integrate trust-building activities (explainability sessions, peer reviews) into daily routines.

Note: Embedding such behaviours turns cultural alignment from rhetoric into routine. They anchor mindset as a living process; one that reshapes leadership language, daily interactions, and the organisation's collective comfort with uncertainty.

5.3.2 Cluster 2: Technical and Operational Readiness

Technical and operational readiness represents the extent to which organisations possess the infrastructure, data architecture, and operational capacity to integrate AI effectively. The literature consistently identifies data quality, interoperability and technical scalability as the backbone of readiness (Jöhnk et al., 2021; Deloitte, 2024; Yen et al., 2024). Mature organisations are expected to maintain unified data ecosystems, modular platforms, and processes for continuous validation (Sadiq et al., 2021). Nevertheless, empirical evidence from this study suggests that while digital foundations are often in place, the ability to translate them into adaptive, AI-enabled operations remains limited.

Across participants, technical readiness was frequently described as strong in infrastructure but weak in knowledge. A public-sector respondent observed that "internally, the potential is there, but it depends on structured data … the challenge is that we suffer from a lack of knowledge bases." This distinction between digital maturity and AI capability echoes Jamil et al. (2025), who highlight that organisations often conflate IT modernisation with intelligent automation. Many interviewees equated readiness with the presence of data systems, rather than the human and procedural capabilities required to convert data into intelligence.

Legacy architectures further constrain integration. As one aviation participant noted, "We have solid systems, but we're still migrating from older architectures. We'd rely on suppliers to guide integration; we don't have the people internally who can retrain or validate the model." This dependence illustrates what Deloitte (2024) calls reactive readiness: infrastructure robustness without internal competence to sustain innovation. Similar patterns appeared across manufacturing and public-sector contexts, where integration was possible in principle but hindered by regulatory caution, data fragmentation, or unclear ownership structures.

The supplier perspective corroborated these findings. A representative from an AI vendor explained that "clients assume that because they have modern systems, they're AI-ready. In reality, they're data-rich but insight-poor." Another supplier remarked that technical readiness is often superficial: clients focus on acquiring tools but neglect post-deployment maintenance and learning. Such asymmetry between technical capacity and operational maturity mirrors Yen et al. (2024), who argue that readiness without internal governance produces brittle adoption cycles, efficient at launch but unsustainable over time.

Even highly digital organisations acknowledged the trade-off between compliance and agility. As one aviation engineer reflected, "Our digital infrastructure is robust, but every change has to go through strict validation cycles. That slows down innovation but keeps quality high." This paradox demonstrates that technical readiness alone does not guarantee adaptability; instead, it can entrench procedural inertia when risk management outweighs experimentation.

Beyond governance, data quality itself emerged as a recurring weakness. Several participants described confidence in their datasets despite limited visibility into accuracy or lineage. As one interviewee reflected, "We trust that the data is clean because our vendors say so." Such blind trust undermines both technical and ethical integrity: data that is incomplete, inconsistent, or biased can encode unfairness long before algorithms are deployed. Ensuring high-quality, representative, and well-documented data is therefore not only a technical prerequisite but also a foundation for fairness and accountability.

Taken together, both literature and empirical evidence confirm that technical and operational readiness are necessary but insufficient conditions for AI maturity. Many organisations are digitally enabled yet operationally immature: their infrastructures are robust, but their capacity to sustain, validate, and evolve AI systems remains underdeveloped. The maturity gap lies not in technology but in translation, the ability to convert digital infrastructure into intelligent, adaptive, and self-improving ecosystems.

Technical readiness becomes sustainable only when paired with a mindset of internal ownership. The following levers help translate technical capacity into adaptive operational practice.

BOX 5.6 PRACTICAL LEVERS FOR TECHNICAL STEWARDSHIP MINDSET

- Ensure robust, resilient architecture, including a detailed review of data quality
- Pair engineers with business users to co-own AI performance reviews.
- Encourage in-house retraining on vendor platforms to reduce dependence.
- Document the lessons learned from the integration of cross-functional and team learning.
- Promote accountability for data quality through shared metrics.
- Embed fairness and explainability checks in data preparation and model development to identify and mitigate bias early

Note: These measures reframe technical readiness from tool possession to organisational learning, embedding maturity through stewardship and fairness-aware design.

5.3.3 Cluster 3: Ethical and Regulatory Readiness

Ethical and regulatory readiness concerns how organisations interpret, internalise, and operationalise emerging frameworks for trustworthy AI. The literature converges on the view that compliance is a necessary but insufficient condition for responsible adoption. Floridi et al. (2018) describe AI ethics as a "systemic responsibility" that extends beyond technical safeguards, requiring value alignment and continuous reflection. Vakkuri et al. (2021) similarly argue that ethical governance must be embedded throughout the AI lifecycle, while Thiebes et al. (2020) emphasise explainability and accountability as determinants of sustained trust.

Across cases, awareness of regulation was generally high, yet its practical implementation lagged. A senior public-sector interviewee reflected that "we've already

assessed both the AI Act and GDPR and are carrying out extensive risk assessments. What's missing is explicit guidance, guardrails that make risk identification more accurate, but no one has actually put those guardrails in place." This echoes what Vakkuri et al. (2021) describe as procedural readiness: policy interpretation without operational translation. Ethical frameworks exist on paper, but their enforcement remains reactive and fragmented.

Regulatory authorities confirmed this pattern. A representative from the regulator observed, "We often see strong technical compliance but weak awareness of ethical or societal implications. Compliance is not the same as readiness." The distinction is significant: organisations may satisfy documentation requirements while neglecting the reflective and deliberative processes that underpin fairness and accountability. As Floridi et al. (2018) note, responsible AI demands more than rule following; it requires internalising ethics as an organisational norm.

Suppliers reinforced these findings from the implementation side. An AI vendor explained that "clients talk about responsible AI, but they rarely ask about it during implementation. It's treated as an afterthought, something to document at the end, not design at the start." This perspective aligns with PwC's (2021) Responsible AI Framework, which emphasises that ethics must be built in from the outset through governance, interpretability, and stakeholder inclusion. When ethical considerations are deferred to post-deployment audits, accountability becomes retrospective rather than preventive.

Systematic fairness testing and bias detection should accompany every stage of model development and retraining. Fairness cannot be inferred from technical performance alone; it requires deliberate validation of both datasets and algorithmic logic. Embedding these checks within technical workflows transforms fairness from a principle into a measurable, auditable practice.

Sectoral contrasts further illustrate how ethical maturity varies by context. An aviation respondent noted, "We already operate under strict regulation; adding AI just means adapting those same principles. The difference is that AI shifts the line of responsibility, who's accountable when something goes wrong?" Here, an established compliance culture provides a foundation for responsible innovation, yet the locus of accountability blurs when algorithmic systems mediate human decisions. This shift reflects the evolving regulatory challenge described by Floridi et al. (2018): as agency disperses between humans and machines, responsibility must be redistributed through transparent governance mechanisms.

The findings reveal that while ethics and compliance are now part of strategic discourse, their implementation remains inconsistent. Suppliers described governance as reactive and fragmented, and regulators highlighted the absence of sustained accountability mechanisms. This gap between intention and institutionalisation underscores the need for practical tools that help organisations translate ethical aspirations into structured routines.

The checklist shown in Table 5.2 synthesises these insights with leading responsible-AI frameworks, offering stepwise guidance for integrating ethics and compliance into everyday practice.

Together, the evidence shows that ethical and regulatory readiness remains underdeveloped. Most organisations demonstrate awareness of obligations but do not

TABLE 5.2 Ethics and compliance readiness checklist

Readiness dimension	*Objective*	*Practical actions*	*Empirical & Framework references*
Policy integration	Align AI ethics with organisational values and regulatory obligations.	• Adopt a board-endorsed AI ethics policy. • Map AI use cases to relevant legislation (EU AI Act, GDPR).	PwC Responsible AI Framework (2021). Regulator interview – *"Governance evolves with technology."*
Governance structure	Establish accountability and clarify ethical oversight roles.	• Appoint designated ethics leads for AI projects. • Create internal ethics review boards for high-risk applications.	Microsoft RAI MM – Governance Tier; Supplier interview – *"Governance isn't the finish line."*
Training and culture	Build awareness and ethical literacy throughout teams.	• Deliver role-based Responsible AI training. • Discuss ethical risks during design and post-mortems.	PwC RAI – Principle 3: Fairness; Supplier interview – *"Clients underestimate the emotional side of transformation."*
Risk and impact assessment	Identify and mitigate harm proactively.	• Conduct AI impact assessments before deployment. • Document risk mitigations and accountability chains.	EU AI Act; Regulator interview – *"Assurance must evolve continuously."*
Monitoring and auditing	Ensure continuous accountability and transparency.	• Monitor model drift, bias, and unintended outcomes. • Conduct periodic ethical audits and publish findings.	Vakkuri et al. (2021). Supplier interview – *"After the launch, attention drops."*

Source: Authors' own elaboration, based on Floridi et al. (2018), Vakkuri et al. (2021), Thiebes et al. (2020), PwC Responsible AI Framework (2021), and Microsoft Responsible AI Maturity Model (2022).

Note: Table 5.2 integrates insights from interview data with established AI. It provides a stepwise approach to embedding ethics and compliance into AI readiness, bridging aspirational values with actionable governance.

integrate ethics into practice. Regulation is interpreted as a compliance checklist rather than a cultural framework, and responsibility remains reactive. The stepwise actions outlined in Table 5.2 provide a pathway for moving beyond this gap, aligning technical and legal governance with moral accountability. Embedding ethical deliberation, fairness metrics, and explainability into every stage of the AI lifecycle enables compliance to evolve into genuine readiness, where organisations move from meeting standards to modelling trustworthiness.

Alongside the Ethics and Compliance Readiness Checklist, organisations can cultivate an ethical mindset through daily micro-practices that humanise compliance.

BOX 5.7 REINFORCING ETHICAL MINDSET IN PRACTICE

- Discuss one ethical dilemma per team meeting to normalise reflection.
- Encourage staff to flag fairness or bias concerns without fear.
- Integrate "ethical impact" as a criterion in project approvals.
- Conduct systematic fairness and bias testing at each stage of model training and retraining, documenting findings and mitigation steps (like PWC's Algorithmic Impact Assessment).
- Provide targeted training for teams on interpreting fairness and algorithmic output tests, ensuring that technical and ethical oversight remain connected

Note: These micro-practices turn compliance into active ethical engagement, ensuring fairness becomes a measurable discipline rather than an abstract principle.

5.3.4 Cluster 4: Human–AI Interaction Readiness

Human–AI interaction and trust readiness refer to the degree to which people can understand, rely on, and critically evaluate AI systems within their decision-making environments. The literature conceptualises trust as both a psychological state and a behavioural calibration between human and machine performance (Lee & See, 2004). Too little trust results in disuse; too many leads to overreliance. Thiebes et al. (2020) argue that successful AI adoption requires maintaining this equilibrium through transparency, explainability, and meaningful human oversight. Vakkuri et al. (2021) extend this view by framing trust as a governance property: a system-level condition built through interfaces, feedback loops, and accountability structures rather than an individual emotion.

Across all interviews, participants emphasised that trust depends on understanding. A public-sector respondent explained, "People will trust AI when they can understand its logic. If they can't explain it, they'll bypass it." This statement reflects the cognitive component of trust, the need for interpretability and clarity before behavioural reliance. As Ryan (2020) notes, explainability bridges epistemic uncertainty and moral accountability, allowing humans to justify reliance on non-human agents.

In technical and safety-critical environments, the human–AI relationship was marked by ambivalence. A maintenance engineer described, "Engineers feel confident when the system confirms what they already know. When it contradicts them, they tend to ignore it." This echoes the confirmation bias described by Lee and See (2004), in which trust aligns with prior beliefs rather than with objective performance. Even in contexts with mature digital infrastructures, human operators often interpret AI recommendations through the filter of professional experience, accepting outputs that reinforce intuition and discounting those that challenge it.

The AI supplier echoed these patterns, observing the volatility of user trust once systems move from novelty to routine. One expert noted, "The irony is that when the system is right too often, users stop checking it. When it's wrong once, trust collapses completely." This reflects the asymmetric fragility of trust in automation: reliability builds confidence slowly but erodes rapidly after error. The supplier interviews also confirmed

that many clients still lack a structured concept of meaningful human oversight, often defaulting to either full automation or complete rejection, both extremes of mistrust. Perceived fairness amplifies or erodes this trust: users are less likely to rely on systems they suspect of bias, regardless of accuracy. Transparency about how data quality and fairness are managed thus becomes integral to trust calibration.

Overall, the findings indicate that readiness for human–AI interaction necessitates more than just accuracy. They also reveal that human–AI interaction readiness requires trust calibration, rather than technical literacy. Users must learn to navigate the boundary between control and delegation, confidence and scepticism. Building this readiness requires more than user training; it demands systemic mechanisms that reinforce transparency, feedback, and human interpretability. As the literature suggests (Lee & See, 2004; Thiebes et al., 2020; Vakkuri et al., 2021), readiness in this cluster matures only when trust becomes an iterative process: designed into workflows, validated through oversight, and sustained through accountability.

Building trust in AI is not only a design issue but a cognitive discipline. The following levers strengthen users' interpretive confidence and calibrate trust.

BOX 5.8 PRACTICAL LEVERS FOR TRUST CALIBRATION MINDSET

- Include explainability sessions where users challenge and validate AI outputs, during all phases of implementation.
- Use paired decision logs: human rationale vs AI rationale.
- Train teams to analyse failure cases to rebuild informed trust.
- Rotate oversight roles so multiple users experience the AI system critically.
- Discuss fairness perceptions during training and onboarding, clarifying how bias mitigation and data-quality controls reinforce reliability.

Note: These levers promote reflective trust, a balance between confidence and caution, while linking fairness and transparency to user confidence. Essential for resilient human–AI collaboration.

5.3.5 Cluster 5: Client Preparedness and Change Management

Client preparedness and change management readiness concern how organisations anticipate, communicate, and adapt to the behavioural and structural shifts introduced by AI. The literature consistently highlights that technical and strategic readiness are insufficient if the human system does not evolve in parallel (Jöhnk et al., 2021; Sadiq et al., 2021). Butler et al. (2021) emphasise that organisational adoption depends on both cognitive and emotional engagement, on how employees interpret change, and on whether they perceive it as empowering or threatening. Readiness, therefore, involves the capacity to manage uncertainty, foster learning, and maintain alignment between technological ambition and workforce confidence within a change management process.

Empirical findings underscored this interdependence, including uneven levels of staff preparedness. A public-sector respondent observed, "If automation is introduced too

quickly, it can create frustration, anxiety, and even an existential crisis for staff. Preparation is critical. Right now, I don't think we're handling that as well as we should." This reflection echoes Sadiq et al.'s (2021) argument that organisational readiness for AI is relational progress in one dimension that can be undermined by resistance or fear in another. Emotional readiness thus becomes a determining factor of adoption, particularly when employees perceive AI as a replacement rather than an aid.

Similar sentiments appeared in the aviation sector, where structured communication was seen as essential for sustaining trust. One participant explained, "Reactions among staff are mixed. Some are enthusiastic about AI, while others remain sceptical. Early engagement and clear communication are critical to managing expectations." Jöhnk et al. (2021) likewise emphasise that leadership commitment must be accompanied by open dialogue and participatory mechanisms to translate awareness into motivation. Where such mechanisms are absent, enthusiasm dissipates after pilot phases and projects fail to institutionalise learning.

In professional service environments, resistance was linked to personal identity and perceived indispensability. A wealth-management participant reflected, "It will be difficult, and we'll need to manage it carefully. The challenge comes down to how indispensable someone believes their role is; the more they feel defined by it, the harder it will be to get them to change." This dynamic aligns with Butler et al. (2021), who highlight that self-efficacy and perceived professional value are central to employees' willingness to adopt AI-augmented workflows. Here, preparedness depends on how leaders frame transformation: whether AI is introduced as a threat to expertise or as an extension of it.

Smaller and mid-sized organisations expressed a quieter form of unreadiness, characterised less by resistance than by passivity. As one manufacturing respondent noted, "While there is no outright resistance, the concept of AI still feels somewhat alien, meaning employees are not actively suggesting use cases or showing strong enthusiasm." Such inertia points to a missing link between awareness and participation: readiness falters when individuals cannot visualise the role AI plays in their work.

The AI supplier corroborated these insights, observing that the technical and emotional dimensions of change often diverge. One expert explained, "Clients underestimate the emotional side of transformation. They prepare technically but not psychologically." This confirms that organisational change for AI adoption requires what Sadiq et al. (2021) describe as interdependence readiness: an integration of technical competence, emotional adaptation, and continuous communication.

Together, these findings demonstrate that client preparedness and change management readiness hinge on the organisation's ability to synchronise its mindset with its methods. Communication, participation, and reinforcement transform initial awareness into sustainable behaviour. Without these, readiness remains declarative, a state of knowing without doing.

In practical terms, advancing change-management readiness involves deliberate leadership interventions: linking AI projects to shared organisational goals, recognising early adopters who model new behaviours, and creating short feedback loops to keep dialogue alive after implementation. Structured learning sessions where employees can interrogate outputs or co-design improvements help convert anxiety into agency. Over time, such interventions normalise adaptation and reduce the emotional friction that typically accompanies technological disruption.

Interviewees repeatedly linked readiness failures to unaddressed fears and weak communication. To operationalise a change-management mindset, leaders can implement small, visible interventions that transform uncertainty into engagement.

BOX 5.9 PRACTICAL LEVERS FOR CHANGE MINDSET AND WORKFORCE ADAPTATION

- Co-design experimentation, Proof of Concepts and pilot projects with staff who will use AI outputs.
- Frame AI as augmentation, not replacement, through internal storytelling and role-modelling.
- Establish feedback channels where employees can question or suggest improvements.
- Recognise adaptive behaviour publicly and link it to professional growth.
- Provide psychological-safety training for managers, facilitating transformation.

Note: These levers convert abstract notions of "resistance management" into concrete social rituals that sustain trust and motivation during AI transitions.

5.3.6 Cluster 6: Post-Deployment and Continuous Adaptability

This cluster captures the organisation's capacity to sustain, recalibrate, and evolve AI systems after implementation. The literature stresses that readiness cannot be static: technological, regulatory, and societal conditions evolve continuously, requiring organisations to embed learning and feedback into the lifecycle of AI systems (Butler et al., 2021; Sadiq et al., 2021). Jöhnk et al. (2021) similarly highlight that maturity depends on maintaining adaptive processes that transform one-time implementation into institutional learning. Without such mechanisms, early success gives way to erosion, as systems age faster than the organisations that deploy them.

Empirical evidence confirms that adaptability remains underdeveloped. A public-sector interviewee explained, "We need to keep auditing the models after they go live. The data environment changes fast, and without periodic checks, accuracy declines." Another added, "We're still learning how to build feedback loops. People think once a model works, it's done. That mindset must change." These reflections illustrate what Weiner (2009) conceptualises as the dynamic interplay between capacity and commitment; organisations move forward or regress as their social and technical conditions evolve. Adaptability, therefore, is not a supplementary stage of readiness but its living expression. It also aligns with what Felemban et al. (2024) describe as resilience readiness: the ability not only to deploy but also to absorb disruption and regenerate performance as conditions change.

The AI supplier underscored the same point from an implementation perspective. One expert observed, "Deployment isn't the finish line. The hardest part is keeping the model relevant. Clients often forget that data drift is inevitable." Another noted, "After the launch, attention drops. Organisations celebrate success but stop collecting insights, and performance starts decaying." Their accounts align with Sadiq et al.

(2021) and Vorvoreanu et al. (2023), who emphasise that AI adoption must be treated as an iterative ecosystem rather than a linear project. Without continuous monitoring and organisational learning, systems stagnate, and the cognitive link between human judgment and algorithmic behaviour weakens.

Regulators framed the issue as one of evolving assurance whilst also linking it to the EU AI Act obligations for ongoing monitoring and risk management of high-risk applications. During the interview, one regulator explained, "Governance doesn't stop at certification. Systems evolve, and oversight must evolve with them." This perspective reflects the shift toward continuous accountability discussed by Vakkuri et al. (2021), where ethical and operational assurance become cyclical rather than episodic. As discussed in Cluster 3, when ethical oversight is deferred to post-deployment audits, accountability becomes retrospective rather than preventive. Cluster 6 thus extends that logic: adaptability ensures that governance remains live, anticipatory, not reactive.

Some industries already embody this mindset structurally. As noted earlier in Cluster 2, the aviation sector's reliance on iterative validation cycles provides a strong example of post-deployment discipline. Each system update undergoes review to safeguard reliability, illustrating how adaptability and assurance can coexist. However, this procedural rigour, while ensuring stability, may also slow innovation, revealing the delicate balance between safety and agility that other sectors must navigate as they mature.

Taken together, these insights show that post-deployment readiness is both a technical and cognitive capacity. It requires organisations to view AI not as a finished product but as a living system, one that demands maintenance, monitoring, and recalibration. Continuous adaptability transforms readiness from a pre-implementation condition into an ongoing capability, sustaining alignment between data, governance, and performance. In this sense, adaptability represents the most tangible expression of maturity: it reflects an organisation's willingness to learn from itself.

Practically, fostering adaptability requires embedding structured learning cycles into operational routines. Organisations can schedule periodic "model health" reviews that include both technical and ethical checks, encourage teams to document lessons learned from each iteration, and make results visible across departments. Treating post-deployment reviews as shared learning events rather than compliance exercises helps cultivate the reflex of recalibration that defines maturity.

Sustaining maturity depends on cultivating a learning mindset that normalises iteration and reflection. The next set of levers shows how organisations can embed continuous adaptability into everyday routines.

BOX 5.10 PRACTICAL LEVERS FOR A CONTINUOUS-LEARNING MINDSET

- Conduct periodic "model-health" and ethics reviews as shared learning events.
- Create cross-functional retrospectives after each deployment cycle.
- Use internal dashboards to visualise lessons learned and corrective actions.
- Rotate staff through data, ethics, and operations roles to broaden perspective.
- Celebrate updates and recalibrations as milestones of progress, not signs of failure.

Note: Such practices institutionalise curiosity, "health checks", and accountability, ensuring that readiness remains dynamic rather than episodic.

5.3.7 Cluster 7: Risk and Security Readiness

This cluster encompasses the mechanisms through which organisations identify, assess, and mitigate risks associated with AI systems, including data protection, model integrity, and systemic resilience. The literature consistently highlights that AI-specific risk management should consider beyond traditional cybersecurity: it demands ongoing governance of uncertainty, bias, and unintended consequences (Jöhnk et al., 2021; Sadiq et al., 2021; Vakkuri et al., 2021). Butler et al. (2021) note that preparedness involves integrating security and risk frameworks into organisational routines rather than treating them as isolated technical safeguards.

Across the interviews, participants acknowledged the growing complexity of AI-related risks but revealed divergent levels of preparedness. A public-sector representative explained, "We have good cybersecurity measures, but keeping pace with AI-specific risks is difficult. Threats evolve faster than the processes that contain them." This statement reflects what Felemban et al. (2024) describe as the resilience gap: the widening divide between evolving risks and the adaptive capacity of governance frameworks. The same interviewee noted, "Risk appetite is still conservative. We prefer to delay implementation rather than risk reputational damage from a system failure." This cautious posture underscores the cultural dimensions of risk readiness: as organisations attempt to protect trust, they may inadvertently inhibit innovation, confirming Jöhnk et al.'s (2021) observation that maturity in risk governance requires balancing assurance with agility.

From a regulatory standpoint, they highlighted a persistent issue of risk perception. One representative explained, "The EU AI Act introduces risk categories, but organisations often misclassify themselves as low risk to avoid regulatory effort. That's a cultural issue, not just technical." This echoes Deloitte's (2024) findings that governance maturity is uneven across sectors, with many firms adopting compliance-driven strategies that obscure rather than address systemic risk. The same participant emphasised that "security isn't just about systems; it's about trust. Assurance requires continuous oversight, not just penetration testing." This reinforces Vakkuri et al.'s (2021) argument that accountability for AI must extend beyond security audits toward continuous accountability, a state in which oversight, assurance, and ethical responsibility evolve together.

The AI supplier confirmed the existence of operational blind spots. One respondent noted, "Clients assume that because the system runs in a secure environment, they're covered. However, data pipelines, APIs, and model storage create unseen vulnerabilities." Another stressed the importance of shared accountability: "Risk needs shared ownership. The provider manages the platform, but the client owns the data and the decisions. Many still don't see it that way." These perspectives align with Vorvoreanu et al. (2023), who argue that AI ecosystems distribute responsibility across multiple actors, making clarity of ownership essential to maintaining security and trust. Without mutual understanding of these boundaries, risk becomes a moving target rather than a managed process.

The aviation sector offered a contrasting picture, where risk and security are deeply embedded in operational culture. As one participant explained, "Everything we do runs through safety cases. That discipline protects us, but it also makes experimentation difficult." Here, maturity in risk management provides a model of reliability but at the expense of flexibility, highlighting the inherent trade-off between control and innovation that many sectors face as they seek to embed AI responsibly.

Finally, risk and security readiness are inseparable from trust. As one financial services professional remarked, "Security is critical because we handle sensitive client data. AI tools can't compromise that. One breach would destroy trust." This reiterates that security is not purely a technical property but a relational one: a breach of data integrity or accountability is also a breach of legitimacy.

Together, these findings show that risk and security readiness converge at the intersection of governance, culture, and adaptability. Organisations must treat risk not as a constraint but as a dynamic indicator of maturity; one that evolves alongside systems and stakeholders. The most advanced actors demonstrate not just compliance, but anticipatory assurance: the ability to predict vulnerabilities, adapt governance, and sustain public confidence as technology and threat landscapes change.

Strengthening risk and security readiness, therefore, involves cultivating a proactive risk culture. Leaders can integrate AI-specific risks into existing enterprise risk frameworks, establish cross-functional security forums that include data scientists and compliance officers, and conduct scenario exercises to test organisational responses to model failure or data breach. Embedding these practices transforms risk management from a defensive posture into a continuous assurance habit, reinforcing trust internally and externally.

Maturity in this cluster is reflected not only in technological resilience but also in demonstrable assurance practices recognised in the literature, such as adherence to industry standards and certifications (e.g. ISO 27001), which reinforce stakeholder confidence and institutional accountability. A mature risk culture depends on a proactive mindset that treats uncertainty as a signal, not a setback, and on cautious oversight rather than one-off audits. The following levers help cultivate anticipatory assurance.

BOX 5.11 PRACTICAL LEVERS FOR RISK AND RESILIENCE MINDSET

- Run scenario simulations of model failure or data breach (including infrastructure failure).
- Pursue recognised information security certifications (e.g. ISO 27001)
- Reward early risk reporting to reinforce transparency.
- Integrate ethics and security teams in joint incident reviews.
- Visualise risk trends on shared dashboards for collective awareness.

Note: Embedding these practices nurtures vigilance as a shared norm, shifting risk management from defensive reaction to forward-looking resilience.

5.3.8 Cluster 8: AI Governance, Accountability, and Assurance

Governance and assurance readiness represent the organisational capacity to ensure accountability, traceability, and ethical integrity throughout the AI lifecycle. The literature consistently identifies governance as the unifying dimension of readiness, linking strategy, culture, risk, and ethics into a coherent oversight system (Jöhnk et al., 2021; Sadiq et al., 2021; Butler et al., 2021). Deloitte (2024) defines governance as a "living framework" that evolves alongside technology, while PwC's Responsible AI Framework (2021) advances a similar view: responsible AI requires operationalising principles of fairness, transparency, and accountability within decision-making processes, rather than treating them as mere compliance artefacts.

Across interviews, participants echoed the need to move from governance as control to governance as learning. A public-sector leader affirmed, "Governance sits with us. That responsibility cannot be outsourced." Adding further, "We've introduced a data governance framework, but AI adds another layer. The governance model needs to expand beyond compliance to include explainability and accountability." These statements reflect what Sadiq et al. (2021) describe as embedded governance: oversight that evolves within the organisational fabric rather than being applied externally.

Regulators framed governance as a continuous process of assurance rather than a one-time evaluation. A representative from the interviewed regulator explained, "Our role is to ensure that governance evolves with technology. Assurance isn't a certificate; it's a living process." This sentiment resonates with Vakkuri et al. (2021), who conceptualise continuous accountability as the cornerstone of trustworthy AI: assurance must accompany systems through their lifecycle, adapting to new risks, data sources, and societal expectations. As Vorvoreanu et al. (2023) observe, adequate assurance emerges from transparency and stakeholder alignment, requiring governance that balances technical reliability with human interpretability.

The AI supplier also highlighted governance as a shared construct. One AI vendor explained, "Governance has to be a partnership. The client defines the business context; we ensure the models follow ethical and technical standards. One without the other doesn't work." This view aligns with the PwC (2021) and Deloitte (2024) frameworks, which both emphasise collaborative accountability among AI providers, users, and regulators. Without shared responsibility, assurance collapses into fragmented compliance; strong on documentation but weak on practice. As the supplier cautioned, "Some organisations still treat governance as audit theatre; it looks impressive on paper, but nothing changes in how systems are designed or monitored." This observation captures the maturity gap between performed accountability and practised governance, underscoring that genuine assurance depends on behavioural change, not procedural display.

Highly regulated industries already model aspects of mature assurance. As an aviation professional noted, "Every procedure has a checklist, audit trail, and sign-off. That's how we maintain traceability. AI governance will need the same discipline if it's to be trusted." Similarly, in professional services, transparency emerged as the linchpin of assurance: "Clients need to see how decisions are made and by whom. Without that, there's no assurance." These examples demonstrate that governance maturity is achieved when transparency becomes operational rather than rhetorical, a condition in which accountability is both visible and verifiable.

To close the readiness loop, governance must not only define rules but also nurture an ethical mindset that keeps accountability alive across decision levels. The following levers illustrate how organisations can embed conscience into governance practice.

BOX 5.12 EMBEDDING ETHICAL MINDSET IN GOVERNANCE

- Incorporate "ethics moments" into governance-board meetings.
- Recognise individuals who consistently implement controls and governance daily, rather than only doing so during an audit.
- Link performance incentives to transparent decision-making and responsible AI metrics.
- Require teams to publish explainability statements alongside technical documentation.
- Include external observers or ethics advisors in periodic assurance reviews.
- Publicly report on corrective actions taken after audits to model integrity.

Note: These steps transform governance from a compliance structure into a behavioural system of trustworthiness, where accountability is visible and continuous.

Together, these findings confirm that governance and assurance readiness form the integrative core of AI maturity. Governance aligns technical and ethical capacity, while assurance sustains public trust over time. Organisations that treat assurance as ongoing learning, regularly auditing, validating, and communicating how AI decisions are made, demonstrate both capability and conscience. In this sense, governance readiness closes the readiness loop: it transforms fragmented practices into a self-regulating system where accountability, transparency, and ethical integrity evolve continuously with technology.

The cluster analysis highlights the interdependence of readiness and maturity, with progress in one area often reinforcing or constraining progress in the other. The findings show that organisations cannot treat dimensions such as strategy, governance, ethics, risk, and culture in isolation. Instead, they must be understood as part of an integrated framework in which mindset and maturity evolve together. The following section consolidates these insights into a unified perspective on AI readiness and maturity. Without genuine governance and accountability, organisations risk hollow adoption, regulatory breaches, and reputational collapse.

5.4 Synthesis of Cluster Analysis

Four cross-cutting insights emerged. Primarily, mindset proved as decisive as technical or structural readiness. Across clusters, openness to experimentation, willingness to recalibrate, and tolerance for uncertainty often determined whether maturity advanced or stalled.

The findings show that mindset operates as a latent lever; the cognitive and cultural condition that translates potential into performance. In strategic and cultural readiness, it manifested as strategic conviction: leadership's capacity to move from rhetorical ambition to enacted change. In terms of technical and operational readiness, the mindset emerged as one of ownership, marking a shift from dependence on external

expertise to internal learning and stewardship. Within ethical and regulatory readiness, the mindset reflected moral awareness, evolving from compliance to conscience. In human–AI interaction, mindset change emerged as trust calibration, the psychological readiness to balance confidence and control. Together, these patterns confirm that mindset shapes not only how readiness develops but how maturity stabilises.

Subsequently, governance, accountability, and ethics repeatedly emerged as areas of weakness. Organisations frequently expressed principles without embedding them into practice, echoing the literature on symbolic compliance and the "say–do gap." Supplier critiques, such as "audit theatre", reinforced the fragility of governance when it was treated as a performative exercise.

Adaptability also emerged as a defining capability. Organisations that treated deployment as a continuous cycle of monitoring and recalibration demonstrated greater resilience to disruption. Still, this adaptability depended as much on culture and leadership commitment as on technical safeguards.

Finally, progress across clusters was uneven, with advances in technical and strategic areas often outpacing those in ethics, governance, and cultural integration.

The empirical synthesis also revealed a set of recurrent barriers that cut across clusters, highlighting why progress remains uneven and often fragile. Table 5.3 summarises these challenges and the corresponding actions identified in the study, linking structural and behavioural obstacles to practical organisational responses. These insights translate the preceding analysis into a diagnostic tool for strengthening both maturity and mindset.

Patterns across actors reinforced these insights. Adopters typically advanced in infrastructure and technical experimentation but lagged in embedding governance, ethics, and cultural change. The AI supplier acted as both an enabler and a critic: providing technical capacity and tools where internal expertise was lacking, but also exposing weaknesses, warning against checklist compliance, and highlighting the risks of vendor dependency. Regulators shaped adoption as both drivers and enablers: on the one hand, setting clear obligations through the EU AI Act; on the other, supporting SMEs through guidance and awareness initiatives that reframed compliance as a trust-building measure.

Figure 5.2 illustrates these patterns, underscoring the uneven interplay of readiness, maturity, and mindset, and confirming that adoption remains fragile when mindset lags structural progress. Together, these insights reaffirm the value of the AI-RMM framework, which positions mindset and maturity as dynamic levers: mindset initiating change through belief and behaviour, and maturity consolidating through governance and continuous learning. Sustainable adoption occurs only when these two evolve in synchrony.

5.4.1 Moving from Baseline to Demonstrated Maturity: A Dynamic Approach

The AI-RMM's dynamic feedback logic is most visible when organisations must adapt to regulatory or contextual change. Unlike static models that capture readiness at a single point in time, the AI-RMM treats maturity as an evolving capacity rather than a linear destination. Baseline maturity represents the regulatory and infrastructural minimum for responsible AI use, while demonstrated maturity emerges when feedback loops integrate learning, ethics, and performance into continuous adaptation.

TABLE 5.3 Implementation challenges and practical solutions across AI readiness clusters

Barrier / Challenge	*Description (from findings and literature)*	*Practical Response/ Organisational Action*	*Relevant Cluster(s)*
Cultural fragility and fear of AI	Employees associate AI with job loss or loss of control; risk aversion limits experimentation.	• Frame AI as augmentation, not replacement. • Create safe "sandbox" pilots. • Encourage, recognise and reward experimentation. • Address ethical concerns early and relate them to organisational values.	Strategic & Cultural (1)
Symbolic adoption/ "Pilot trap"	Projects remain at the proof-of-concept stage and have not been scaled; there is a lack of post-deployment feedback.	• Ensure that pilot projects move to production environments and are "operationalised" • Establish post-deployment review loops. • Assign ownership for scaling and maintenance. • Integrate lessons into governance dashboards. • Use communication dashboards or briefings to share pilot outcomes across departments	Post-Deployment Adaptability (6)
IT Systems and Technology Infrastructure	Legacy systems can be a barrier to effective AI implementation. Technology Infrastructure must be secure against cyber-attacks and be resilient.	• Plan and implement modernisation of IT systems and infrastructure • The technology infrastructure must be resilient and secure.	Technical (2), Risk & Security (7)
Data silos and poor interoperability	Fragmented data architectures impede analytics and automation.	• Establish data governance process. • Build shared data platforms and APIs. • Nominate cross-functional data stewards. • Adopt enterprise data governance standards.	Technical (2), Risk & Security (7)
Low AI literacy and skill asymmetry	Staff lack understanding of AI capabilities, limits, and ethics.	• Implement tiered AI literacy programmes. • Pair domain experts with data scientists/architects. • Promote "explainability sessions" led by AI teams.	Strategic & Cultural (1), Human-AI Interaction (4)

Barrier / Challenge	*Description (from findings and literature)*	*Practical Response/ Organisational Action*	*Relevant Cluster(s)*
Reactive governance, compliance theatre, and fairness metrics	Ethics frameworks are adopted superficially; compliance is treated as box-ticking.	• Embed accountability in project lifecycles. • Look for and mitigate "say–do" gaps. • Use dynamic registers for risk and ethics tracking. • Conduct regular internal audits. • Embed accountability and fairness metrics in project lifecycles	Ethical & Regulatory (3), Governance & Assurance (8)
Lack of cross-functional collaboration	Departments pursue AI in isolation; no shared goals or language.	• Form AI steering committees including IT, legal, operations, and HR. • Map dependencies before project approval.	Strategic & Cultural (1), Technical (2)
Leadership ambiguity and inconsistent vision	Senior leaders endorse AI rhetorically but fail to prioritise resources or cultural change.	• Develop a formal AI strategy endorsed by the board. • Integrate AI KPIs into leadership performance reviews. • Ensure that the strategy and business plan adapt to new evolving scenarios.	Strategic & Cultural (1)
Fragmented communication	Limited internal dialogue and poor connection with sectoral networks or regulatory fora reduce visibility on evolving standards, ethical debates, and compliance updates.	• Establish communication routines that include periodic briefings, updates from regulators, and participation in professional chambers or industry associations. • Maintain internal newsletters or AI governance bulletins to cascade new developments.	Strategic & Cultural (1), Ethical & Regulatory (3), Governance & Assurance (8)
Limited risk and resilience planning	Organisations are unprepared for model drift, bias, or system failure.	• Establish risk registers and scenario testing. • Integrate AI resilience into business continuity plans.	Risk & Security (7), Post-Deployment Adaptability (6)
Vendor dependency	Adopters rely on suppliers for technical and ethical guidance, creating knowledge gaps.	• Require transparency clauses in contracts. • Build internal AI competency centres. • Co-design governance with suppliers.	Technical (2), Governance & Assurance (8)

Barrier / Challenge	*Description (from findings and literature)*	*Practical Response/ Organisational Action*	*Relevant Cluster(s)*
Uneven mindset maturity	Technical capacity outpaces behavioural readiness; mindset lags behind tools.	• Apply mindset diagnostics at project inception. • Facilitate peer learning across departments. • Align incentives with learning and adaptability.	All clusters (transversal)

Source: Authors' own elaboration, based on interview data and literature reviewed in Chapter 5, including Vakkuri et al. (2021), Floridi et al. (2018), Thiebes et al. (2020), and Jamil et al. (2025).

Note: Table 5.3 consolidates recurring barriers and corresponding actions derived from interview data and literature across the eight readiness clusters. It operationalises the unified framework by linking structural, cultural, and governance challenges with tangible organisational responses, supporting continuous progression from readiness to maturity.

While Table 5.3 identifies recurrent barriers and actions, the progression from readiness to maturity remains underexplored in most frameworks. The AI-RMM reconceptualises this movement not as a stage-jump but as a continuous loop of recalibration. Box 5.13 illustrates how organisations advance from baseline maturity, where readiness and compliance coexist, to demonstrated maturity, where feedback and learning are embedded into daily operations.

BOX 5.13 MOVING FROM BASELINE TO DEMONSTRATED MATURITY: A DYNAMIC APPROACH

- Step 1: Diagnose and Map. Conduct a multi-cluster self-assessment to identify uneven development across readiness, maturity, and mindset. Use early indicators such as leadership engagement, trust levels, and governance coverage.
- Step 2: Calibrate. Focus interventions on the lowest-performing clusters. For instance, if Cluster 1 (Cultural Readiness) lags, initiate cross-departmental trust-building sessions before scaling technical projects.
- Step 3: Implement and observe. Deploy AI pilots with embedded governance checkpoints. Monitor mindset indicators, such as employee confidence and transparency practices, alongside performance metrics.
- Step 4: Reflect and reassess. Feed results into Cluster 6 (Post-Deployment Adaptability), capturing lessons learned, bias incidents, or governance improvements. Document what has changed and why.
- Step 5: Iterate and Scale. Apply insights across clusters. Demonstrated maturity is achieved when organisations institutionalise these reflexes; adapting governance, retraining models, and recalibrating behaviour whenever external or internal conditions evolve.

Note: This cyclical progression differentiates the AI-RMM from static models that treat readiness as a checklist. It positions AI governance as a living process, where

every external change, be it regulatory, ethical, or technological, becomes an opportunity for recalibration.

This dynamic process is illustrated in the following example (see Box 5.14), which shows how a financial-services organisation applied these steps in practice when responding to regulatory change, leading to a dynamic progression of recalibration across clusters; a mechanism that underpins the transition from baseline to demonstrated maturity.

BOX 5.14 ILLUSTRATIVE EXAMPLE: DYNAMIC RECALIBRATION IN FINANCIAL SERVICES

A mid-sized financial-services organisation, initially assessed at baseline maturity in 2025, had achieved regulatory compliance under the EU AI Act through risk assessments and standard governance procedures. When new transparency requirements were introduced, the organisation's readiness was tested. In a static framework, the firm would perform a one-time compliance adjustment, rewrite its policies, and update its disclosure forms. Under the AI-RMM's dynamic model, the process becomes iterative:

1. Trigger: The compliance change prompts a reassessment in Cluster 3 (Ethical and Regulatory Readiness).
2. Feedback: Findings flow to Cluster 8 (Governance and Assurance), where senior leaders evaluate the implications for accountability for interpretability statements, explainability metrics and communication.
3. Recalibration: Cross-functional teams adapt workflows in Cluster 2 (Technical and Operational Readiness) to log and review explanations within the model interface.
4. Cultural reinforcement: Cluster 1 (Strategic and Cultural Readiness) revisits leadership messages and training to reframe transparency as a trust-building asset, not a compliance burden.
5. Loop closure: Monitoring results in Cluster 6 (Post-Deployment Adaptability) data to inform new governance dashboards, creating a living feedback loop that strengthens both compliance and culture.

Note: Within this dynamic process, the framework transforms regulatory change from a disruption into a developmental pulse. Maturity advances not by stage completion but through continuous recalibration between structure and mindset.

5.5 Implications for Responsible AI Adoption: Insights from the AI-RMM Framework

AI adoption is not a linear progression of technical steps, but a dynamic process where maturity and mindset must advance in tandem. For organisations, this means technical investment and compliance are only a starting point. Lasting adoption requires cultural openness, willingness to experiment, and commitment to continuous recalibration. Treating deployment as an endpoint leaves systems fragile.

For the AI supplier, the role is twofold: enabling adoption with technical expertise while challenging clients when governance is weak or compliance is symbolic. Adoption cannot be outsourced wholesale. Real progress depends on knowledge transfer that strengthens internal maturity and mindset.

For regulators, the EU AI Act sets the floor, but adoption moves further when compliance is framed as trust-building rather than box-ticking. Guidance, capacity-building, and sector-specific support help especially SMEs embed AI responsibly and strategically.

Responsible adoption is therefore shared. Organisations, the AI supplier, and regulators each hold part of the puzzle. Without a mindset, maturity risks becoming hollow. Without maturity, the mindset remains an aspiration. The AI-RMM shows that adoption becomes sustainable only when both evolve together, reinforced by governance, ethics, and external critique.

Fairness and explainability emerge as unifying principles across the AI-RMM. They connect the technical, ethical, and human layers of readiness: from data quality and bias mitigation in system design, to systematic fairness testing during model training, to transparent communication that sustains user trust. When fairness is embedded as both a technical and behavioural norm, explainability ceases to be an afterthought; it becomes a daily accountability practice. Together, they turn governance from a static obligation into a living conversation between technology and its human stewards.

The sectoral distinctions outlined earlier (Section 3.3) revealed that readiness conditions and adoption pathways vary sharply between high-risk and commoditised industries. These differences resurface in practice when the AI-RMM's feedback loops are applied to real organisational contexts, showing how the same framework adapts across divergent regulatory and ethical landscapes.

The same dynamic logic observed in the financial sector applies with greater intensity in high-risk environments, where ethical responsibility and human oversight define the limits of automation. To illustrate how the AI-RMM operates under such conditions, the following example draws on interview evidence from the healthcare sector. It demonstrates how the framework's feedback loops help balance innovation with accountability when decision-making carries direct implications for human life.

BOX 5.15 ILLUSTRATIVE EXAMPLE: DYNAMIC READINESS IN A HEALTHCARE CONTEXT

A regional healthcare provider had achieved baseline maturity by 2026, using AI-assisted imaging to support diagnostics in accordance with existing medical-device regulations. When new EU health data rules introduced tighter standards for explainability and patient consent, the organisation activated its AI-RMM feedback loop.

Step (1) Ethical and Regulatory Readiness: The compliance trigger prompted an audit of patient-data governance.
Step (2) Governance and Assurance: Leadership reassessed accountability between clinicians, data scientists, and compliance officers.
Step (3) Human-AI Interaction: Clinicians and data teams co-designed new explainability dashboards to clarify how AI models contributed to diagnoses.

Step (4) Strategic and Cultural Readiness: Leadership reframed transparency as part of patient trust rather than administrative overhead.
Step (5) Post-Deployment Adaptability: Lessons were logged into the organisation's monitoring system, embedding feedback and continuous learning.

Within six months, the hospital had advanced toward demonstrated maturity: explainability protocols were integrated into routine clinical review meetings, and patient feedback became a regular input to model retraining. The regulatory change thus evolved from a compliance event into a cultural learning cycle.

Note: This example builds on the financial-services scenario presented earlier (see Box 13), showing how the AI-RMM's dynamic structure enables organisations to convert external shocks, regulatory, ethical, or technical, into adaptive cycles of improvement.

The process described in Box 5.15 is visualised in Figure 5.3. The diagram highlights the "lit-up" clusters within the AI-RMM that became active during this recalibration cycle. Each orange-numbered step traces the flow of adaptation triggered by new regulation, from ethical review through governance, human–AI collaboration, strategic reframing, and post-deployment learning. The remaining clusters maintain baseline maturity, ensuring continuity while the active ones adjust. Together, they illustrate how readiness and maturity evolve through feedback loops, and how mindset, manifested in transparency, collaboration, and accountability, transforms compliance into continuous learning.

The patterns observed across the eight clusters reaffirm and refine the unified AI Readiness and Maturity Framework introduced in Section 5.1. They confirm that readiness and maturity are interdependent, continuously influenced by mindset and governance as transversal enablers. The empirical findings provide not only validation but calibration: they illustrate how the framework's theoretical dimensions manifest in practice, where structural progress often outpaces cultural and ethical integration. The AI-RMM thus captures readiness as a dynamic state; fluid across time, actors, and contexts, offering a practical lens through which organisations can diagnose strengths, identify interdependencies, and plan for adaptive, ethically grounded AI adoption.

Across sectors, the AI-RMM reframes compliance as an engine of learning. Every external demand becomes a trigger for recalibration; proof that readiness is not a state to be reached, but a rhythm to be maintained.

5.6 Summary of Section 5

Section 5 presented the unified AI-RMM framework, integrating insights from literature and empirical research into a single, adaptive model for responsible AI adoption. The framework reconceptualises readiness and maturity as interdependent states connected through mindset, emphasising governance, ethics, and adaptability as cross-cutting levers. Its eight clusters, spanning strategic, technical, ethical, human, and governance domains, demonstrate that sustainable adoption depends on iterative feedback loops rather than linear progress. Empirical findings from industry interviews

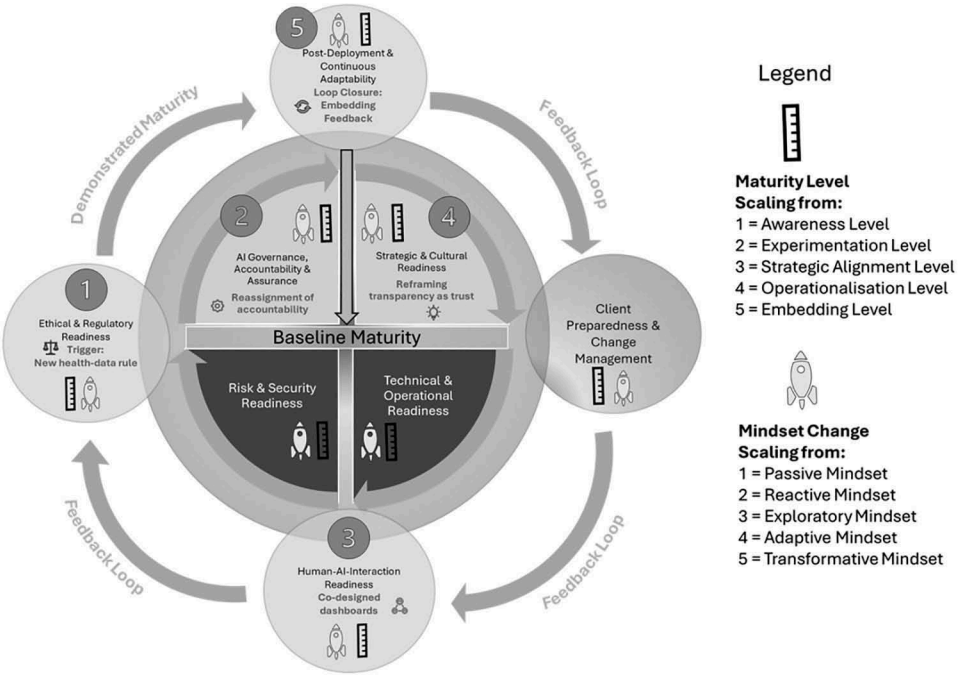


FIGURE 5.3 Dynamic readiness cycle illustrated through the healthcare case
Note: The figure visualises how the AI-RMM feedback loops activate selected clusters in response to regulatory change. Orange-numbered steps trace the progression from ethical trigger (1) to embedded learning (5), demonstrating the transition from baseline to demonstrated maturity as mindset and governance evolve through continuous recalibration. (Based on Box 5.15)
Source: Designed by the authors.

validated the framework's multidimensional structure, showing that maturity advances unevenly across clusters and that mindset often lags behind structural capability. By positioning governance and ethical alignment as the mechanisms that sustain equilibrium between readiness and maturity, the AI-RMM provides a practical blueprint for continuous, responsible transformation.

6 Conclusion

This study has limitations. The interviews covered a cross-section of industries but remain exploratory in scope. More work is needed to broaden the empirical base, extend into additional sectors, and follow organisations over time as regulatory change takes hold. These steps would strengthen the generalisability of the findings and deepen understanding of how readiness, maturity, and mindset evolve in practice. The findings also derive directly from the frameworks analysed and applied. While care was taken to capture the most relevant and widely cited approaches to AI readiness and maturity, it is possible that additional frameworks, had they been included, might

have yielded different or complementary insights. This constitutes a limitation of the study and highlights the importance of continuous engagement with emerging models.

Even with these limits, the study makes a clear contribution. It advances the literature by consolidating readiness, maturity, and mindset into a unified framework, offering practical value by providing adopters, AI suppliers, and regulators with a shared lens for diagnosing gaps and guiding adoption.

The implications point forward. AI adoption will remain fragile if progress in maturity and mindset occurs unevenly. Organisations that build infrastructure without cultural change risk symbolic adoption; those that embrace openness without capacity risk aspiration without delivery. The AI-RMM framework shows how to bridge this divide. By aligning the technical, organisational, and cultural dimensions and reinforcing them through governance, ethics, and external critique, AI adoption can become both sustainable and trustworthy.

The study's contribution is therefore both conceptual and practical. Conceptually, the unified AI Readiness and Maturity Framework illustrated in Figure 5.1 captures readiness as a dynamic, interdependent system shaped by mindset, maturity, and governance. Practically, the interventions outlined in Table 5.3 translate these dimensions into actionable guidance, offering organisations a basis for diagnosing cultural barriers and designing targeted leadership responses. Together, these outputs present the AI-RMM as both an analytical lens and a developmental instrument, enabling organisations to move from awareness to sustained behavioural change in their AI adoption approach.

When readiness, maturity, and mindset evolve in tandem, AI adoption advances beyond compliance to become resilient, ethical, and transformative. This integrated pathway provides organisations with a foundation not only for today's challenges but also for the uncertainties of tomorrow.

References

Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications, 11*(1), 1568. doi:10.1057/s41599-024-04044-8.

Ajzen, I. (1991). The theory of planned behaviour. *Organisational Behaviour and Human Decision Processes*, 50(2), 179–211.

Butler, T., Espinoza-Limón, A., & Seppälä, S. (2021). Towards a capability assessment model for the comprehension and adoption of AI in organisations. *Journal of AI, Robotics & Workplace*. doi:10.69554/KLJJ4414.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. doi:10.2307/249008.

Deloitte. (2024). Deloitte's integrated approach to AI. Deloitte Insights. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/public-sector/ai-readiness-and-management-framework.pdf.

Felemban, H., Sohail, M., & Ruikar, K. (2024). Exploring the readiness of organisations to adopt artificial intelligence. *Construction Management, Computers & Digitisation*, 14(8), 2460. doi:10.3390/buildings14082460.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People: An ethical framework for a good AI society, Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. doi:10.1007/s11023-018-9482-5.

Jamil, K., Zhang, W., Anwar, A., & Mustafa, S. (2025). Exploring the influence of AI adoption and technological readiness on sustainable performance in Pakistani export sector manufacturing small and medium-sized enterprises. *Digital Transformation in Sustainable Supply Chain and Service Operations*, 17(8), 3599. doi:10.3390/su17083599.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not? Exploring digital readiness in the context of AI adoption. *Business & Information Systems Engineering*, 63(5), 329–343. doi:10.1007/s12599-020-00676-7.

Kelm, K., & Johann, M. (2024). Artificial intelligence in corporate communications: Determinants of acceptance and transformative processes. *Corporate Communications: An International Journal*, 30(6), 1–21. doi:10.1108/CCIJ-04-2024-0051.

Lee, J. D., & See, J. E. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. doi:10.1518/hfes.46.1.50.30392.

McGee, F. (2024). Approaching emergent risks: An exploratory study into artificial intelligence risk management within financial organisations. *Computers and Society*. doi:10.48550/arXiv.2404.05847.

PricewaterhouseCoopers (PwC). (2021). *Responsible AI: A framework for building trust*. PwC. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/responsible-ai.html.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Ryan, M. (2020). Ethics of artificial intelligence and robotics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020 ed.). Stanford University. https://plato.stanford.edu/archives/win2020/entries/ethics-ai/.

Sadiq, R. B., Safie, N., Rahman, A. H. A., & Goudarzi, S. (2021). Artificial intelligence maturity model: A systematic literature review. *PeerJ Computer Science*, 7, e661. doi:10.7717/peerj-cs.661.

Tadimalla, S. Y., & Maher, M. L. (2024). AI literacy for all: Adjustable interdisciplinary socio-technical curriculum. In *Proceedings of the 2024 IEEE Frontiers in Education Conference (FIE)* (pp. 1–9). IEEE. doi:10.1109/FIE61694.2024.10893159.

Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, 31(2), 1–18. doi:10.1007/s12525-020-00441-4.

Vakkuri, V., Jantunen, M., Halme, E., Kemell, K.-K., Nguyen-Duc, A., & Abrahamsson, P. (2021). Time for AI (ethics) maturity model is now. *Computers and Society*. doi:10.48550/arXiv.2101.12701.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. doi:10.1287/mnsc.46.2.186.11926.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. doi:10.2307/30036540.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. doi:10.2307/41410412.

Vorvoreanu, M., Heger, A., Passi, S., Dhanorkar, S., Kahn, Z., & Wang, R. (2023). Responsible AI maturity model: Mapping your organisation's goals on the path to responsible AI. Microsoft Research. https://www.microsoft.com/en-us/research/publication/responsible-ai-maturity-model/.

Weiner, B. J. (2009). A theory of organisational readiness for change. *Implementation Science*, 4(67), 1–9. doi:10.1186/1748-5908-4-67.

Yen, W. T. M., Hong, T. L., & Yen, Y. Y. (2024). A proposed framework for assessing the readiness of AI in small and medium enterprises (SMEs) in Malaysia. *Journal of Ecohumanism*, 3(5), 254–260. doi:10.62754/joe.v3i5.3893.

6

A FRAMEWORK FOR ASSESSING AI READINESS IN BUSINESSES

Hossam Kamel

Introduction

AI refers to the ability of machines, particularly computer systems, to simulate human intelligence processes, including reasoning, learning, planning, and perception (Butler et al., 2023). The rapid evolution of AI has positioned it as a core driver of digital transformation across industries. AI capabilities include understanding, reasoning, planning, communication, decision-making, and perception, among others. Machine learning represents a popular subset of AI in which computers automatically learn from data and perform tasks with minimal human input. AI's strategic role and transformational potential in business, coupled with its rapidly evolving landscape, make an AI readiness concept essential for transitioning to an AI-enhanced business (Diallo et al., 2025).

However, existing AI readiness frameworks show important limitations. Many are either overly generic, focusing narrowly on technical dimensions without linking them to strategic or organizational contexts (e.g., capability maturity models), or highly specialized, addressing only specific domains such as data readiness or process management. As a result, they often fail to capture the holistic integration challenges organizations face—such as aligning AI initiatives with corporate strategy, managing cross-functional readiness, or scaling pilots into enterprise-wide transformations (Jöhnk et al., 2021; Butler et al., 2023). This gap is particularly visible in the lack of frameworks that combine the classification of different adopter types with a step-by-step roadmap for readiness and adoption.

This chapter addresses that gap by integrating insights from capability-based and service-management perspectives into a comprehensive AI Readiness and Adoption Framework. The framework makes two main contributions:

1. It introduces a four-dimensional classification of companies that adopt AI, based on the depth of integration, functional focus, strategic intent, and industry sector.

DOI: 10.4324/9781003632429-9
This chapter has been made available under a CC-BY-NC-ND 4.0 license

2. It connects this classification to a seven-stage readiness and adoption roadmap, offering organizations a structured method to evaluate maturity, prioritize use cases, identify gaps, and implement governance.

By explicitly bridging classification and readiness stages, the framework not only enriches the academic literature but also provides a practical tool for decision-makers to plan, prioritize, and execute AI-driven transformations in a scalable and sustainable way.

Thus, the emergence of AI presents tremendous opportunities to support decision-making and gain competitive advantages (Butler et al., 2023). However, firms face strategic and implementation challenges that hinder the successful adoption of AI. The concept of "AI readiness" aims to understand an organization's preparedness for AI adoption, allowing it to plan accordingly (Jöhnk et al., 2021).

Overview of AI Readiness Frameworks

The growing usage of the term AI in the business arena has led to numerous AI maturity and preparedness frameworks prevalent in the consulting community. Most of these frameworks can be categorized into two main approaches (Aldoseri et al., 2024). The first, a Capability Assessment Model, adopts a classic business lens and builds on previous analytical frameworks for emerging technologies reported in the academic literature (Sadiq et al., 2021). These are a set of criteria to guide interpretation and application to specific use cases. The second approach provides a simple AI heatmap based on familiar service management frameworks, including people, processes, and technology (Spring et al., 2022). Each approach offers a distinct perspective on assessing AI readiness within organizations. A clear framework can help understand the organizational environment and guide investments and discoveries aligned with future objectives.

In recent times, numerous scholars and practitioners in the research domain have devoted significant time and effort to developing innovative frameworks. These frameworks are specifically tailored to help firms effectively deploy, comprehensively implement, and seamlessly integrate a range of AI capabilities across their diverse operations. By focusing on these strategic frameworks, the aim is to enhance the overall efficiency and productivity of organizations as they adopt advanced technologies, ultimately leading to a more streamlined and competitive business environment (Bawack et al., 2021).

Classification Framework for Companies Adopting AI

Companies can be classified along four complementary dimensions. These lenses can be used individually or combined for a nuanced analysis.

1 Depth of Integration and Maturity (How Deeply Is AI Embedded?)

This dimension concentrates on the stage and scope of AI adoption within the organization.

AI Experimenters/Piloters

Companies categorized as AI Experimenters/Piloters are those that test AI within isolated, small-scale projects, such as deploying a chatbot for a single marketing campaign or running a predictive maintenance pilot on one machine line. Their primary focus is on learning, proving return on investment (ROI), and gradually building internal AI capabilities, without yet integrating AI as a core element of operations. These organizations typically operate with limited budgets, assign small, dedicated teams—often within innovation labs—depend heavily on external vendors or consultants, and lack a clear strategy for scaling AI initiatives. This group includes many traditional small and medium-sized enterprises (SMEs) as well as larger companies still in the early stages of AI adoption (Michael et al., 2023)

AI-Core Value Providers

AI-Core Value Providers are companies that have progressed beyond the pilot stage to deploy AI solutions across specific functions or processes. These organizations actively use AI to increase effectiveness, reduce costs, or enhance existing offerings within well-defined operational boundaries. They typically have dedicated AI or data science teams, established data pipelines, and a clear focus on scaling successful pilots into broader applications. They tie their AI initiatives to measurable operational key performance indicators (KPIs) to track and optimize their impact. Examples include banks leveraging AI for fraud detection and credit scoring, retailers using AI-driven demand forecasting to optimize supply chains, and manufacturers implementing AI-powered quality control systems (Davenport & Ronanki, 2018).

Recent cases illustrate this trend. For instance, JPMorgan Chase has reported that its COIN (Contract Intelligence) platform processes over 12,000 commercial contracts annually, cutting review time from 360,000 hours of lawyer work to seconds (JPMorgan Chase, 2025). In retail, Walmart and Carrefour use AI-driven demand forecasting to reduce stockouts by up to 30%, improving supply chain resilience (Sheehan et al., n. d.). In manufacturing, Siemens utilizes computer vision for quality control, achieving a 90% defect detection accuracy in pilot production lines. At the same time, BMW integrates predictive maintenance across its plants to significantly reduce downtime costs (Patil, 2024). These examples demonstrate how value providers embed AI not as experiments but as scalable, performance-driven tools that directly enhance efficiency and competitiveness.

AI Transformers/Systemic Adopters

AI Transformers/Systemic Adopters are companies that utilize AI in various key areas, transforming how they conduct business, what they offer, and their workplace culture. In these organizations, AI serves as a core strategic driver rather than a mere operational tool. They demonstrate CEO and board-level commitment to AI, invest significantly in AI-focused research and development, and integrate AI into core products and services. These companies foster a data-centric culture and re-engineer processes to fully leverage AI capabilities. Examples include technology giants such as

Google, Amazon, Meta, and Microsoft—where AI powers search, recommendations, advertising, and cloud services—alongside companies like Netflix, which uses AI for content recommendation and production, and Tesla, which applies AI in autonomous driving and manufacturing (Brynjolfsson & McAfee, 2017).

AI Innovators/Creators

AI creators/pioneers are companies that not only utilize AI extensively but also develop foundational AI technologies, platforms, or novel applications, thereby pushing the boundaries of AI capabilities themselves. These organizations make heavy investments in fundamental AI research, often employing teams of PhD-level experts and operating dedicated AI research labs. They develop proprietary AI models, chips, or platforms and frequently license their AI technologies or offer them as services to other businesses. Notable examples include OpenAI, known for developing the GPT family of models and DALL·E; DeepMind, a Google subsidiary recognized for breakthroughs such as AlphaFold and AlphaGo; NVIDIA, which produces AI-focused hardware and software platforms; and Anthropic, developer of the Claude AI models (Stanford University, 2023).

2 Functional Focus (Where Is AI Applied?)

This dimension categorizes companies based on the primary business functions where AI is most actively deployed.

Customer-Facing AI Adopters

Customer-Facing AI Adopters are companies that strategically leverage AI to enhance customer experience, deliver personalization, and optimize sales and marketing activities. Core applications include chatbots and virtual assistants, recommendation engines, targeted marketing campaigns, sentiment analysis, and dynamic pricing models. For example, Amazon reports that its recommendation system drives 35% of total sales revenue, highlighting the transformative role of AI in personalization (Michael et al., 2023). Similarly, Spotify's AI-driven recommendation ecosystem—spanning Discover Weekly, Day Lists, and the AI DJ—has transformed personalization into a core retention driver. AI-powered users spend 140 minutes/day, compared to 99 minutes for non-AI users, with 40% higher overall retention. The AI DJ feature alone boosted retention by 15%, while personalized playlists contributed to 46% of premium conversions (Space, 2025). These AI-driven tools enable organizations to deepen customer engagement, improve service efficiency, and boost conversion rates (Kaplan & Haenlein, 2019).

Customers interacting with AI now expect the technology not only to automate routine tasks—such as chat support, recommendations, or self-service—but also to deliver measurable business value and reliability. McKinsey's 2024 survey shows that about 65% of organizations are regularly using generative AI (gen AI), particularly in functions directly touching customers such as marketing and sales, and they report revenue increases and cost savings from these uses (McKinsey, 2024). Combined with findings from Deloitte that customers favor AI tools that resolve issues quickly,

personalize engagement, and operate transparently (Deloitte, 2024a, 2024b), these insights imply that customer-facing AI adopters must emphasize both performance metrics (e.g., retention, conversion, revenue uplift) and ethical dimensions (data privacy, transparency, trust) to foster sustained customer loyalty.

Operations and Process AI Adopters

Operational AI adopters are companies that leverage AI to optimize internal processes, supply chains, manufacturing, and logistics with the goals of improving efficiency, reducing costs, and enhancing quality. Common AI applications in this category include predictive maintenance to prevent equipment downtime, supply chain optimization for better demand and resource planning, robotic process automation (RPA) to streamline repetitive tasks, quality control using computer vision systems, and AI-powered inventory management to reduce waste and stockouts. By embedding AI into these operational domains, organizations can achieve measurable performance gains and strengthen their competitive advantage (Wamba et al., 2017).

Decision Support and Analytical AI Adopters

Decision-Support AI Adopters are organizations that employ AI to enhance human decision-making by delivering data-driven insights, advancing risk assessment, and strengthening strategic planning. Their applications span financial forecasting and risk modeling in banking and insurance, fraud detection, pharmaceutical drug discovery and clinical trial analytics, market research, and strategic scenario planning. Recent studies highlight that AI-driven systems not only enhance analytical capabilities and competitive responsiveness but also raise important concerns regarding transparency, trust, and ethical oversight (Csaszar et al., 2024). For example, qualitative and quantitative evaluations indicate that AI recommendation tools can enhance strategic visioning and responsiveness, although challenges persist in terms of transparency and trust. These findings suggest that while AI can significantly augment decision-making, the design of human–AI interaction and the retention of human judgment are crucial to the effective and responsible deployment of AI.

Product/Service Core AI Adopters

Core AI adopters are firms where AI is not just an enabler but the central foundation of their offerings—ranging from autonomous driving and AI-powered drug discovery to generative content and intelligent cybersecurity. As shown in Chapter 10 of Business Model Innovation: Game Changers and Contemporary Issues (Aagaard & Tucci, 2024), AI integration reshapes nine business model components. It redefines value propositions through personalization, drives value creation via automation and learning, and transforms customer relationships with trust and predictive engagement. It also enables new revenue models (subscriptions, pay-per-use), lowers cost structures, and shifts key activities toward AI training and deployment. Supported by data and algorithms as key resources and strengthened through ecosystem partnerships, firms ultimately reposition themselves through distinct AI-driven advantages.

TABLE 6.1 AI-driven business model innovation: real-world company cases

Business Model Component	*Traditional Business Model*	*AI-Driven Business Model*	*Example Company*	*Use Case (Brief Description)*
Customer Segments	Broad, demographic-based targeting	Micro-segmentation & personalization via AI	Netflix	Uses AI to analyze viewing patterns and cluster users into micro-segments for personalized recommendations.
Value Propositions	Generic products/ services	AI-enhanced, predictive, and personalized offerings	DeepMind	Developed Streams app to identify patients at risk of acute kidney injury, improving outcomes in healthcare.
Channels	Conventional retail/distribution	AI-optimized digital assistants & predictive distribution	Domino's Pizza	Uses Dom, an AI-powered assistant, to take voice/text orders across multiple channels.
Customer Relationships	Manual, reactive service	Proactive, predictive, and AI-personalized engagement	Sephora	Offers Virtual Artist AR app for customers to try makeup virtually, enhancing personalized shopping.
Revenue Streams	Static pricing, subscriptions	Dynamic pricing, AI-as-a-Service, predictive upselling	Uber	Implements AI-based surge pricing to adjust fares in real-time based on demand and supply conditions.
Key Resources	Physical assets, staff	Data, algorithms, AI infrastructure	Caterpillar	Uses AI-driven predictive maintenance on heavy machinery to reduce downtime and create new service models.
Key Activities	Production, marketing, logistics	AI training, deployment, analytics	IBM Watson, H2O.AI	Enhance vertical processes such as supply chain management and diagnostics with AI.
Key Partnerships	Suppliers & distributors	Ecosystems with cloud & AI providers	OpenAI–Microsoft	Partnership integrates generative AI into Microsoft Azure for enterprise-scale deployment.
Competitive Positioning	Price, quality, scale	AI-driven differentiation & innovation speed	Tesla	Leverages AI for autonomous driving, providing a strong innovation-based market advantage.

Source: Authors' own elaboration, adapted from Aagaard, A. & Tucci, C. (Eds.) (2024). Business model innovation: Game changers and contemporary issues (Chapter 10).

To clarify these dynamics, we adapted the chapter into a comparative table contrasting traditional and AI-driven models.

This classification clarifies an organization's AI posture—AI-Core Value Provider versus AI Integrator—and sets the starting point for readiness. Building on this positioning, the seven readiness stages operationalize how each type aligns strategy, builds capabilities, and scales AI adoption.

3 Strategic Intent (Why Is AI Being Adopted?)

This dimension classifies companies based on the primary business driver for their AI investments.

Efficiency and Cost-Reduction Seekers

Efficiency-Driven AI Automator's are organizations that deploy AI primarily to automate tasks, lower labor costs, reduce errors, and optimize resource use. Sectors such as manufacturing, logistics, back-office operations (including finance and HR), and traditional service industries are home to these adopters. Their AI applications include intelligent process automation—integrating RPA with AI and machine learning—to streamline workflows and improve accuracy. For instance, a recent case study demonstrated that utilizing generative AI with intelligent document processing (IDP) reduced the time required to process expenses by more than 80%, increased accuracy, and resulted in fewer errors in corporate finance tasks (Jeong et al., 2025). In manufacturing, AI and robotics support predictive maintenance that reduces maintenance costs by 18–25% and cuts unplanned downtime by up to 50% (Viorel, 2025). In supply chain and logistics, AI optimizes operations through smarter demand forecasting, route planning, and autonomous warehouse systems, leading to ~15% reductions in logistics costs, improved inventory accuracy by up to 35%, and service-level gains of around 65% (McKinsey, 2024). Amazon's expanded use of robotics and AI in fulfillment centers has the potential to cut order fulfillment costs by up to 25%, translating into estimated annual savings of $10 billion by 2030, even as human workers continue to handle more complex tasks (Stone, 2025).

Revenue Growth and Innovation Drivers

AI-driven innovators are organizations that strategically employ AI not only as an instrument, but as the central engine behind new products, services, and market positioning. These firms leverage AI to enhance customer satisfaction, unlock new revenue streams, and drive innovation to gain a competitive edge. Examples include providers of autonomous vehicles, AI-powered drug discovery platforms, AI-generated content tools, and intelligent cybersecurity offerings. Recent research highlights that generative AI (GenAI) serves as a powerful catalyst for business model transformation, fueling product and service innovation, reshaping market entry strategies, and enabling entirely new revenue models through "creative destruction" dynamics within industries (Singh et al., 2024). McKinsey's latest findings concur, showing how AI can be deployed to accelerate innovation, making it possible to develop completely new goods and services that boost R&D output.

Risk Mitigation and Compliance Focus

AI-Powered Risk Guardians are organizations that harness AI as a strategic asset to enhance security, detect fraud, ensure regulatory compliance, manage operational risks, and improve safety. Financial services, healthcare, cybersecurity, and other highly regulated sectors typically house these entities. Cutting-edge AI applications—such as hybrid architectures for deep learning combining recurrent neural networks, transformers, and autoencoders in a "Mixture of Experts"—have achieved exceptional fraud detection results while supporting AML and KYC compliance (Vallarino, 2025). Explainable AI (XAI) combined with federated learning further offers privacy-preserving, interpretable models for fraud detection that build regulatory trust (Awosika et al., 2023). A broader 2025 review shows how AI-powered platforms—especially cloud-based ones—are transforming financial security and compliance workflows by automating risk monitoring and boosting operational efficiency (Bartosiewicz, 2024). Across fintech, AI systems that integrate behavioral analytics and XAI are significantly reducing false positives and financial losses, while increasing customer trust (Oyelade & Yusuff, 2025).

4 Industry Sector (Which Industries Are Leading?)

While AI is pervasive, adoption patterns, drivers, and challenges vary significantly across different industries.

High-Tech and Digital Natives

High-tech and digital natives are companies operating in sectors such as internet services, software, and cloud computing, where AI is a core component of their products and services. These organizations demonstrate the highest level of AI maturity, benefiting from substantial investments in research and development, data-rich environments, and advanced infrastructure. They extensively integrate AI into their offerings, including recommendation engines, intelligent analytics, and generative content creation. Recent studies indicate that digital maturity significantly facilitates a business's ability to implement AI-assisted solutions. Effectively (Brătucu et al., 2024), while research on multinational corporations indicates that AI adoption fosters team innovativeness, thereby enhancing overall innovation performance (Javed et al., 2025).

Financial Services

Financial Services AI Adopters are institutions—such as banks, insurers, and investment firms—that heavily focus on AI to bolster areas like risk management, fraud detection, regulatory compliance, algorithmic trading, personalized financial services, and automated customer support. These organizations benefit significantly from their AI maturity, as illustrated by their sector's prominent digital transformation score that demonstrates the vital function of AI in financial innovation (Gkika et al., 2025). Cutting-edge AI-powered advancements in algorithmic trading, including quantitative

investment pipelines enhanced by large language models (LLMs) and deep learning, are transforming the way investment decisions are made—from pattern recognition to autonomous strategy execution (Cao et al., 2025). At the same time, financial institutions must address the "how" of responsible AI deployment; recent qualitative research indicates that, despite being an early adopter, the financial sector still grapples with translating principles such as fairness, accountability, and transparency into practical, operational governance (Skouloudis & Venkatraman, 2025). Together, these findings highlight both the pioneering AI-driven capabilities of financial services firms and the ongoing challenges in ensuring responsible, trustworthy deployment.

Healthcare and Life Sciences

Healthcare and Life Sciences AI Adopters are organizations in pharmaceuticals, biotechnology, and healthcare services (such as providers and hospitals) that leverage AI across essential functions—ranging from drug discovery and diagnostics (particularly medical imaging analysis) to personalized medicine, operational efficiency in clinical settings, and clinical trial optimization. AI's transformative impact is evident in its acceleration of drug development by predicting molecular structures, identifying targets, and designing clinical trials more effectively (Kant et al., 2025). In diagnostics and surgery, convolutional and recurrent neural network–powered AI systems are enhancing diagnostic accuracy and surgical performance, particularly in pathology and imaging (Shin et al., 2025). Deep learning models are also streamlining clinical trial workflows: AI-driven stratification, adverse-event forecasting, and patient recruitment are being used to personalize treatment plans and reduce trial inefficiencies (Anuyah et al., 2024). In dermatology, generative AI and multimodal machine learning are advancing diagnosis and therapeutic decision-making for inflammatory skin conditions, enabling deep phenotyping and tailored care (Tang et al., 2025).

Retail and Consumer Packaged Goods (CPG)

Retail and CPG AI adopters operate across e-commerce, brick-and-mortar stores, and consumer goods manufacturing, where AI is central for personalization, demand forecasting, supply chain optimization, dynamic pricing, and customer service automation. Cutting-edge models—such as neural architectures like Temporal Fusion Transformers and Time-Series Mixers—are advancing retail demand forecasting by delivering higher accuracy across stores and product categories (Theodoridis & Tsadiras, 2025). At the same time, foundation models tailored for supply chains employ a dual-strategy ensembling approach—hierarchical segmentation and architectural diversity—to enhance forecasting performance across hierarchical levels (Yang et al., 2025). Hybrid AI systems that combine advanced visual tracking (e.g., YOLOv8—You Only Look Once, version 8—is a real-time object detection model for identifying items in video streams with object trackers like BOT-SORT and ByteTrack—both are multi-object tracking methods—and sequence models (GRU—Gated Recurrent Unit—is a neural network for sequence modeling) empower smart retail analytics to improve inventory and customer flow understanding (Hossam et al., 2024). Amazon exemplifies operational AI leadership with agentic warehouse robots that manage multiple

functions—such as unloading and retrieval—alongside generative-AI-enhanced mapping tools for delivery and hyper-local demand prediction.

Amazon provides concrete results from such deployments: its DeepFleet foundation model for robot coordination reduced travel time by 10% across warehouse fleets, while the Sequoia system accelerated inventory stowing by up to 75% (Durham, 2025; Robotics, 2025). In forecasting, Amazon's LLMForecaster achieved gains of 32–105 basis points in seasonal demand prediction accuracy, improving hyper-local forecasts and reducing stock imbalances (H. Zhang et al., 2024).

Manufacturing and Automotive

Manufacturing and Automotive AI Adopters are organizations leveraging AI in sectors such as automotive, aerospace, and industrial equipment to power predictive maintenance, computer-vision-driven quality control, supply chain optimization, autonomous systems (including robots and vehicles), and generative product design. AI-powered predictive maintenance—integrating IoT sensors, real-time analytics, and machine learning—has become instrumental in reducing downtime, extending machinery lifespan, and optimizing resources (Patil, 2024). AI-based quality inspection systems now detect defects with pixel-level precision, helping to slash downtime by up to 50% in smart factories. Deep learning frameworks, such as DeepInspect, combine different neural networks and generative networks to enable real-time detection of manufacturing flaws, thereby reducing waste and costs (Kumbhar et al., 2023). In precision manufacturing, hybrid ML models that fuse convolutional feature extraction to accuracy—improving measurement precision by 73% (Samadi et al., 2025). The emerging Industrial AI Foundation Framework—comprising modular knowledge, data, and modeling layers—strengthens the deployment of these technologies across complex manufacturing systems (Lee & Su, 2025). Furthermore, AI-generated and optimized composite designs underscore the growing transformative power of generative AI in manufacturing design processes.

AI Readiness and Adoption Assessment Framework

The AI Readiness and Adoption Assessment Framework is structured into seven sequential stages, beginning with defining the scope and objectives and concluding with the creation of standardized tools to ensure consistency in future assessments. Each stage builds upon the previous one, forming a continuous cycle of evaluation, planning, execution, and improvement. The stages move from understanding the organization's current AI maturity to discovering and prioritizing use cases, closing capability gaps, creating a roadmap, implementing governance, and equipping teams with reusable instruments. Table 6.2 summarizes these stages, highlighting their numbering, names, and key areas of focus.

Scope and Objectives (Pre-Work)

Before any AI assessment can begin, the company must establish a clear strategic intent for AI adoption. This phase sets the boundaries, goals, and decision-making authority for the entire process. Without a well-defined scope and stakeholder alignment, assessments risk producing generic or misaligned recommendations. This stage

TABLE 6.2 The AI Readiness and Adoption Assessment Framework structure

Stage	*Updated Label*	*Description (with Business Impact)*
Stage 1	Assess the Current State	Evaluate organization's AI maturity across eight dimensions (e.g., data, talent, governance, culture). Provides a diagnostic baseline that enables leadership to prioritize investments and transformation pathways (Silva & Alahakoon, 2021)
Stage 2	Define Strategic Alignment	Clarify how AI initiatives fit into organizational goals and digital transformation agendas. Ensures resources are allocated to projects with measurable business value (Fountaine et al., 2019)
Stage 3	Define Metrics	Identify KPIs and success criteria for AI initiatives. Links AI outcomes directly to value creation and ROI tracking for executives.
Stage 4	AI Adoption Roadmap	Develop talent, infrastructure, and governance mechanisms for AI deployment. Strengthens organizational capacity and reduces barriers to scaling AI solution (Mayer et al., n.d.)
Stage 5	Pilot & Scale	Test AI projects in controlled environments and progressively expand successful use cases. Drives organizational learning, accelerates adoption, and reduces transformation risk.
Stage 6	Instruments & Templates	Create standardized tools and repeatable methods for AI readiness evaluation. Provides consistent benchmarks across units, reducing costs of adoption and enhancing predictability (Accenture, 2025).
Stage 7	Continuous Improvement & Transformation	Embed AI readiness assessment into ongoing transformation processes. Supports long-term agility, business innovation, and resilience in dynamic markets (Accenture, 2025).

Source: Authors' own elaboration, based on Reim et al. (2020), Silva & Alahakoon (2021), Fountaine et al. (2019), Mayer et al. (n.d.), and Accenture (2025).

ensures everyone understands *why* the company is pursuing AI, *what* will be included, and *how* success will be measured (Reim et al., 2020).

Objective

Align on strategic goals, success KPIs, and boundaries before assessment begins.

Activities

- Conduct executive interviews to clarify AI-related business outcomes.
- Define which domains, processes, or products will be included.
- Identify constraints (budget, compliance, market deadlines).

Outcome

- Assessment Charter
- Stakeholder Map
- Scope and Timeline Document

Reflection on Business Impact

A clear scope ensures that resources are used effectively and prevents scope creep, aligning the AI assessment with strategic goals from the outset.

How It Can Be Achieved

We hold structured kickoff workshops with the leadership, utilizing predefined templates to identify priorities and constraints.

Stage 1: Assess the Current State (Eight Dimensions)

This stage assesses the company's current readiness for AI across multiple capability areas, including strategy, data, technology, people, processes, governance, operations, and innovation. The aim is to establish a factual baseline using both quantitative maturity scoring and qualitative evidence gathering (Hiniduma et al., 2024). This creates a shared, evidence-based understanding that informs the gap analysis and roadmap phases.

Objective

Establish a baseline of the company's AI maturity across strategy, data, technology, people, processes, governance, operations, and innovation.

Activities

- Score each dimension (0–5 maturity scale).
- Collect evidence through workshops, data audits, policy reviews, and architecture analysis.
- Apply Artificial Intelligence Data Readiness and Integration Network (AIDRIN) for data readiness, Machine Learning Operations (MLOps) maturity checks, and governance audits.

Outcome

- AI Readiness Heatmap
- Detailed Evidence Notes for each dimension

Reflection on Business Impact

It provides a factual, shared view of where the organization stands today, avoiding assumptions and setting a realistic foundation for AI planning.

How It Can Be Achieved

By combining quantitative surveys with qualitative interviews and document reviews to ensure an evidence-based maturity score.

Stage 2: Define Strategic Alignment

Once the current state is clear, the company can identify high-value AI opportunities. This stage focuses on gathering ideas, evaluating their alignment with business strategy, and prioritizing them based on feasibility, data readiness, and risk. Prioritization ensures that early efforts deliver visible wins while laying the groundwork for more complex, strategic initiatives.

Objective

Identify and rank AI opportunities that align with strategy, feasibility, and readiness.

Activities

- Conduct workshops to gather ideas for AI use cases.
- Score each on value, feasibility, data fitness, risk, and time-to-impact.
- Apply Technology–Organization–Environment (TOE)–DOI (TOE–DOI) considerations for SMEs and resource-constrained contexts.

Outcome

- Prioritized Use-Case Portfolio
- Quick Wins, Mid-Term, and Strategic Bet Categorization

Reflection on Business Impact

Focuses energy on projects with the highest likelihood of measurable success, creating early momentum and organizational buy-in.

How It Can Be Achieved

By using a scoring matrix in collaborative tools (e.g., Miro, Excel) so stakeholders can visually see and agree on priorities.

Stage 3: Define Metrics

This stage translates the maturity assessment into a targeted improvement plan. For each capability area, the team compares current performance with desired future states, identifies gaps, and proposes actionable steps. The output is not just a list of weaknesses but a prioritized set of interventions with clear ownership and measurable outcomes.

Objective

Compare current maturity levels with target goals, identify capability gaps, and propose targeted actions.

Activities

- For each dimension, document the current score, target score, and gaps.
- Recommend people, process, and technology interventions.
- Assign ownership, timelines, and KPIs for tracking progress.

Outcome

- Gap Analysis Matrix
- Recommendations Report (by dimension)

Reflection on Business Impact

Turns assessment findings into an actionable improvement plan, ensuring investments target the most critical capability gaps.

How It Can Be Achieved

By mapping each gap to a clear action plan using a RACI chart to clarify responsibilities and timelines.

Stage 4: AI Adoption Roadmap

The roadmap converts recommendations into a phased execution plan. It accounts for dependencies, risks, and business priorities, ensuring the organization moves from pilot projects to scaled AI capabilities systematically. This stage also secures leadership commitment by providing a clear investment plan and timeline.

Objective

Develop a phased plan for the adoption and scaling of AI.

Activities

- Organize initiatives into three phases: Foundation, Scale, and Optimize.
- Sequence actions, considering dependencies, risk, and ROI.
- Create RACI charts and investment plans.

Outcome

- AI Adoption Roadmap (with phases, milestones, KPIs)
- Resource Allocation Plan

Reflection on Business Impact

Gives decision-makers a realistic, time-bound execution plan that balances speed with risk management, maximizing ROI over time.

How It Can Be Achieved

By using Gantt charts or roadmap software (e.g., MS Project) to visualize dependencies and timelines.

Stage 5: Pilot & Scale

AI initiatives require ongoing monitoring and adaptation. This stage embeds governance structures, KPIs, and review processes into the organization's operating rhythm. It ensures AI systems remain compliant, performant, and aligned with business objectives over time.

Objective

Ensure continuous tracking of AI performance, risk, and value realization.

Activities

- Define operational, business, and responsible AI KPIs.
- Set cadence for steering committee reviews and quarterly reassessments.
- Implement continuous improvement loops for data, models, and processes.

Outcome

- Governance Framework
- KPI Dashboard
- Quarterly Maturity Review Process

Reflection on Business Impact

It fosters transparency, accountability, and adaptability, ensuring that AI initiatives remain aligned with business priorities as conditions change.

How It Can Be Achieved

By implementing KPI dashboards in BI tools (e.g., Power BI, Tableau) linked to live operational and model-monitoring data.

Stage 6: Instruments and Templates

Standardized tools accelerate assessment and ensure consistency. This stage builds and deploys reusable templates, checklists, and prioritization frameworks, reducing the time and effort needed to repeat the AI assessment process across different units or projects.

Objective

Provide standardized tools to streamline the assessment and execution process.

Activities

- Develop ready-to-use checklists (Data Readiness, MLOps Maturity, GRC).
- Build prioritization matrices and benefits tracking sheets.
- Provide workshop templates and report formats.

Outcome

- Complete AI Assessment Toolkit
- Standardized Evaluation and Reporting Templates

Reflection on Business Impact

The system reduces assessment time, improves consistency, and enables repeatable evaluation across business units or subsidiaries.

How It Can Be Achieved

By storing all tools in a centralized repository (e.g., SharePoint, Notion) and training teams in their consistent application.

Methodology Note

The framework is developed as a conceptual synthesis grounded in an extensive review of recent peer-reviewed literature (2020–2025) and leading industry reports. This integrative approach consolidates dispersed knowledge into a coherent structure, ensuring both theoretical rigor and practical relevance. Future research directions include applying the framework across industries to enable empirical validation and generate comparative insights.

The technical models referenced in this study build on recent benchmarks. YOLOv8, the latest addition to the YOLO family, is primarily evaluated on the COCO (Common Objects in Context) dataset for large-scale object detection. Multi-object tracking methods such as ByteTrack and BOT-SORT, both introduced in 2022, have set state-of-the-art performance standards on MOTChallenge datasets (Y. Zhang et al., 2022; Aharon et al., 2022). GRU remains a widely applied recurrent model for

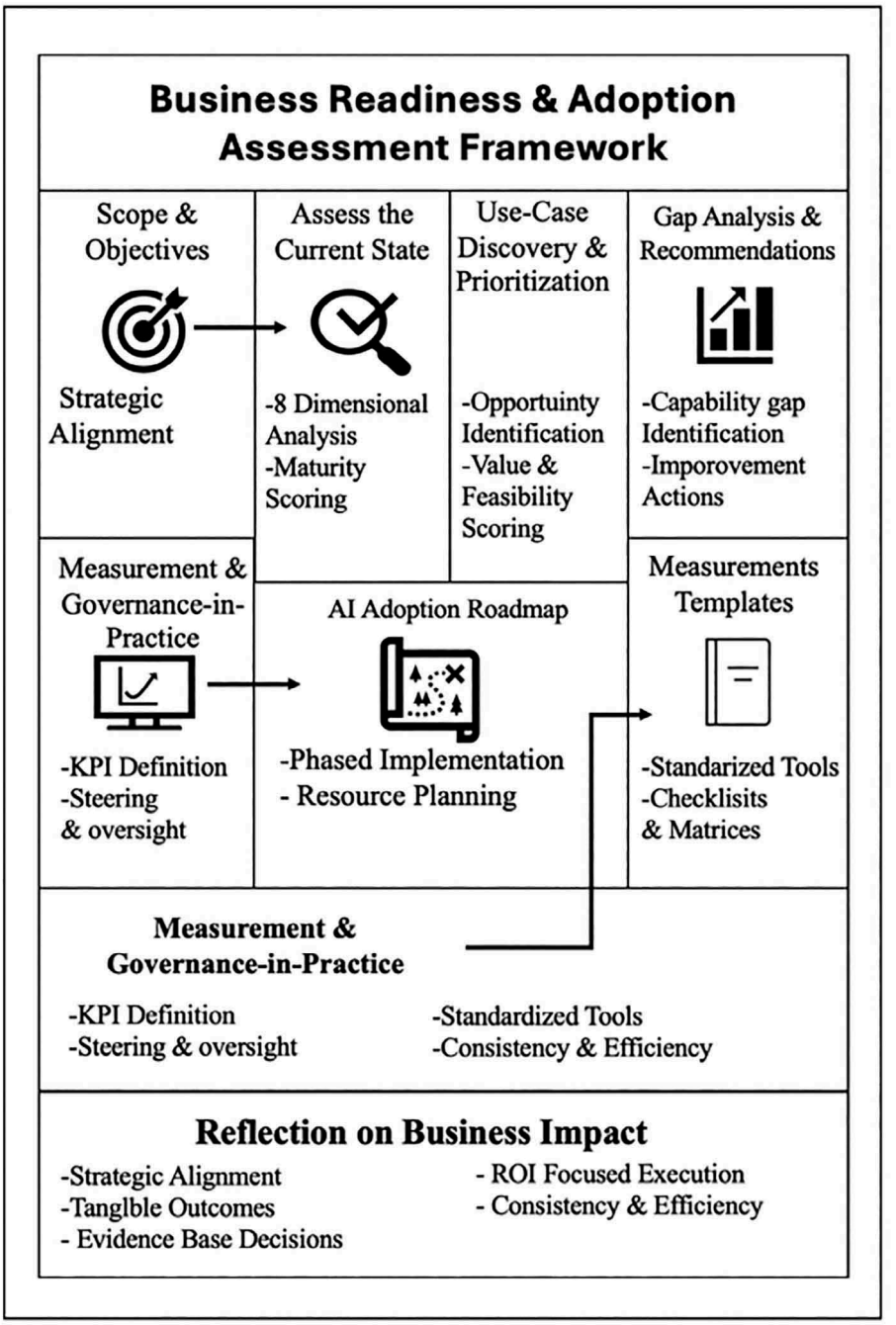


FIGURE 6.1 AI Readiness and Adoption Assessment Framework
Source: Authors' own elaboration, based on Reim et al. (2020), Silva & Alahakoon (2021), Fountaine et al. (2019), and Accenture (2025).

time-series forecasting, although newer transformer-based approaches are increasingly adopted. In this chapter, these models are used as illustrative references rather than implementations with new proprietary datasets.

Summary

The AI Readiness and Adoption Assessment Framework as shown in Figure 6.1 is a structured methodology designed to guide organizations through successful AI-driven transformation. It recognizes that effective AI adoption requires more than technology—it demands alignment of strategy, capabilities, and governance. The framework comprises seven sequential stages that move from defining intent to enabling repeatable execution.

Stage 0 sets the foundation by defining scope, objectives, stakeholders, and success criteria. Stage 1 assesses the current state across eight dimensions—strategy, data, technology, people, processes, governance, operations, and innovation—to establish a factual baseline. Stage 2 identifies and prioritizes AI use cases based on value, feasibility, and risk. Stage 3 conducts a gap analysis to compare current maturity with target levels and recommend capability improvements. Stage 4 develops a phased adoption roadmap linking actions, milestones, and KPIs. Stage 5 embeds measurement and governance to ensure ongoing performance, compliance, and value realization. Finally, Stage 6 delivers standardized instruments and templates to ensure consistent, repeatable assessments.

By combining careful planning with real-world actions, the framework helps organizations connect their AI projects with business goals, fill skill gaps, and make the best use of resources. Its structured approach ensures scalability, sustainability, and long-term impact, making it a comprehensive tool for guiding AI readiness and adoption efforts.

References

Aagaard, A., & Tucci, C. (2024). AI-Driven Business Model Innovation: Pioneering New Frontiers in Value Creation. In A. Aagaard (Ed.), *Business Model Innovation: Game Changers and Contemporary Issues* (pp. 295–328). Springer International Publishing. doi:10.1007/978-3-031-57511-2_10.

Accenture. (2025). Front-Runners' Guide to Scaling AI. https://www.accenture.com/us-en/insights/data-ai/front-runners-guide-scaling-ai.

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking (arXiv:2206.14651). arXiv. doi:10.48550/arXiv.2206.14651.

Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2024). Methodological Approach to Assessing the Current State of Organizations for AI-Based Digital Transformation. *Applied System Innovation*, 7(1), Article 1. doi:10.3390/asi7010014.

Anuyah, S., Singh, M. K., & Nyavor, H. (2024). Advancing Clinical Trial Outcomes Using Deep Learning and Predictive Modeling: Bridging Precision Medicine and Patient-Centered Care. *World Journal of Advanced Research and Reviews*, 24(3), 1–25. doi:10.30574/wjarr.2024.24.3.3671.

Awosika, T., Shukla, R. M., & Pranggono, B. (2023). Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection (arXiv:2312.13334). arXiv. doi:10.48550/arXiv.2312.13334.

Bartosiewicz, B. (2024). Countering Counter-Urbanization: A Comparative Analysis of Resiliency and Redevelopment in Postindustrial Cities of the United States and Poland | Springer Nature Link. https://link.springer.com/chapter/10.1007/978-981-99-5744-6_28.

Bawack, R. E., Wamba, S. F., Carillo, K. D. A., & Akter, S. (2022). Artificial Intelligence in E-Commerce: A Bibliometric Study and Literature Review. *Electronic Markets*, 32(1), 297–338. doi:10.1007/s12525-022-00537-z.

Brătucu, G., Ciobanu, E., Chițu, I. B., Litră, A. V., Zamfirache, A., & Bălășescu, M. (2024). The Use of Technology Assisted by Artificial Intelligence Depending on the Companies' Digital Maturity Level. *Electronics*, 13(9), 1687. doi:10.3390/electronics13091687.

Brynjolfsson, E. & McAfee, A. (2017,July 18). The business of artificial intelligence: What it can—and cannot—do for your organization. *Harvard Business Review*. https://hbr.org/2017/07/the-business-of-artificial-intelligence.

Butler, T., Espinoza-Limón, A., & Seppälä, S. (2023). Towards a Capability Assessment Model for the Comprehension and Adoption of AI in Organizations (arXiv:2305.15922). arXiv. doi:10.48550/arXiv.2305.15922.

Cao, B., Wang, S., Lin, X., Wu, X., Zhang, H., Ni, L. M., & Guo, J. (2025). From Deep Learning to LLMs: A survey of AI in Quantitative Investment (arXiv:2503.21422). arXiv. doi:10.48550/arXiv.2503.21422.

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors. *Strategy Science*, 9(4), 322–345. doi:10.1287/stsc.2024.0190.

Davenport, T. H. & Ronanki, R. (2018). Artificial Intelligence for the Real World. *Harvard Business Review*, 96*(1), 108–116*.

Deloitte (2024a). AI in Post-Sales Customer Experience | Deloitte US. Retrieved September 22, 2025, from https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/articles/ai-in-customer-experience.html.

Deloitte (2024b). Earning trust as gen AI takes hold: 2024 Connected Consumer Survey. Deloitte Insights. Retrieved September 22, 2025, from https://www.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey.html.

Diallo, K., Smith, J., Okolo, C. T., Nyamwaya, D., Kgomo, J., & Ngamita, R. (2025). Case Studies of AI Policy Development in Africa. *Data & Policy*, 7, e15. doi:10.1017/dap.2024.71.

Dresser, S. (2025, July 1). Amazon Launches a New AI Foundation Model to Power Its Robotic Fleet and Deploys Its 1 Millionth Robot. https://www.aboutamazon.com/news/operations/amazon-million-robots-ai-foundation-model.

Durham, J. (2025). Amazon builds first foundation model for multirobot coordination. Amazon Science. https://www.amazon.science/blog/amazon-builds-first-foundation-model-for-multirobot-coordination

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. https://hbr.org/2019/07/building-the-ai-powered-organization?utm_source=chatgpt.com.

Gkika, E. C., Kargas, A., Salmon, I., & Drosos, D. (2025). Unveiling Digital Maturity: Key Drivers of Digital Transformation in the Greek Business Ecosystem. *Administrative Sciences*, 15(3), 96. doi:10.3390/admsci15030096.

Hiniduma, K., Byna, S., Bez, J. L., & Madduri, R. (2024). AI Data Readiness Inspector (AIDRIN) for Quantitative Assessment of Data Readiness for AI. *Proceedings of the 36th International Conference on Scientific and Statistical Database Management*, 1–12. doi:10.1145/3676288.3676296.

Hossam, A., Ramadan, A., Magdy, M., Abdelwahab, R., Ashraf, S., & Mohamed, Z. (2024). Revolutionizing Retail Analytics: Advancing Inventory and Customer Insight with AI (arXiv:2405.00023). arXiv. doi:10.48550/arXiv.2405.00023.

Javed, H., Goncalves, M., & Thirunavukkarasu, S. (2025). Innovative Pathways: Leveraging AI Adoption and Team Dynamics for Multinational Corporation Success. *Businesses*, 5(3), 28. doi:10.3390/businesses5030028.

Jeong, C., Sim, S., Cho, H., Kim, S., & Shin, B. (2025). E2E Process Automation Leveraging Generative AI and IDP-Based Automation Agent: A Case Study on Corporate Expense Processing (arXiv:2505.20733). arXiv. doi:10.48550/arXiv.2505.20733.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors. *Business & Information Systems Engineering*, 63(1), 5–20. doi:10.1007/s12599-020-00676-7.

JPMorgan Chase. (2025). JPMorgan's COiN (Contract Intelligence) Platform: Using AI in Mergers & Acquisitions and Commercial Lending. Emre Ates. https://www.emreates.co.uk/research-2/jpmorgan's-coin-(contract-intelligence)-platform%3A-using-ai-in-mergers-%26-acquisitions-and-commercial-lending?utm_source=chatgpt.com.

Kant, S., Deepika, & Roy, S. (2025). Artificial Intelligence in Drug Discovery and Development: Transforming Challenges into Opportunities. *Discover Pharmaceutical Sciences*, 1(1), 7. doi:10.1007/s44395-025-00007-3.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in My Hand: Who's the Fairest in the Land? on the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons*, 62(1), 15–25. doi:10.1016/j.bushor.2018.08.004.

Kumbhar, A., Chougule, A., Lokhande, P., Navaghane, S., Burud, A., & Nimbalkar, S. (2023). DeepInspect: An AI-Powered Defect Detection for Manufacturing Industries (arXiv:2311.03725). arXiv. doi:10.48550/arXiv.2311.03725.

Lee, J., & Su, H. (2025). Rethinking industrial artificial intelligence: A unified foundation framework. *International Journal of AI for Materials and Design*, 2(2), 56. doi:10.36922/IJAMD025080006.

Mayer, H., Yee, L., Chui, M., & Roberts, R. (n.d.). AI in the Workplace: A Report for 2025 | McKinsey. Retrieved September 25, 2025, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work?utm_source=chatgpt.com.

McKinsey. (2024). Succeeding in the AI Supply-Chain Revolution | McKinsey. https://www.mckinsey.com/industries/metals-and-mining/our-insights/succeeding-in-the-ai-supply-chain-revolution.

Michael, C., Lareina, Y., Bryce, H., Alex, S., Alexander, S., & Company, M. (2023). The State of AI in 2023: Generative AI's Breakout Year. *NASWA Workforce Technology*. https://library.naswa.org/doi/10.5555/20.500.11941/5019.

Oyelade, K., & Yusuff, M. (2025, February 12). AI-Driven Fraud Detection in Fintech: Enhancing Security and Customer Trust. https://www.researchgate.net/publication/389354397_AI-Driven_Fraud_Detection_in_Fintech_Enhancing_Security_and_Customer_Trust.

Patil, D. (2024). Artificial Intelligence-Driven Predictive Maintenance in Manufacturing: Enhancing Operational Efficiency, Minimizing Downtime, and Optimizing Resource Utilization (SSRN Scholarly Paper 5057406). doi:10.2139/ssrn.5057406.

Reim, W., Åström, J., & Eriksson, O. (2020). Implementation of Artificial Intelligence (AI): A Roadmap for Business Model Innovation. *AI*, 1(2), 180–191. doi:10.3390/ai1020011.

Sadiq, R. B., Safie, N., Abd Rahman, A. H., & Goudarzi, S. (2021). Artificial Intelligence Maturity Model: A Systematic Literature Review. *PeerJ Computer Science*, 7, e661. doi:10.7717/peerj-cs.661.

Samadi, H., Ahsan, M. M., & Raman, S. (2025). Hybrid Machine Learning Framework for Predicting Geometric Deviations from 3D Surface Metrology (arXiv:2508.06845). arXiv. doi:10.48550/arXiv.2508.06845.

Sheehan, E., Natalie Martini, Brian McCarthy, Lupine Skelly, Bryn Walton, & Oliver Vernon-Harcourt. (n.d.). (2026) Retail Industry Global Outlook. Deloitte Insights. Retrieved April 18, 2026, from https://www.deloitte.com/us/en/insights/industry/retail-distribution/retail-distribution-industry-outlook.html.

Shin, Y., Lee, M., Lee, Y., Kim, K., & Kim, T. (2025). Artificial Intelligence-Powered Quality Assurance: Transforming Diagnostics, Surgery, and Patient Care—Innovations, Limitations, and Future Directions. *Life*, 15(4), 654. doi:10.3390/life15040654.

Silva, D. D., & Alahakoon, D. (2021). An Artificial Intelligence Life Cycle: From Conception to Production (arXiv:2108.13861). arXiv. doi:10.48550/arXiv.2108.13861.

Singh, N., Chaudhary, V., Singh, N., Soni, N., & Kapoor, A. (2024). Transforming Business with Generative AI: Research, Innovation, Market Deployment and Future Shifts in Business Models (arXiv:2411.14437). arXiv. doi:10.48550/arXiv.2411.14437.

Skouloudis, A., & Venkatraman, A. (2025). Scratching the Surface of Responsible AI in Financial Services: A Qualitative Study on Non-Technical Challenges and the Role of Corporate Digital Responsibility. *AI*, 6(8), 169. doi:10.3390/ai6080169.

Space, P. (2025, May 21). How Does Spotify Use AI: Case Study [Substack newsletter]. *Product Space Newsletter*. https://theproductspace.substack.com/p/how-does-spotify-use-ai-case-study.

Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from Artificial-Intelligence-based systems in professional service operations. *Journal of Operations Management*, 68(6–7), 592–618. doi:10.1002/joom.1215.

Stone, M. (2025). Amazon's Robots Could Help It Save $10 Billion a Year by 2030, Morgan Stanley Analysts Say. *Business Insider*. https://www.businessinsider.com/amazon-robots-could-save-10-billion-a-year-2025-2.

Tang, A., Wei, M., Haemel, A., La, C., Sirota, M., & Lee, E. Y. (2025). Artificial Intelligence-Enabled Precision Medicine for Inflammatory Skin Diseases (arXiv:2505.09527). arXiv. doi:10.48550/arXiv.2505.09527.

Theodoridis, G., & Tsadiras, A. (2025). Retail Demand Forecasting: A Comparative Analysis of Deep Neural Networks and the Proposal of LSTMixer, a Linear Model Extension. *Information*, 16(7), 596. doi:10.3390/info16070596.

Vallarino, D. (2025). Detecting Financial Fraud with Hybrid Deep Learning: A Mix-of-Experts Approach to Sequential and Anomalous Patterns (arXiv:2504.03750). arXiv. doi:10.48550/arXiv.2504.03750.

Viorel, P. (2025, February 14). Predictive Maintenance: Cutting Costs & Downtime Smartly. *IIoT World*. https://www.iiot-world.com/predictive-analytics/predictive-maintenance/predictive-maintenance-cost-savings/.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities. *Journal of Business Research*, 70, 356–365. doi:10.1016/j.jbusres.2016.08.009.

Yang, W., Cao, D., & Liu, Y. (2025). Foundation Models for Demand Forecasting via Dual-Strategy Ensembling (arXiv:2507.22053). arXiv. doi:10.48550/arXiv.2507.22053.

Zhang, H., Arvin, C., Efimov, D., Mahoney, M., Perrault-Joncas, D., Ramasubramanian, S., Wilson, A., & Wolff, M. (2024). LLMForecaster: Improving Seasonal Event Forecasts with Unstructured Textual Data. *Amazon Science*. https://www.amazon.science/publications/llmforecaster-improving-seasonal-event-forecasts-with-unstructured-textual-data.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box (arXiv:2110.06864). arXiv. doi:10.48550/arXiv.2110.06864.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART III

AI in Action Across Business Functions: Operational Excellence through Intelligent Automation



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

7

BEYOND THE BUZZ

Practical AI Implementation in Business Functions

Mohamad Abou El Hassan

1 Introduction: The Transformative Imperative of Artificial Intelligence in the World of Business

While the world is witnessing an unprecedented era of technological advancement, a few incidents have profoundly challenged the disruptive and constructive power of artificial intelligence (AI). What was once restricted to the realms of science fiction has rapidly infused the conceptual framework of our daily lives and, more critically, has become a crucial force in reshaping the core of international commerce (Nyberg et al., 2025). AI, in its myriad forms, has emerged from a futuristic concept to a reality. This transformative dynamism is fundamentally reshaping business structures, functional processes, and operational paradigms across numerous industries globally (Samokhvalov, 2024).

The modern business landscape is characterized by unparalleled volatility, uncertainty, complexity, and ambiguity (VUCA) (Makkar & Rani, 2024). In such an environment, relying solely on historical data and human intuition is no longer sufficient to navigate complex challenges and seize new opportunities (Al-Mamary, 2025). This is where AI plays a crucial role, offering powerful tools that can process large datasets, uncover hidden patterns, make highly accurate predictions, and execute complex tasks with speed and precision far beyond human capabilities (Stoyanova, 2023). Whether used as a tool or an agent, AI represents a strategic necessity, driving significant shifts in how enterprises perceive, plan, execute, and sustain operations.

This chapter entitles a comprehensive investigation into the strategic incorporation of AI across various core business areas. The primary objective is to critically examine the theoretical and practical impact of AI on critical decision-making processes, operational efficiency and effectiveness, and the structural trajectories of innovation within enterprises. In other words, it combines the required tools to promote an interchangeable model of innovative entrepreneurship, sustained by profound intrapreneurial principles and values. By adopting a multidisciplinary approach, this chapter will investigate how specific AI tools, such as natural language processing (NLP),

DOI: 10.4324/9781003632429-11
This chapter has been made available under a CC-BY-NC-ND 4.0 license

machine learning (ML), and predictive analytics (PA), are not merely assisting tools but also actively revamping many conventional business practices and projected business affairs. This reform is driven by a pragmatic will to implement structural solutions to persistent business challenges, moving beyond superficial enhancements to effect fundamental changes.

Ultimately, this chapter explores how to associate the indispensable gap between the rapidly advancing theoretical capabilities of AI and their tangible impact on business sustainability. The chapter will illustrate how AI efficiently and effectively minimizes the gap between raw data-driven insights and deliverables, transforming abstract information into distinguished competitive advantages. From customized customer experiences in marketing to optimizing complex supply chains in operations, and from revamping talent management in human resources to bolstering fraud detection in compliance and finance, the unmatched influence of AI is on a solid track to crafting the future of enterprises (Cano-Marin, 2024). Beyond these significant implications, the chapter will also explore AI's transformative impact on marketing, human resources (HR), and finance, as well as on the global efficiency of supply chains across various industries. Moreover, the process indirectly addresses and acknowledges multiple barriers, such as ethical concerns, data privacy, and labor displacement, while promoting a community and environment that supports accountable AI governance and AI literacy (Prieto, 2025).

1.1 The AI Revolution in the Contemporary Business Landscape

The past decade was a turning point in the exponential surge in AI adoption, transforming it from a narrow technological utility into a strategic asset. This rapid propagation is attributable to various converging factors: the unprecedented availability of vast datasets (big data), the significant development in computational power, and the continuous refinement of sophisticated algorithms. At an international scale, businesses are now recognizing that AI is not merely a tool for automation, but a fundamental shift in how value is created, delivered, and sustained (Babashahi, et al., 2024).

The "AI revolution" signifies a departure from traditional, reactive business models towards proactive, predictive, and adaptive ones. For instance, the ability of machine learning algorithms to discern complex patterns in data far beyond human capacity has enabled businesses to anticipate market shifts, forecast demand more accurately, and identify emerging risks before they materialize. This foresight transforms decision-making from an art into a science driven by data (Teodorescu & Ovreiu, 2024).

Furthermore, the captured capacity of AI for automation soars beyond routine tasks. Advanced AI systems, specifically those incorporating NLP and computer vision, can now handle more sophisticated customer inquiries, optimize logistical networks, and even contribute to creative processes. This has led to significant gains in operational efficiency and, in many cases, has redefined the very nature of work within organizations (Teodorescu & Ovreiu, 2024). Accenture has published a report in 2024 indicating that when enterprises integrated AI into their primary operations, they witnessed a tangible increase in productivity that ranged between 15% and 20%, with a significant reduction in operational costs that reached 12% within the first three years of integration (Roark et al., 2025). It is believed that the numbers may vary

across industries, considering the scale of implementation. The prevailing trend points towards substantial economic prosperity.

However, the revolution was not triggered without complexities. The trend of acquiring Artificial Intelligence also requires a critical and scientific examination of its societal implications, including ethical considerations, job displacement, and the need for robust data governance (Babashahi, et al., 2024). Businesses are now more concerned about responsible innovation and productivity across various areas. For instance, enterprises are now seeking to implement sound policies that ensure fairness, increase transparency, and validate accountability in their algorithmic prompts and deployments (Al-Mamary, 2025).

2 The Foundational Pillars of AI in Business

A set of sophisticated tools and instruments sustains the universal effect of AI on modern business functions. These foundational pillars, while differentiated in their conceptual mechanisms, often operate synergistically to automate processes, formulate insights, and optimize decision-making (Butt, 2024). Understanding these core constituents is mandatory for recognizing AI's strategic utility in a business environment. This section of the chapter will explore the primary AI tools—ML, NLP, and PA—illustrating their fundamental principles and relevance to modern, rapidly evolving business operations (Jovanović et al., 2025). The emphasis will outline the evolution and innovations that have led to their widespread adoption.

2.1 Machine Learning (ML) and its Algorithmic Prowess

With minimal or negligible human intervention, ML is a core subset of AI that enables systems to learn from existing, prompted, or generated data, identify patterns, and make remarkable predictions or decisions (Jovanović, 2025). ML is an algorithmic approach that learns from experience, sustainably improving its performance based on the level of exposure to vast datasets, without any explicit instructions, coding, or prompts for every task, as in traditional programming (Teodorescu & Ovreiu, 2024). This learning process typically falls into three primary categories (Lenhof et al., 2024):

- **Supervised Learning:** It involves training a given algorithm on a labeled dataset where both input and output are provided, allowing the model to learn the relative mapping of generated output.
- **Unsupervised Learning:** In contrast, this process addresses analyzing unlabeled data to unveil hidden patterns, relationships, and structures, such as clustering customer segments based on their purchasing behavior without preset categories.
- **Reinforcement Learning:** It enables a prospective agent to learn optimal sequential decisions by exhibiting actions in a corresponding environment to maximize aggregate rewards, commonly exhibited to optimize sophisticated logistical channels or dynamic pricing strategies.

The algorithmic competence of ML enables enterprises to extract unprecedented value from their massive data reserves. The ability of ML to capture significant correlations

and predict future trends converts raw data into actionable, subtle intelligence. For example, in fraud detection, models can identify anomalies in transactional patterns that human specialists might miss in real time (Jovanović et al., 2025). Additionally, ML can enhance predictive maintenance in manufacturing by analyzing data from machinery sensors to predict failures, promote a proactive approach to repairs, and help reduce idle downtime.

Hence, deep learning was founded as a remarkable area of recent development in ML. It is considered a subfield that utilizes artificial neural networks with multiple layers, referred to as "deep networks," to learn from data labeled with various levels of abstraction. This has entangled inventions in speech processing, translating, and interpreting, image recognition, and complex pattern recognition. This finding contributed to the development of complex applications, such as facial recognition for security purposes or automated quality assurance (QA) on manufacturing production lines, by integrating powerful, fast, and efficient graphics processing units (GPUs) with cloud computing resources (Jovanović et al., 2025).

2.2 Natural Language Processing (NLP) for Enhanced Communication and Insights

NLP is another main branch of AI that equips computers to recognize, understand, compile, interpret, generate, and manipulate human language. As enterprises increasingly interact with customers and record massive amounts of textual data (complaints, emails, app chats, social media posts, and documents), NLP has become a central pillar for extracting and generating valid insights and automating language-based tasks. (Mittapally, 2024) NLP algorithms can process natural language in various forms, including text and speech. Its main functional tasks include:

- **Text Classification:** Classifying text into predefined categories and classes (e.g., top-urgent communication, spam detection, feedback analysis of customer reviews).
- **Named Entity Recognition:** Capturing, identifying, and extracting key entities from existing text, such as names, family names, email addresses, organizations, locations, and others.
- **Sentiment Analysis:** Identifying the sentiment or emotional tone transmitted in a unified piece of communication (positive, negative, neutral). This is crucial for understanding customer feedback and for taking the corresponding actions.
- **Language Translation:** Translating and interpreting text, written communication, phone calls, or speech from one language to another.
- **Question Answering and Summarization:** Enabling systems to respond to open questions based on a preset text or to generate meaningful, coherent, and concise feedback and responses.
- **Natural Language Generation (NLG):** Generating human-like text from clustered, organized, or structured data, used in provided reports or personalized communications.

NLP has significant implications for how enterprises communicate both internally and externally, as well as for generating valuable insights from text data (Nguyen et al.,

2023). Internally, the integration of NLP could enhance communication efficiency by performing intelligent document management, automating routine tasks such as meeting transcription and email categorization, and strengthening human capital interactions through chatbots and deployed virtual assistants. Enterprises use NLP to understand customer feedback better, analyze sentiment in reviews and social media interactions, and personalize marketing communications at scale. Moreover, the ability to process and analyze massive volumes of clustered text data, such as customer service meeting transcripts, survey-collected data, and other similar data from digital channels, enables enterprises to uncover patterns and trends, detect urgent issues, and make data-driven decisions with exceptional speed and precision. NLP is becoming an essential pillar of business intelligence strategies in the modern world, optimizing the quality of processes, operations, strategic planning, and customer engagement innovation (Butt, 2024).

2.3 Predictive Analytics (PA) for Foresight and Strategic Decision-Making

PA is a branch of AI that leverages statistical algorithms and machine learning techniques to estimate the probability of future outcomes based on historical and current data, integrating multiple variables (Jovanović et al., 2025). This sophisticated approach surpasses reporting events labeled as "descriptive analytics" and their underlying cause, labeled as "diagnostic analytics" (Razbornik & Todosijević, 2024). It reinforces enterprises with the advantage to project trends, mitigate risks, and make more proactive scientific decisions. The application and integration of this model into the enterprise's core duties requires the fulfillment of all of the following:

- **Data Collection and Preparation:** Collecting, archiving, and storing relevant historical data after cleaning to ensure integral quality (Stoyanova, 2023).
- **Model Development:** Applying advanced statistical models such as regression analysis and time series analysis, or machine learning algorithms such as binomial trees, decision trees, and neural networks to capture patterns and validate relationships within the collected and labeled dataset (Zheng & Brintrup, 2025).
- **Model Validation and Diagnosis:** A fundamental and technical evaluation of a model's precision requires validating its reliability, integrity, and accuracy on a given dataset that has not been previously tested. This process is mandatory to ensure the model's capacity to generalize its adaptation to alien data, thereby affirming its robustness and predictive adaptability in real-world scenarios (Cano-Marin, 2024).
- **Deployment and Monitoring:** Integrating a predictive model into enterprise operations is a multi-layer process that extends beyond its conceptual development. The integration process requires the coherent validation of the model's outputs within adopted workflows, accompanied by continuous monitoring of their performance (Teodorescu & Ovreiu, 2024). This sustainable engagement is crucial to ensure that the model's forecasts provide accurate and relevant output over time, adapting to rapidly changing data and dynamic conditions.
- **Demand Forecasting:** In the contemporary entrepreneurial landscape, accurately forecasting future demand for products and services designates a significant

operational efficiency (Butt, 2024). The ability to predict future demand is not merely an academic exercise; it is a critical function that directly affects inventory management, optimizes production schedules, and reinforces the strategic allocation of resources. Anticipating demand can help organizations proactively avoid the costly pitfalls of stockouts, which can result in lost profitable sales opportunities and customer dissatisfaction (Al-Mamary, 2025).

- **Risk Management:** PA is instrumental in risk measurement, quantification, mitigation, and avoidance. By exploring and inspecting large datasets, financial institutions can accurately assess their standing, as well as that of their clients and counterparty, which may constitute direct or indirect risks (Mettrick, 2024). Additionally, financial institutions can benefit from these processes by assessing clients' creditworthiness, evaluating the likelihood of loan defaults, and detecting fraudulent transactions, thereby protecting their assets and maintaining market integrity and accountability.
- **Consumer Behavior Prediction:** Enterprises deploy predictive analytics to possess and monitor consumer behavior, enabling them to anticipate which ones are likely to churn, purchase new items, or interact and respond to a marketing initiative (Benevity, 2024). This advantage enables enterprises to deploy highly focused, accurately personalized interventions (Ayadi & Achour, 2025). By projecting customer needs and behaviors, enterprises can tailor their interactive offerings and communications to optimize customer loyalty, increase sales, and maximize return on investment.
- **Workforce Planning:** In the field of human resources, predictive analytics is an influential tool for succession and strategic workforce planning. By analyzing data, predictive analytics models can forecast future talent needs, proactively identify elements at risk of attrition, and align staffing levels with projected demand (Joseph, 2025). This advantage enables enterprises to transition from reactive hiring and retention policies to a proactive, scientific approach, while ensuring the right people are in the right roles at the right time to support business growth and preserve a competitive edge.

3 AI's Pervasive Impact across Core Business Functions

Having unveiled the importance of the three main pillars that underpin Artificial Intelligence's core relevance to enterprises and outlined the evolution and innovations that have led to its extensive adoption, this section will elaborate on Artificial Intelligence's capacity to permeate and transform every aspect of a business. From internal operations and administrative tasks to all customer and public interactions, AI is set to fundamentally revolutionize processes, optimize decision-making, and unleash new areas for value creation (Davenport & Ronanki, 2021). This section also provides a detailed investigation of AI's pervasive impact across core business functions, illustrating its strategic deployment with practical applications, significant and relevant data, and comprehensive insights from recent research (Butt, 2024). The aim is to articulate how AI solutions are introduced, not just to solve old problems, but also to create new possibilities that were once impossible.

3.1 Revolutionizing Marketing: Personalized Experiences and Engagement

The marketing landscape for enterprises has undergone a profound transformation, shifting from mass-market campaigns to deeply individualized interactions. AI leads the way at the forefront of this transformation, entangling sophisticated algorithms to optimize customer experience at an unprecedented level (Kumar et al., 2021). By integrating vast amounts of consumer data, AI reshapes raw insights to project individual preferences, anticipate future behaviors, and consequently enhance interaction, communication, and engagement. This improved application of AI surpasses traditional segmentation, enabling hyper-personalization that engages deeply with individual consumers, leverages stronger relationships, and yields superior outcomes (Benevity, 2024).

Impact

The impact on business marketing is versatile, covering every level of the customer experience from awareness to after-sales services and follow-up. It allows the marketing team to override traditional demographics by understanding individual flagships, preferences, and real-time behavior (Kumar et al., 2021). This translates into more relevant communications, optimized campaign performance, and a deeper understanding of the customer's "voice." According to a recent report by *Harvard Business Review*, a survey of 5,000 global consumers conducted by Boston Consulting Group found that more than 80% of respondents seek and expect more personalized experiences. Also, more than 65% of respondents had an unsatisfactory experience with inaccurate, inappropriate, or invasive personalization. The report also showed that enterprises can utilize AI to innovate, develop, and refine personalization, enabling customers to access their products more quickly, easily, and affordably (Abraham & Edelman, 2024).

Practical Applications

- **AI-driven Personalization:** YouTube, Netflix, Amazon, Spotify are some of the trendy and famous categorical engines of entities that managed to reach out to clientele and potential prospects by analyzing historical preferences, saved material, usage, and behavior, and similar user data to render and promote related content, products, or services relating to individual interests. This significantly optimizes user experience and increases cross-selling attempts and upselling opportunities (Mittapally, 2024).
- **Targeted Advertising and Programmatic Purchasing:** One of AI's algorithmic capacities is to optimize ad placements, real-time dynamic and creative variation, bidding strategies, ensuring that ads and promotional material are broadcast to reach out to the most relevant audiences ideally. This process enhances the efficiency and effectiveness of digital advertising expenditure (Abraham & Edelman, 2024).
- **Consumer's Journey Mapping and Optimization:** To map sophisticated customer experience, utilization, and behavior, personalize trajectory and dynamic

satisfaction, capture bottlenecks; AI algorithm processes a significant volume of interaction data across various touchpoints with the customers, such as website visits, social media, and customer service communication, whether by phone or in text (Kordzadeh & Ghasemaghaei, 2021).

- **Content Generation and Curation:** Generative AI tools (as referenced in Section 2.2, Natural Language Processing (NLP) for Enhanced Communication and Insights) can help in drafting marketing material, social media content, digital newsletters, and even customized video scripts, while accelerating the creation and ensuring integrity and consistency (Jovanović et al., 2025).
- **Sentiment Analysis and Brand Monitoring:** NLP Generative AI tools consistently monitor mega platforms over social media, digital news, blogs, and vlogs, and review sites to identify significant public sentiment towards a given brand, product, or campaign, enabling rapid interaction and response to critical reputational issues or identification of evolving trends (Davenport & Ronanki, 2021).

Recommended Case Studies

Netflix, YouTube, Spotify, Starbucks, and Sephora are all good examples and remarkable case studies. Sophisticated ML algorithms are integrated behind every page, click, and button to help the website or mobile app deliver the virtual shelves for exploration, checking, and requesting its offerings and programs. The systems are restlessly active, optimizing and refining the user experience, in many cases accommodating more than 80% of the communicated and accessed material and content (Abraham & Edelman, 2024).

3.2 Optimizing Operations: Efficiency, Resilience, and Supply Chain Excellence

Operations, the backbone of enterprises, encompass the processes that convert business inputs into demanded outputs. The role of AI extends beyond conventional automation to optimize efficiency, precision, speed, accountability, integrity, and accuracy in complex transactions and processes (Stoyanova, 2023). Moreover, AI has contributed to an era of intelligent automation that minimizes costs, bias, and systemic redundancies, while profoundly enhancing supply chain resilience, improving production cycles, and achieving record production levels to outperform human capabilities (Al-Mamary, 2025). Day after day, the strategic integration of AI into business is becoming a necessity to minimize waste, control spending, utilize resources efficiently, prevent and mitigate outages, and achieve seamless service delivery and cost savings (Teodorescu & Ovreiu, 2024).

Impact

AI has contributed to a shift in how business is conducted, promoting proactive decision-making and problem-solving over the conventional reactive approach. For instance, AI technology can process and analyze large volumes of operational data and inputs, including those generated by production machinery, logistics information, and supply chain metrics (Zheng & Brintrup, 2025). This breakthrough has helped

businesses to plan realistic financial budgets, identify weaknesses and inefficiencies, anticipate equipment failures, optimize resource allocation, and streamline workflows. The impact and outcomes enhanced output levels, lowered operational costs, improved quality control, and optimized supply chain agility (Sharma, 2025). In increasingly globalized and volatile markets, AI solutions have become a critical pillar for maintaining a competitive edge and embracing entrepreneurial resilience. AI has reframed the modern business yards to adopt smart structural alteration by embracing new ways of doing things smarter—not only faster.

Practical Applications

- Predictive Maintenance: AI technology helps detect potential failures before they occur by analyzing real-time sensor data from deployed robots and machinery. This approach promotes proactive maintenance, prevents high and unplanned costs, minimizes outages and downtime, and extends the planned lifespan of assets (Sharma, 2025).
- Robotic Process Automation (RPA) and Intelligent Automation: RPA roles promote automating rule-based and automated digital tasks. Integrating RPA with AI elevates the game to a higher level of development, enabling it to handle more complex cognitive processes, including decision-making, financial planning, problem-solving, and unstructured data. In large corporations, this represents a headache-free solution for automating customer onboarding, invoice processing and auditing, data entry, validation, and compilation across disparate systems (Ayadi & Achour, 2025).
- Inventory Optimization: AI technology contributed to the innovation of algorithms that analyze historical sales data, promotional impacts, seasonal trends, and micro and macro external factors that include the effects of disasters, weather, economic forecasts, wars, and others (e.g., weather, economic forecasts). This technological advancement enabled the prediction of demand, the planning of supply, the optimization of inventory levels, the procurement of raw materials, the hiring of the required human capital, the reduction of holding costs, the minimization of stockouts, and effective waste management (Praveen & Bhati, 2024).
- Supply Chain Resilience and Risk Management: AI powers innovative and advanced models that analyze universal data streams to detect potential supply chain failure or disruptions related to geopolitical events, natural disasters, routing complications, insurance concerns, and supplier solvency issues. This technological advancement enables enterprises to monitor routes proactively, reroute shipments, identify alternative suppliers, and invest in resilient supply networks (Uddin et al., 2025).

Recommended Case Studies

Siemens: The supply chain team at Siemens has successfully integrated AI, using ML, to predict and detect maintenance requirements for gas turbines, building management systems, and train fleets, resulting in reduced unplanned outages and optimized service schedules (Eckl-Dorna, 2024).

3.3 Revolutionizing Human Resources: Talent, Performance, and Workforce Optimization

The role of human resources (HR) has always been limited to administrative duties, tasks, and compliance. The progressive revolution of AI contributed to a profound reshaping of both its theoretical and professional roles (Nurski & Ruer, 2024). AI advanced tools have introduced new, efficient ways and methodologies for acquiring new talents, thereby establishing new paradigms for evaluating human capital. These tools optimize workforce delegation and task management by leveraging sophisticated analytics to enable reliable predictive modeling (Jatobá et al., 2019). Based on this transformation, HR departments will be able to benefit from the ability of AI models to process vast amounts of structured and unstructured data derived from human capital, thereby becoming more objective, scientifically oriented, and integral to HR strategies (Babashahi et al., 2024).

Impact

AI's impact on HR helps redefine the role of enterprises in attracting, developing, managing, and retaining great talent. AI models can help HR departments and professionals shift from conventional roles to strategic partners who lead enterprises to success by capturing the required insights into workforce dynamics, individual employee needs, and potential biases. This strategic transition is considered essential for constructing a resilient, dynamic, and adaptable workforce in an era of rapid economic and technological change (Gunning, 2023). AI enhances its capacity to analyze complex talent data, identify skills gaps, personalize learning pathways, and predict attrition, while ensuring that investments in human capital yield optimal returns. AI models also enable HR departments to use hiring and promotion data to identify indirect biases in processes and also create more fair opportunities.

Practical Applications

- **AI-Powered Talent Acquisition:** AI tools power the automation process of various stages of recruitment. This advantage includes resume screening models to identify the best candidates by analyzing keywords and skills, NLP to examine job descriptions for bias, and responsive chatbots to answer candidate queries and optimize the candidate experience. Some models are even designed to prepare and analyze video interviews for nonverbal cues.
- **Predictive Analytics for Employee Retention and Attrition:** ML models analyze different factors such as performance, remuneration, supervisor's evaluation, tenure, and others. This enables the HR department to identify employees at risk of attrition and proactively engage them with personalized retention strategies, development plans, or special recognition programs.
- **Workforce Planning:** AI models project future talent needs depending on business growth planning, technological developments, and market trends. These models enable the HR department to optimize staffing levels, identify critical roles, and allocate resources across different departments and surging projects.

Recommended Case Study

Cisco, one of the world's leading telecommunications, cybersecurity, and networking corporations, has adopted an innovative approach to integrating technology with human capital management, development, and retention, rewarding its people for the company's high-performing results (Randstad Enterprise, 2019). Cisco possessed a business transformation approach to encourage employee participation rates in its social impact programs. A three-year study concluded that employees who participated in Cisco's social impact program remained with the company longer, received higher bonuses, and received more promotions (Benevity, 2024).

Also, at Cisco, the HR team has been investing in integrating talent analytics processes and HR data tools to better understand how its human capital collaborates and networks. For instance, they launched an initiative titled "Human Network Analysis" that examines how members of different teams communicate to accomplish assigned tasks. For example, by mapping the interactions of a given role both within and beyond its function, the HR team can identify and chart pathways that empower employees to attain greater success. The process also generates predictive analytics and signals to help identify employees likely to experience attrition and leave the company. (Aviana, 2025). The latter enables HR leaders and business line managers to intervene proactively with innovative retention strategies for retaining high-quality human capital (Dhingra et al., 2021).

3.4 Transforming Finance and Accounting: Fraud Detection, Risk Management, and Investment Optimization

The broader pillar and foundational anchor of the business world is the financial sector. When it comes to enterprise financial records, the volume of data is massive and highly susceptible to risk and fraud. In this context, sophisticated AI tools have significantly contributed to effective, efficient, fast, and reliable fraud detection (Mettrick, 2024). Moreover, AI tools are helping more in risk identification, quantification, and assessment, as well as the integral optimization of decision-making when promoting strategic investment opportunities (Hussain et al., 2025). This growing phenomenon is poised to transform traditional fiscal management, reshaping the infrastructure for a more agile financial system characterized by enhanced security, efficiency, and transparency (Garg et al., 2025).

Impact

AI has helped create new tools designed to transform the structure of financial operations by embedding live analysis of transactional data, enabling the immediate identification of anomalies associated with fraudulent indications or emerging risks, and optimizing sophisticated investment strategies that surpass human cognitive capacities. The AI tools possessed a differentiated capacity to process vast, sophisticated, clustered, and disparate variables and datasets (Hussain et al., 2025). These advantageous capacities enabled the AI tools to learn from historical events and patterns, engaging with financial institutions in a renewed, rather than traditional, rules-

based and rigid approach. The new environment prompted various leading financial institutions to explore new markets, optimize existing ones, and reduce economic losses by quantifying associated risks more accurately, optimizing capital allocation, and promoting a more regulatory and legally compliant framework.

Practical Applications

- **Real-time Fraud Detection:** AI tools are supported by algorithms that can analyze transactional data streams in blinks, detecting patterns indicative of fraudulent behaviors or activities such as suspicious spending patterns, unusual locations, or high-velocity transactions. These algorithmic modules can flag or restrict transactions in real time, thereby tangibly reducing financial losses from electronic payment fraud, money laundering, terrorism financing, or cyber fraud (Garg et al., 2025).
- **Credit Scoring and Loan Underwriting:** AI tools are becoming more capable of assessing the creditworthiness of commercial (Legal) or individual (Moral) credit facility applicants. The enhanced modules of artificial intelligence can analyze a broader range of data points and variables than traditional methods, including gender, age, spending behavior, reputation, business sector, social media data and reviews, professional history, and many others. This process enables achieving higher accuracy in risk assessments, faster approvals, and potentially more secure lending (Lining et al., 2023).
- **Algorithmic Trading and Portfolio Optimization:** AI tools led to enhancing existing and establishing new algorithms that enable the investment analysts, wealth and portfolio managers, advisors, and consultants to execute high-speed trades with protected parameters, identify arbitrage opportunities, and optimize the return of investment portfolios based on accurate predictive market analysis, steady risk tolerance, and real-time market turbulence. The new approach created a satisfactory, mutually beneficial environment, delivering superior returns and efficient risk management in dynamic, volatile markets (Cohen, 2022).
- **Anti-Money Laundering (AML) and Compliance:** Integrating AI tools into a dynamic database and advanced systems that can detect suspicious activities to monitor vast amounts of financial transactions that might involve money laundering or terrorist financing. By identifying complex, layered patterns, AI tools enhance the effectiveness of compliance units and minimize false positives (Khan et al., 2024).
- **Risk Management and Stress Testing:** Risk management is becoming one of the most sophisticated, critical, and growing concerns for financial institutions at a global level. Deploying AI models can help simulate various, yet complex, market scenarios and quantify the potential impact on investment portfolios, enabling leading institutions to assess, manage, and optimize their exposure to credit risk, market risk, operational risk, and systemic risk (Lining et al., 2023).

Recommended Case Study

PayPal & Visa: PayPal and Visa are integrating AI-powered security models and solutions as the financial services industry considers AI as a security hazard and

increasingly enables AI to manage security risks. They employ advanced algorithms to detect real-time fraud by analyzing a massive volume of transactions to identify and prevent suspicious and fraudulent activities (Adams, 2025).

4 Conclusion: Embracing the Challenges and Ethical Factors of Implementing Artificial Intelligence in Modern Business

Integrating AI into enterprise operations has become a vital necessity for health projection, sustainability, resilience, and growth. Nevertheless, the adoption of AI arms and tools cascades a vast list of challenges and complex ethical and social considerations. Discussions of the latter controversial topics and associated governance are deemed mandatory to ensure a smooth deployment with respect to social and ethical sustainability and responsibility. The discussions have always culminated in vicious loops addressing the normalization of potential job displacements, paramount ethical complications, and unmatched data privacy considerations that required advanced handling, responsible governance, and respect for widely accepted moral principles.

Adopting AI in businesses poses a significant challenge and resistance to change for professionals. Algorithmic bias raises numerous sophisticated ethical concerns, as AI systems can inadvertently discriminate against specific outcomes arising from implicit bias in training data (Kordzadeh & Ghasemaghaei, 2021). Interpreting incomplete, partial, or weak datasets during the training of AI models can reveal biases that resurrect a substantial risk, leading to discriminatory or invalid conclusions (Jospeh, 2025). Executives may confront ethical puzzles when relying on AI models to perform intended processes and deliver desired outcomes. The dilemma arises from the natural and logical integration of the learning cycle, as evidenced by the transferable biases inherited by human resources, experience, and cultural differentiation (Razbornik & Todosijević, 2024).

Addressing the sophisticated challenges, incorporating ethics, employment, privacy, and many others, has been a significant concern for policymakers and business leaders. The adoption of AI in modern business requires solid frameworks that respect widely accepted ethical and moral principles and values, ensuring clear accountability and wise, sustainable governance.

Overseeing and controlling AI-generated outcomes requires direct human intervention, arising from the associated risks that encompass all random biases that could combine to produce hazardous outcomes. These chains of uncontrolled biases can accumulate, creating tangible drawbacks that undermine an enterprise's values. An enhanced governance structure and adequate supervision can always mitigate potential risks and unleash the potential that leads to sustainable development. In this field, studies, forums, and constructive projects involving experts from diverse backgrounds in technology, ethics, the social sciences, and law could best help minimize adverse events and promote a promising, responsible, and sustainable environment for development and growth.

References

Abraham, M., & Edelman, D. C. (2024). Personalization Done Right. *Harvard Business Review*, 102(6), 104–115. Retrieved June 19 2025.

Adams, J. (2025, July). PayPal and Visa Embrace AI for Payment Crime Fighting. Retrieved from American Banker: Amercianbanker.com

Adelakun, B. O. (2022, October 23). Ethical Considerations in the Use of AI for Auditing: Balancing Innovation and Integrity. *European Journal of Accounting, Auditing and Finance Research*, 10(12), 31–108. doi:10.37745/ejaafr.2013/vol10n1291108.

Al-Mamary, Y. (2025). The Transformative Power of Artificial Intelligence in Entrepreneurship: Exploring AI's Capabilities for The Success of Entrepreneurial Ventures. *Future Business Journal*, 11(1). doi:10.1186/s43093-025-00533-7.

Aviana. (2025). Cisco. Retrieved from Aviana.com: https://avianaglobal.com/portfolio/cisco/.

Ayadi, S., & Achour, S. (2025, June). The Role of Intelligent Process Automation in Modern Marketing Strategies. *Journal of Administrative and Financial Science*, 9(1), 52–69. Retrieved July 21, 2025.

Babashahi, L., Barbosa, C. E., Lima, Y., Lyra, A., Salazar, H., Argôlo, M., … Souza, J. M. (2024, June). AI in the Workplace: A Systematic Review of Skill Transformation in Tte Industry. *Administrative Sciences*, 14–127. doi:10.3390/admsci14060127.

Benevity. (2024, October 22). The Business Case for Social Impact. Retrieved July 24, 2025, from Benevity.com: https://hub.benevity.com/cisco-social-impact-business-case.

Butt, J. S. (2024). AI-Driven Economies: How Artificial Intelligence (AI) is Reshaping Business Administration and Decision-Making. *Acta Universitatis Danubius: Œconomica*, 20(6), 106–120. Retrieved July 29, 2025.

Cano-Marin, E. (2024, May). The Transformative Potential of Generative Artificial Intelligence (GenAI) in Business: A Text Mining Analysis on Innovation Data Sources. *Esic Market Economics and Business Journal*, 55(2). Retrieved August 10, 2025, from https://orcid.org/0000-0002-7948-1657.

Cohen, G. (2022, February 10). Artificial Intelligence in Trading the Financial Markets. *International Journal of Economics and Business Administration*, 10(1), 101–110. doi:10.35808/ijeba/751.

Davenport, T., & Ronanki, R. (2021). Artificial Intelligence for the Real World. *Harvard Business Review*, Special Issue (Winter), 44–53. Retrieved March 16, 2025.

Dhingra, N., Samo, A., Schaninger, B., & Schrimper, M. (2021). Help Your Employees Find Purpose or Watch Them Leave. McKinsey & Company. Retrieved July 24, 2025, from https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/help-your-employees-find-purpose-or-watch-them-leave#/.

Eckl-Dorna, W. (2024, July 8). Siemens Energy Sees Boost From Data Centers Supplying Own Power. (Bloomberg) Retrieved July 1, 2024, from Bloomberg.com

Garg, A., Singh, M., & Kumar, M. (2025, February). The Intersection of Quantum Computing, Artificial Intelligence and Financial Risks: A Bibliometric Analysis of the Modern Financial Sector. *Journal of Information Technology Management*, 17 (Special Issue), 3–23. doi:10.22059/jitm.2025.100694.

Gunning, J. (2023). Embracing AI Needed To Shape Future Workforce: Emerging Tech Will Bring Revolution in Basic Skills. *Journal of Business*, 38(24), 8. Retrieved April 6, 2025.

Hussain, H., Ramzan, F., Hamza-Khan, A., & Khan, S. (2025, January). Examining the Role of AI in Detecting Fraudulent Activities, Managing Financial Risks, and Enhancing Decision-Making in Investment Strategies. *Review of Applied Management and Social Sciences (RAMSS)*, 8(1), 207–221. doi:10.47067/ramss.v8i1.450.

Jatobá, M., Santos, J., Gutierriz, I., Moscon, D., Fernandes, P. O., & Teixeira, J. P. (2019). Evolution of Artificial Intelligence Research in Human Resources . *Procedia Computer Science*, 164, 137–142. doi:10.1016/j.procs.2019.12.165.

Joseph, E. (2025). Leveraging AI to Inspire Innovation in Traditional and Digital Business Ecosystems. *Journal of Business Ecosystems*, 6(1), 1–18. doi:10.4018/JBE.383049.

Jovanović, S., Dražeta, L., Dražeta, B., & Petrović, A. (2025, April). Business Perspective on Exploring the Intriguing Vistas of Artificial Intelligence. *European Journal of Applied Economics*, 22(1), 1–17. doi:10.5937/EJAE22-55443.

Khan, H., Malik, M.-Z., & Nazir, S. (2024). Identifying The AI-Based Solutions Proposed for Restricting Money Laundering in Financial Sectors: Systematic Mapping. *Applied Artificial Intelligence*, 38(1). doi:10.1080/08839514.2024.2344415.

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic Bias: Review, Synthesis, and Future Research Directions. *European Journal of Information Systems*, 31(3), 388–409. doi:10.1080/0960085X.2021.1927212.

Krause, D. S. (2025, January 25). Addressing the Challenges of Auditing and Testing for AI Bias: A Comparative Analysis of Regulatory Frameworks. p. 42. Retrieved August 15, 2025, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5050631.

Kumar, V., Robert-P, L., & Kumar, S. (2021, September). Production and Operations Management Call for Papers Special Issue: Transformative Marketing and Operations Management. *Production and Operations Management Society*, 30(9), 3317–3319. doi:10.1111/poms.13584.

Lenhof, K., Eckhart, L., Rolli, L.-M., & Lenhof, H.-P. (2024). Trust Me If You Can: A Survey on Reliability and Interpretability of Machine Learning Approaches for Drug Sensitivity Prediction in Cancer. *Briefings in Bioinformatics*, 25(5). doi:10.1093/bib/bbae379.

Lining, Y., Karl-Härdle, W., Borke, L., & Benschop, T. (2023). An AI Approach to Measuring Financial Risk. *The Singapore Economic Review*, 68(5), 1529–1549. doi:10.1142/S0217590819500668.

Makkar, M., & Rani, R. (2024). Strategies to Respond to VUCA World. *International Journal of Logistics: Research & Applications*, 15(2), 37–49. Retrieved April 19, 2025.

Mettrick, G. (2024). The AI Handbook for Financial Services Leaders: Tips and Tactics for Mastering AI in Banking and Finance. *Journal of Securities Operations & Custody*, 16(3), 286–296. doi:10.69554/sesy2749.

Mittapally, B. K. (2024). Artificial Intelligence and Predictive Analytics for Business Growth. *Workforce Solutions Review*, 2, 14–17.

Nguyen, D. K., Sermpinis, G., & Stasinakis, C. (2023, March). Big Data, Artificial Intelligence and Machine Learning: A Transformative Symbiosis in Favor of Financial Technology. *European Financial Management*, 29(2), 517–548. doi:10.1111/eufm.12365.

Nurski, L., & Ruer, N. (2024). Exposure to Generative Artificial Intelligence in The European Labor Market. Bruegel. Retrieved July 20, 2025, from http://www.jstor.org/stable/resrep58150.

Nyberg, A. J., Schleicher, D. J., Bell, B. S., Boon, C., Cappelli, P., Collings, D. G., … Ingo. (2025, July). A Brave New World of Human Resources Research: Navigating Perils and Identifying Grand Challenges of the GenAI Revolution. *Journal of Management*, 51(6), 2677–2718. doi:10.1177/01492063251325188.

Praveen, R., & Bhati, R. (2024). Revolutionizing Supply Chains: The Transformative Influence of Artificial Intelligence. *International Conference on Ongoing Research in Management & IT*, 610–622.

Prieto, B. (2025, July). Artificial Intelligence, Complexity, and Quantum Project Management: A Transformative Approach. *PM World Journal*, 14(7). Retrieved July 28, 2025.

Randstad Enterprise. (2019, May 6). How Cisco Drives Employee Engagement, Talent Attraction and Retention. Retrieved July 22, 2025, from randstadenterprise.com: https://www.randstadenterprise.com/insights/randstad-enterprise-insights/how-cisco-drives-employee-engagement-talent-attraction-retention/.

Razbornik, I., & Todosijević, S. (2024, June). Transformational Ai (Trai): Revolutionizing The Modern Business World. *Economy & Market Communication Review*, 14(1), 265–275. doi:10.7251/EMC2401265R.

Roark, C., Jamison, S., Crowley, J., Castagnino, T., & Millan, K. (2025). The Productivity Payoff: Unlock Competitiveness with Gen AI. Accenture. Retrieved June 13, 2025, from https://www.accenture.com/us-en/insights/strategy/productivity-payoff.

Samokhvalov, K. (2024, January). The Transformative Impact of Artificial Intelligence on the Management Consultancy Sector. *Management Consulting Journal, 7.1*. doi:10.2478-mcj-2024-0006.

Sharma, A. (2025, January). Artificial Intelligence for Sense Making in Survival Supply Chains. *International Journal of Production Research*, 63(2), 435–458. doi:10.1080/00207543.2023.2221743.

Stoyanova, M. (2023). Document Process Automation with Artificial Intelligence for Logistics Sector. *Izvestia Journal of the Union of Scientists – Varna Economic Sciences*, 12(1), 190–197. doi:10.56065/ijusv-ess/2023.12.1.190.

Teodorescu, P.-G., & Ovreiu, S. (2024). Transformative Advances in NLP and AI: Charting the Evolution of Technology. *Informatica Economică*, 24(4). doi:10.24818/issn14531305/28.4.2024.02.

Uddin, M. B., Jayaram, J., Amin, A., & Islamd, M. R. (2025, July). Impact of Supply Chain Organizational Learning on Organizational Role Redefinition and Supply Chain Risk Management: The Moderating Role of Market Turbulence. *International Journal of Production Research*, 1–23. doi:10.1080/00207543.2025.2523526.

Zheng, G., & Brintrup, A. (2025, August). Enhancing Supply Chain Visibility with Generative AI: An Exploratory Case Study on Relationship Prediction in Knowledge Graphs. *International Journal of Production Research*, 1–23. doi:10.1080/00207543.2025.2543964.

8

AI-DRIVEN INVENTORY OPTIMIZATION FOR RESILIENT SUPPLY CHAINS

Doina Vilceanu

We'll integrate fresh 2025 field evidence from the front lines: Amazon's foundational forecasting model (Amazon Staff, 2025), Nestlé's 87% edible-waste reduction pilot (Carrington, 2025), Toro's just-in-time resilience under tariff volatility (Bendeich, 2025), and lead-scoring breakthroughs from Persana AI (2025) and SuperAGI (2025). We'll also touch on Microsoft's Dynamics 365 Copilot and cross-industry signals from healthcare (Capirci, 2025). The result is a **resilient growth framework** that turns first-party data (orders, baskets, customer relationship management (CRM) touchpoints) and external signals (weather, holidays) into the right offer, to the right segment, at the right time—without eroding your margins.

We conclude with a pragmatic executive roadmap (data readiness → pilots → scale), key metrics to monitor (forecast accuracy, waste, incremental margin, conversion lift), and practical guardrails for pricing and promotions.

This chapter proceeds from forecasting and demand sensing to next-order prediction, then to segmentation and margin-safe promotions, and finally to predictive lead scoring—before closing with an implementation roadmap, playbooks, and future trends.

Introduction: The Supply Chain–Marketing Nexus

In boardroom conversations about growth, marketing and supply chain usually sit at opposite ends of the table. Marketing owns the story, the offer, and the media mix. Supply chain owns procurement, inventory, and delivery. Operationally, this division makes perfect sense. Commercially, it can be absolutely fatal.

When supply and demand aren't synchronized, a business loses twice: once in wasted operational spend and again in wasted promotional spend. A promotion that runs with insufficient stock undermines trust and squanders media budgets. Conversely, surplus stock forces sweeping price cuts that teach customers to simply wait for a deal. This inventory misalignment, a silent tax on profitability, is a problem that has grown to a fever pitch.

DOI: 10.4324/9781003632429-12
This chapter has been made available under a CC-BY-NC-ND 4.0 license

This chapter is written from a marketer's standpoint, and the premise is simple: *availability is the new creative*. We'll dive into four specific capabilities that align supply and demand into *one powerful growth system*:

- Predictive inventory optimization
- Next-order prediction
- Segmentation and margin-safe promotions
- Predictive lead scoring

With fresh cases from 2025 (Amazon Staff, 2025; Bendeich, 2025; Carrington, 2025; Persana AI, 2025; SuperAGI, 2025) and a brief healthcare signal (Capirci, 2025), we'll assemble these insights into a resilient growth framework with a pragmatic implementation roadmap, playbooks, and governance.

The volatility of the 2020s—pandemic aftershocks, tariff swings, climate disruptions, and multiple war zones—exposed just how brittle "average-based" planning can be. Historical extrapolation, once a safe baseline, simply stopped working. That is precisely where artificial intelligence (AI) changes the very economics of growth.

Case in point: In June 2025, Amazon unveiled an AI foundational forecasting model (Amazon Staff, 2025) trained on billions of transactions and enriched with external signals like weather and holidays. The reported impact was material: roughly a 10% improvement in long-term national event forecasts and a 20% improvement in regional accuracy, spanning millions of items. For Amazon, this means inventory placement that anticipates real demand. For the rest of us, it proves a commercial point: when supply precision rises, marketing return on investment (ROI) rises right along with it. Campaigns finally land on stocked shelves, not empty promises.

We all know the pain of running a brilliant campaign that drives demand for a product that isn't there. AI finally closes that gap.

With the stakes and structure in view, we start where most leakage begins: forecasting and inventory waste

TABLE 8.1 Comparative AI impact across industries (illustrative synthesis based on industry reports and case studies, 2021–2025)

Industry	*Forecast Accuracy Improvement (%)*	*Stockout Reduction (%)*	*Waste Reduction (%)*	*Lead Conversion Uplift (%)*
FMCG	25	30	35	12
Automotive	18	15	12	10
IT&C (High-Ticket B2C)	22	20	18	15
B2B Wholesale	20	28	25	14

Source: Authors' own elaboration, based on Amazon Staff (2025), Bendeich (2025), Carrington (2025), Persana AI (2025), SuperAGI (2025), and Capirci (2025).

Forecasting and Inventory Waste

Forecasting errors are expensive in both directions. Overestimate demand and you tie up capital in unsold stock, then pay again in markdowns and margin attrition. Underestimate demand and you miss out on revenue while eroding trust; you also waste media spend pointing customers to out-of-stock products.

This is what operations theory calls the **bullwhip effect**: small consumer misalignments amplified into costly supply chain shocks. The bullwhip effect magnifies the damage—a small fluctuation at the consumer end ripples upstream into disproportionate swings in distributor and supplier orders. One careless promotion and suddenly a supplier is panicking, a manufacturer is retooling, and regions are out of sync.

The bullwhip effect's amplification of demand signals becomes clear when visualized (see Figure 8.1), showing how small consumer fluctuations create massive upstream disruptions —precisely where AI's stabilizing effect matters most.

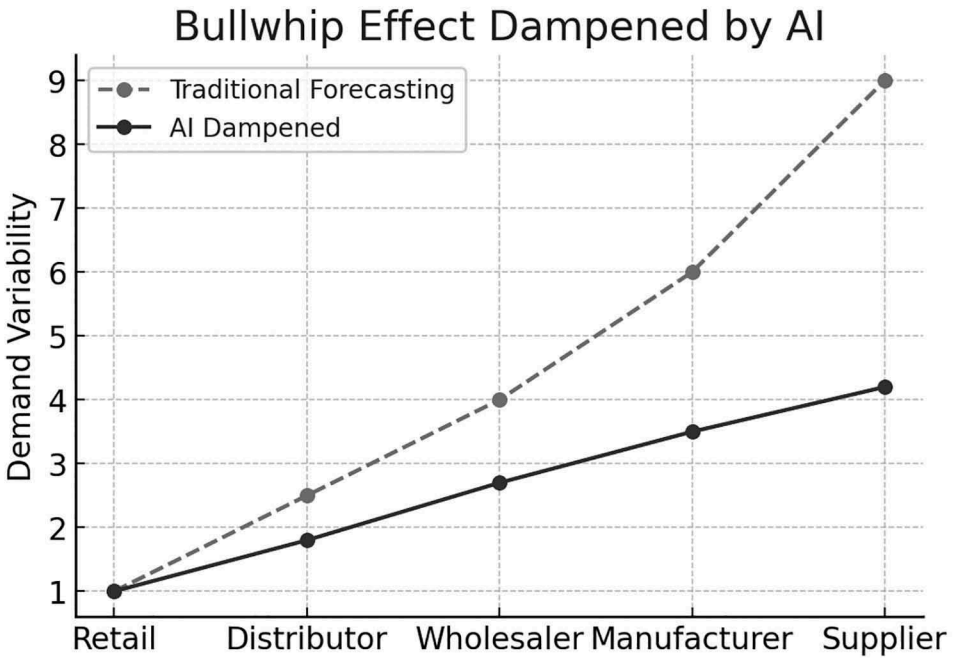


FIGURE 8.1 Bullwhip effect dampened by AI
Note: The values presented are illustrative, designed to highlight the relative amplification of variability across supply chain tiers under traditional forecasting versus AI-enhanced demand sensing.
Source: Conceptual illustration based on the bullwhip effect described by Lee, Padmanabhan, & Whang (1997).

Predictive inventory optimization uses AI to move beyond simple averages. Models ingest first-party signals (orders, stock keeping unit (SKU) histories, promotions, loyalty activity) and external context (weather, holidays, macro indicators; sometimes even competitor activity) to forecast with SKU × location × week granularity. The output is a living signal, not a static plan.

When the forecast improves, marketing can be far more surgical: narrow a discount to a couple of ZIP codes, shift spend to regions with high availability, or re-sequence a national campaign into phased regional rollouts.

Consider three quick snapshots from 2025:

- **Amazon:** Their foundational model improved national and regional forecasts by 10–20%, steering inventory to where demand would actually materialize (Amazon Staff, 2025).
- **Nestlé:** A UK pilot with an AI partner cut edible food waste by a whopping 87% within two weeks (Bendeich, 2025), saving roughly 700 tonnes of food and identifying nearly £14 million in annual cost savings. For a marketer, this means promotions become a steering wheel instead of a blanket clearance hammer—nudging just the right segments and channels where surplus risk exists, all while preserving brand equity and margin.
- **Toro:** They adapted AI for just-in-time resilience amid tariff volatility (Bertini & Koenigsberg, 2025), maintaining launch timing without over-buffering inventory. Marketing calendars stayed intact because operations could flex intelligently under external shocks.

Forecast accuracy isn't just an IT metric. It's a commercial instrument. Better forecasts mean fewer emergency discounts, fewer stockouts during promotions, and a credible promise to customers that what you advertise is truly available. For brands with **ESG** (environmental, social, and governance) commitments, waste reduction isn't just good business; it's also good citizenship.

While baseline AI forecasting improves accuracy, breakthrough performance comes from explaining the surprises—by incorporating external signals traditional models miss entirely.

Demand Sensing: External Signals That Move the Needle

Most forecasting errors aren't about bad math—they're about blind spots in context. Demand doesn't move in neat, linear patterns; it shifts with weather anomalies, cultural events, and macroeconomic jolts. By the time point-of-sale (POS) or enterprise resource planning (ERP) systems detect the signal, the wave has often already passed.

Forward-looking organizations have begun augmenting their sales histories with exogenous signals that help explain why demand veers off the expected path. These signals range from the obvious—temperature spikes driving ice cream sales, school holidays shifting family shopping trips—to the subtle: regional sporting events boosting beverage consumption, or macroeconomic confidence indices nudging discretionary categories. This combination transforms forecasting from a rear-view mirror into a forward-facing radar.

Some factors are common sense: don't promote animal produce during Christian fasting periods, pork in Muslim countries, or veal in India. Rather than relying on commonplace calendar cues, leaders mine higher-frequency indicators—search trends, social virality, live event feeds, and local macro indices—that explain short, sharp demand shifts before POS registers them.

A telling case comes from the grocery sector in the UK. In 2024, retailers found that including weather forecasts and local holiday calendars improved short-term demand predictions by double digits, especially in categories like fresh produce and ready-to-drink beverages (Carrington, 2025). Similarly, US DIY chains that layered in housing permit data and local employment trends saw material improvements in aligning lumber and paint stock with actual renovation activity. In each case, the external signal didn't replace POS data; it explained the surprises—why lettuce wilted unsold after a rainy weekend, or why power tools spiked in one region while remaining flat in another.

The lesson is clear: external signals let leaders phase inventory placement and launch campaigns first where availability will be strongest. Imagine a national beverage brand seeing that a warm spell is forecasted to hit southern Europe two weeks earlier than the north. Instead of blasting a homogeneous campaign, media spend is sequenced: southern markets go live first, supported by pre-positioned stock, while northern campaigns are held back until the weather aligns. The same logic applies to business-to-business (B2B): a packaging supplier anticipating port congestion can pre-warn clients, adjust delivery times, and stagger promotional commitments so sales teams don't overpromise. The decision is no longer defensive—it is strategic, and it's legible to the C-suite.

The best organizations are extending demand sensing even further by weaving in social and digital signals. Search query spikes for "allergy medicine" can precede pharmacy demand by weeks; TikTok-driven fashion trends can tilt apparel sell-through rates almost overnight, and a backlash on Facebook about a new non-inclusive marketing campaign can flop an entire new product launch.

I clearly remember the first time social media became a critical factor in forecasting. At the dawn of the influencer era—long before "viral" was a key performance indicator (KPI)—a client of mine who rented haute couture gowns had a forgotten celebrity post a photo wearing one of their dresses on a Tuesday morning, tagging the website. Traffic exploded to Black Friday levels within minutes, crashing the site. Neither the client, the e-commerce platform, nor the hosting company could have predicted such an impact. In response, we designed a complex contingency plan to forecast sudden spikes and allocate additional hosting resources in advance.

Today, with current AI capabilities, those same signals are not only predicted but closely monitored and fully exploited.

The 2025 Amazon foundational model (Amazon Staff, 2025), for instance, explicitly incorporated non-transactional signals into its training set, delivering 10–20% forecasting improvements at both national and regional levels. The implication: if a global leader like Amazon treats exogenous signals as first-class citizens in its models, smaller organizations can hardly afford to ignore them.

Insight: Demand Sensing Beyond POS

Many retailers still over-index on sales logs. Adding supplier lead times, port congestion alerts, or distributor price sheets helps anticipate pipeline friction. The goal isn't clairvoyance—it's fewer surprises and faster micro-corrections.

Having aligned inventory with reality at the regional level, the next challenge is timing at the individual level—*when* each customer is most likely to buy again.

Next-Order Prediction

Forecasting gets the market ready. **Next-order prediction** gets the individual ready. Traditional CRM knows who a customer is and what they buy; it rarely knows when they are likely to buy again. That timing gap wastes budget on early discounts and misses revenue when outreach arrives too late. The fix is simple to state and brutally hard to execute: predict each customer's next purchase day (NPD) and trigger outreach in that crucial window.

Modern models—often Transformer-based (Grigoras et al., 2023)—examine **cadence** (detergent every 28–35 days), **basket sequences** (pasta → sauce), **seasonality** (beer before holidays), and **context** (paydays, weather, regional events). Academic work has shown Transformers outperforming ARIMA/LSTM baselines (Grigoras et al., 2023) because they capture long-range behavioral dependencies. See Technical Note A for a concise primer on ARIMA/LSTM baselines and why they serve as benchmarks.

In B2B, a hybrid of empirical Bayes and machine learning (De et al., 2024) has been used to predict reorder timing in procurement flows, telling sales reps exactly when to check in—neither too soon nor too late.

The logic is simple: in categories where customer lifetime value is tied to repeat purchases, mistiming even one cycle erodes profit. Send the promotion too early and you waste budget on a customer who doesn't need the product yet. Send it too late and a competitor gets the sale.

I have a scar-tissue reminder of why stock and timing matter. Around 2007, I was Internet Director at Studio Moderna when a TV campaign promoted a pan bundled with a cheap mixer. There was an overstock of "As Seen on TV" pans, but sales were modest. Then the marketing department came up with the bundle idea. The promotion exploded—the bundle sold out quickly while "unaccompanied" pans still filled the warehouses. To avoid a customer revolt, the company air-freighted cheap mixers (yes, the most expensive route) just to fulfill orders already taken.

On paper, I had "nothing to do" with ordering. In reality, marketing and supply chain were welded together by that single promotion. The lesson branded itself into my thinking: without timing intelligence and cross-departmental alignment, even a wildly successful campaign becomes a profit trap. Today, next-order prediction and AI-assisted bundling would have prevented that mismatch—by revealing the true elasticity and attach-rate of "mixer + pan" and by pacing offers to match replenishment capacity.

With NPD in place, three immediate wins appear:

- **Suppression savings.** Don't ping customers who are out of window; move that spend to those who are due.

- **Right-time offers.** Trigger promotions when reorder probability peaks; shift from scatter to scalpel.
- **Mid-cycle cross-sell.** Time accessories (filters, cables, add-ons) for the mid-cycle moment, not randomly.

The full stack, for all its sophistication, is hardly an exotic solution. It's first-party data, a handful of behavioral features, a competent model, and marketing operations that act on the signal.

Recent product launches prove the commercial urgency of NPD. In 2025, Prediko rolled out an AI-driven reorder prediction engine for Shopify merchants, allowing SMEs to time marketing emails and replenishment campaigns with SKU-level precision. evolio.ai extended similar capabilities to Salesforce ecosystems, giving account managers a probability window for each customer reorder. Blue Yonder integrated NPD into its broader end-to-end supply chain planning suite, emphasizing the marketing payoff: better-timed promotions and fewer lost cycles. These tools aren't futuristic—they're already in the hands of frontline sales and e-commerce managers.

Case Study: Automating Technical Bundles with Invoice Data

In industries with highly technical products—plumbing, HVAC, electrical, construction materials—the challenge is not only timing the next order but ensuring the order is complete. A boiler without valves or sensors is a customer service disaster.

A retailer with millions of products and hundreds of thousands of customers tackled this by mining invoices and product catalogues. The goal was to detect "safe bundles"—repeatable product groupings that could be surfaced online without a human adviser.

The roadmap was staged:

- **Invoice Analysis in BigQuery ML.** Historical orders were clustered to reveal co-purchase patterns (e.g., 80% of installations of Boiler A also included Valve B, Kit C, and Sensor D). These became the "Complete Kits" offered online.
- **Prediction with AutoML.** Once bundles were identified, AutoML was trained to predict which kit each customer segment was most likely to need next, and when (NPD at a bundle-level, not just SKU-level). Timing of reorder for consumables (filters, cartridges, cleaning kits). This moved the system from reactive bundling to predictive bundling and just-in-time cross-sell.
- **Specialist Knowledge Integration.** Training manuals and expert rules were codified into simple business rules, acting as guardrails for the machine learning (ML) outputs. For example, the system would never suggest a boiler without the mandatory safety valve.

The result was a system that allowed customers to self-serve online for complex solutions—without increasing the risk of incomplete or unsafe orders. The proof of concept was low-cost: BigQuery SQL queries and clustering produced usable bundles within weeks. Only after success was demonstrated did the team invest in AutoML for predictive personalization and explore knowledge graph models for long-term

sophistication. This staged approach is cheaper, safer, and faster to prove value than building a custom AI recommender from scratch. In other words, don't start with futuristic AI—start with your invoices.

Thought the process is still in early BETA stage, in constant learning loop, and we are missing full YOY numbers, C-level feedback is clear: the program increased order-completeness rates and reduced avoidable returns as customers assembled "complete solutions" online; customer satisfaction improved, and accessory attach-rates rose as cross-sell timing aligned with maintenance cycles.

Insight: Hybrid Modelling for Reorder Precision

Predicting when a customer will reorder isn't just about raw data; it's about capturing behavioral rhythm. Two approaches complement each other: **XGBoost**, a high-performance engine for structured business data that finds nonlinear interactions with minimal feature engineering; and the **Modified-Poisson Gamma (MPG) model**, which explicitly models repeat-purchase cadence across segments (weekly buyers, monthly replenishments, seasonal spikes). MPG gives an interpretable view of "how often" different cohorts return; XGBoost sharpens "who this week" given price, seasonality, and context. Combined, you get forecasts that are both accurate and explainable—so teams can time promotions where purchase likelihood peaks, cutting discount waste and missed windows (Grigoras et al., 2023). New-age companies like Prediko, evolio.ai, and Blue Yonder are already harnessing these multimodal techniques, staking one model on top of the other to generate accurate predictions at scale.

For readers seeking additional algorithmic detail, see **Technical Note B** on XGBoost–MPG hybrids and calibration for operational triggers.

Operationalizing Next-Order Prediction

Treat next-order prediction as a weekly operating rhythm rather than a one-off campaign. Build rolling windows—customers due in the next 7–10 days—and trigger "due soon" journeys while suppressing the rest. This suppression alone can free 10–20% of outbound budget for higher-ROI audiences, while also reducing unsubscribe rates from over-communication. The key is to align NPD signals with inventory availability, so you never push items that are in constraint. This creates a synchronized playbook: marketing pulls the trigger only when both the customer is due, and the product is available.

B2B & SaaS Reorder Automation

In replenishment-driven categories, mid-cycle outreach based on reorder probability feels like service, not selling. A supplier who checks in exactly when inventory is running low strengthens trust and reduces churn. Hybrid models are powerful here: Empirical Bayes + ML hybrids estimate cadence at the segment level, then personalize timing with each account's recent behaviors (De et al. 2024). For more complex sequences, Transformer architectures capture long-range dependencies—such as how a spare-parts order after installation changes the rhythm of future maintenance buys—

explaining why they outperform classic time series on next-purchase timing (Grigoras et al., 2023).

Having solved for "when," the next lever is "who"—directing promotions to the customers whose behavior actually changes because of the offer, not those who would buy anyway.

Customer Segmentation, Promotions, and Margin-Safe Discounts

Every shopper knows the drill: buy two shampoos, get one free. But most retailers know the pain: half of those bottles would have sold anyway—only now at half the margin. Promotions are double-edged. Done well, they drive incremental volume and loyalty. Done poorly, they teach customers to wait, punish margin, and dilute brand equity. In many FMCG categories, a significant share of promotions goes to people who would have bought anyway. That's not growth; that's pure leakage.

AI lets us move from blunt "likelihood to buy" models to **uplift modelling**—predicting who will buy *because of* a promotion. This shift is crucial: avoid discounting loyalists who will purchase at full price and direct incentives to persuadable and surplus-risk regions. Think of two customers. Both have bought your brand ten times. One will keep buying regardless. The other is on the fence and will only come back if nudged. Traditional models treat them the same; uplift models separate them.

The 4 Ps of marketing are getting a facelift. In short life-cycle categories, a Unified Price, Promotion, and Markdown Optimization (UPPMO) view is essential (Gartner, 2024). Price, promo depth, and inventory should be tuned together:

- If inventory is tight, suppress deep discounts and push value via bundles or content.
- If inventory is long, deepen discounts but surgically—only to clusters that are elastic, not to blanket audiences.

The payoff is real. A leading European grocer piloting an AI-driven UPPMO with consulting support achieved a 3–5% revenue uplift with 40% less discount spend. The system didn't increase promotion—it redirected it, with surgical precision.

Personalization at scale is no longer an aspiration. Leaders achieve double-digit revenue lift by tailoring who sees what and when. Generative AI now allows for creative and copy variation at a massive scale, but the center of gravity remains economic: don't erode margin where you don't need to. Use forecast and next-order signals to shape who gets a discount, where, and when.

We should also be honest about **dynamic pricing**. In some markets—mobility, hospitality—it's accepted. In others, "surge pricing" triggers a fierce backlash. The answer isn't to abandon optimization; it's to add guardrails and transparency: fair ranges, clear logic, consistent messaging. The difference between a fair adjustment and a gouge is often a matter of communication. Customers accept optimization when it feels like relevance, not opportunism (Bertini & Koenigsberg, 2024).

We keep dynamic pricing concise here to preserve the chapter's supply-led focus; readers seeking deeper treatment can see **Technical Note C** on transparency guardrails and fairness bands.

Insight: Clustering as the Foundation of Smart Segmentation

Traditional segments—age, income, generic **RFM** (recency, frequency, and monetary value)—are blunt. Clustering discovers behavioral tribes or cohorts hidden in your data: weekly price-sensitive family shoppers, monthly brand-loyal professionals, seasonal bulk buyers. Each tribe gets its own plan—bundles for families, concierge outreach for professionals, timed bulk offers for seasonal buyers. On the supply side, clustering SKUs and regions with similar demand curves simplifies forecasting and allocation. Instead of managing 10,000 items independently, managers can handle 15 coherent clusters—sharpening promotions while reducing complexity.

Once promotions are precision-targeted, the final gating factor for growth is prioritization—ensuring sales effort lands where deals are both likely and deliverable.

Predictive Lead Scoring and AI-Driven Acquisition

Even with inventory aligned and promotions precise, growth stalls if sales teams chase the wrong leads. Old-school scoring systems—open an email = +10 points, attend a webinar = +20—are blunt instruments. They miss the sequence, the context, and most importantly the feasibility of closing a deal.

Modern lead scoring reframes the question: not just "Who is interested?" but "Who is ready to buy now, with what we can actually deliver?" That distinction matters. A prospect can score highly on interest signals but still be a dead end if the product they want is back-ordered for six months. Conversely, a modestly engaged lead may be on the cusp of purchase if stock is available and budget is aligned.

The new generation of models blends behavioral telemetry (pricing page views, demo requests, repeated site visits), firmographics (sector, size, growth rate), and inventory context (are SKUs available, and at what lead time?). The result is not just a likelihood-to-buy score, but a **likelihood-to-buy-now score**—the crucial difference between chasing intent and capturing revenue.

The academic evidence is catching up with practice. A 2025 peer-reviewed study in *Frontiers in Artificial Intelligence* (González-Flores et al., 2025) showed ML-based lead scoring improved conversion in B2B software by learning from journey sequences—not just static attributes. Practitioners report even larger uplifts in live deployments. SuperAGI, in its 2025 client benchmark report, documented an *average 215% increase in qualified leads* across B2B software and enterprise-services clients after adopting sequence-based ML scoring. Persana AI reported a *32% conversion-rate uplift* in complex-sales funnels—primarily SaaS and technology-solutions accounts—after replacing rules-based ranking with ML-driven prioritization (Persana AI, 2025; SuperAGI, 2025). These results highlight how predictive scoring can reallocate sales effort from volume to value.

CRM Integration Playbook

The biggest risk is complexity. A model that requires reps to open three dashboards will die on impact. The most effective systems are invisible—they simply enrich the CRM fields that sellers already use. In practice:

- Surface a single "buy-now likelihood" field in the CRM.
- Attach short reason codes (e.g., "recent pricing page views," "matched ICP," "in-stock SKUs") to boost trust.
- Trigger alerts when a score crosses a threshold and inventory is available.
- Measure outcomes by cohort, so sales sees win-rate lift, not just another number.

The best predictive systems are boring. They don't feel like AI—they feel like the CRM finally works as promised.

Rules vs. AI—When Each Wins

Rules-based scoring is fast, interpretable, and useful for compliance and routing (e.g., "all leads from pharma go to Team X"). But it is brittle: one unexpected sequence, and the rule collapses. Machine learning, by contrast, learns from wins, losses, and unstructured signals to prioritize accounts that rules would miss. The pragmatic blend is best: rules for who *must* see the lead, AI for who *should* see it first (Amazon Staff, 2025).

Case Study: Enterprise SaaS Lead Prioritization

A regional SaaS vendor selling enterprise e-commerce software faced the classic problem: a funnel bursting with leads, but a sales team unable to tell noise from signal. Their rule-based scoring system prioritized activity volume (webinars, downloads), which meant smaller accounts with hyperactive researchers often scored higher than million-dollar prospects quietly exploring. Long explanatory talks with wannabe entrepreneurs left no time for busy C-level RFQ.

By deploying an ML-based lead scoring system inside their CRM, trained on three years of win/loss data, the company completely re-ranked its pipeline. The model weighted journey sequences (demo request → pricing page → stakeholder invite) more than raw activity counts and cross-checked scores against current inventory of licenses and deployment bandwidth.

Within six months:

- Sales accepted 27% fewer leads, but
- Close rates rose by 18%, and
- Average deal size increased because reps spent more time on enterprise accounts that were both ready and serviceable.

For sales managers, the real breakthrough wasn't the algorithm—it was that reps finally trusted the system. Each score came with reason codes ("enterprise size, recent pricing visit, deployment slot available"), turning AI from a black box into a guided compass.

Insight: Transformers vs. Stacking—Different Tools for Different Problems

Executives often hear both terms but rarely see the distinction. A **Transformer** is a powerful deep learning architecture built on self-attention—excellent at finding long-

range dependencies, such as predicting a customer's next purchase day from months of behavioral history. By contrast, **Stacking** is not an architecture but an ensemble strategy: combining multiple models (e.g., XGBoost, logistic regression, neural nets) and training a meta-model to learn when to trust each. In practice, Transformers shine in sequential forecasting (next-order prediction, basket flows), while Stacking excels in structured decision problems (predictive lead scoring, segmentation) where different algorithms capture complementary aspects of the data. Employing both at different stages can skyrocket a business.

While this chapter centers on FMCG, B2B, and high-ticket B2C, healthcare reminds us that reliability is a cornerstone of marketing. Hospitals such as Mayo Clinic and Cleveland Clinic have used AI with IoT sensors to forecast shortages of critical items—surgical consumables, dialysis components—so clinicians are not forced into workarounds (Capirci, 2025). In that world, "always available" is not a tagline; it is life-critical. The lesson translates: credibility in your promise is a supply chain outcome as much as a copywriting skill.

For readers who want the math, see **Technical Note D** for short, non-derivational summaries of Transformers and stacking ensembles.

Having aligned timing, targeting, and prioritization, the question every executive asks is practical: Where to start, and how to scale without losing control?

Implementation Roadmap

Executives always ask the same question: "Where do we start?" The answer is deceptively simple: start small, where it hurts most, and where you can measure. The journey has three phases.

Data Readiness (4–6 weeks)

First, consolidate the data you already have: orders (SKU, location, date), promotions (depth, channel, region), CRM (customer IDs, touchpoints), and external signals (weather, holidays). Establish a weekly refresh cadence so this isn't a one-off exercise. Then, define your baselines: current forecast error, stockout rates, waste levels, promotion ROI, and discount leakage. Without a baseline, improvement is invisible.

Pilot (8–12 weeks)

Choose a single SKU family or a region with both volume and pain. Stand up your first forecast and next-order prediction models; wire the signals into your campaign tooling. Implement a single uplift-based promotion and a supply-aware lead score. This is where you'll see the results emerge: 10–20% forecast lift, fewer stockouts, reduced waste, higher promo ROI, and tangible conversion lift. Crucially, keep the pilot compact—it should succeed or fail fast enough to teach.

Scale (12+ weeks)

Once the pilot demonstrates value, extend to more SKUs and regions. Institutionalize a **Revenue S&OP**—a cross-functional weekly forum that unites marketing, sales, and supply chain around one single set of signals. Add guardrails: fairness ranges for prices, budget caps for discounts, and inventory safety bands. Document feature lineage so you know where predictions come from, and always enable human override. AI should propose; leaders must decide.

Strategic Framework: In-House AI Development vs. External Consulting

Executives often ask: should we build AI in-house, or bring in external partners? In my experience, the best outcomes come from blending both approaches.

Why Internal AI Capabilities Matter

Internal teams ensure cultural fit, data stewardship, and continuity.
They embed AI in everyday operations and protect know-how long-term.

Why External Consultants Are Critical

Kickstarting transformation. External specialists shorten learning curves and de-risk early pilots.
Scaling best practices. Consultants bring hard-won insights from multiple industries and markets—knowledge no single company can generate alone.
Bridging talent shortages. Consultants fill gaps with niche expertise until companies can grow their own teams.

The Hybrid Advantage

The strongest model isn't "internal vs. external," but **internal with external**: consultants accelerate progress, introduce proven frameworks, and train staff on live projects. Internal teams ensure that knowledge sticks, while consultants continue to inject fresh benchmarks and innovations over time.

In fast-moving fields like AI, ongoing access to external benchmarks and emerging practices remains a competitive edge; internal capability anchors execution, while expert partners keep it sharp.

Adoption Challenges and Missteps

AI projects in commerce stall for predictable reasons:

- **Data silos** create a partial truth—forecasts ignore promotions, or promo engines ignore inventory.
- **Pilots lack a business owner** and are run as IT projects without P&L accountability.
- **Governance is an afterthought**—and one pricing misstep can derail trust for years.

The fix is structural. Assign an executive owner, operate a joint Revenue S&OP, and publish clear guardrails. A frequent anti-pattern is **algorithm first**. The winning pattern is **decision first**: identify a recurring decision (how much to stock, whom to discount, which lead to call), define the metric that matters (margin preserved, incremental sales), and then choose the model. This sequence keeps teams honest and focused on commercial outcomes, not shiny models.

Laying the Groundwork: Unifying Data Silos for Strategic AI Use

Before AI can deliver measurable results, businesses must address a more fundamental step: securing and unifying the right data. This goes beyond simply "having data." In practice, valuable insights remain trapped in silos—CRM systems, e-commerce platforms, ERP systems, or BI dashboards—each operating independently. Decisions made on fragmented or inconsistent information aren't just inaccurate; they're fundamentally misleading.

Establishing a central data strategy with clear access, ownership, and governance rules is essential. Only by integrating sources can producers and wholesalers fully leverage AI tools for inventory optimization, next-order prediction, customer segmentation, or predictive marketing. Success begins not with models, but with disciplined data hygiene and strategic alignment.

Once a company identifies the critical datasets, the next step is to ensure its infrastructure can support scale. AI doesn't thrive in legacy environments. It requires scalable storage, high-speed processing, and seamless integration across systems—something traditional IT architectures often fail to provide. This is why modern cloud platforms such as Microsoft Azure, Google Cloud, and Amazon Web Services matter: they offer the flexibility, resilience, and computing power required for real-time AI applications.

By moving to the cloud, businesses not only unlock performance and scalability but also position themselves to consolidate fragmented tools—from CRM and ERP to e-commerce platforms—into a unified, AI-ready ecosystem. The companies that treat data as a strategic asset, rather than a by-product of transactions, will be the ones that turn AI from a buzzword into a bottom-line impact.

Practically, this means your Revenue S&OP runs on one version of truth: the same unified signals drive forecasts, promotions, and sales prioritization.

Measurement and Governance

What gets measured gets improved. Treat AI for supply–demand alignment not as a lab project but as a commercial system. Review a compact scorecard monthly and make it the single source of truth across marketing, sales, and operations.

- **Forecast quality.** Track **WAAPE/MAE** by SKU × region × week. Separate stable vs. promoted periods—averages often mask errors. When error spikes, pause deep promotions, shift spend, and raise resupply tickets. (WAAPE weights percentage errors by sales volume; MAE is the average absolute error.)
- **Availability and waste.** Monitor stockouts (%) and lost-sales estimates. Measure edible or write-off waste (%, €) alongside ESG impacts (environmental, social, and governance). Use uplift-based promotions in surplus-risk clusters instead of blanket clearance.
- **Promotion economics.** Report promo ROI and discount leakage (% of discounts given to customers who would have bought anyway). Uplift modelling bends this curve downward.
- **Acquisition efficiency.** Track qualification rate, win rate, and sales cycle time by lead-score bands. If "very high" scores fail, it's a signal that either the model drifted, or supply was constrained.
- **Margin preserved.** Estimate margin saved vs. last year's promotions after adjusting for mix and seasonality. This keeps teams focused on profitable growth.

Governance prevents optimization from backfiring. Establish transparency and fairness in pricing (Bertini & Koenigsberg, 2024). Set drift monitors to detect when models lose accuracy. Maintain feature lineage to show how predictions were made. And always grant human override—AI is a copilot, not the pilot.

Practical Playbooks

Conceptual frameworks are powerful, but executives often ask: "What does this actually look like in my industry?" The following playbooks illustrate how the same AI capabilities—forecasting, next-order prediction, uplift modeling, and lead scoring—might be applied in practice across different verticals.

- **FMCG.** Regional promotions can be tied directly to regional availability. When AI signals surplus risk, targeted discounts could be pushed to elastic customer clusters. Where stock is tight, marketers might shift media rather than discount. Basket bundles (for instance, pasta with sauce) can lift average order value without unnecessary markdowns.
- **B2B.** Predictive reorder alerts can guide sales reps to reach out precisely when buyers are due, rather than too early or too late. Lead scoring aligned with inventory ensures teams avoid commitments they cannot fulfil. Contracts might also include availability-aware pricing—such as small incentives for flexible delivery windows when supply is tight.
- **High-ticket B2C.** Next-order prediction windows can enable concierge-style upgrades (e.g., smartphones or appliances). Dynamic pricing, when used, should be paired with fairness guardrails such as caps and transparent logic. Accessories and add-ons can be introduced mid-cycle to create value without discounting the primary item.
- **Healthcare.** AI forecasting can underpin an "always available" positioning for critical items. Hospitals or distributors could use availability dashboards to reassure institutional buyers—turning reliability itself into a differentiator.

Trust and Transparency: A Quick Guide

Customers accept optimization when it feels like relevance, not opportunism. Be explicit about the rules: publish fair price ranges; avoid opaque surges; keep messaging consistent (Bertini & Koenigsberg, 2024).

Offer short-window price guarantees to reduce regret and returns. When inventory is tight, communicate availability and lead times upfront. Give sales and service teams manual override authority.

Inside the company, narrate decisions. Attach short "why" notes to price changes, promotion suppressions, and lead-prioritization moves. When colleagues understand the rationale, adoption accelerates—and they provide feedback that improves the models. Transparency isn't just an ethical stance; it is a core adoption strategy.

A Resilient Growth Framework

The capabilities we've discussed aren't standalone pilots; together they form a flywheel. Forecasting informs inventory placement, so campaigns don't outpace stock. Next-order prediction aligns outreach with individual readiness. Segmentation and uplift modelling protect margin by directing discounts only to persuadable clusters. Predictive lead scoring focuses attention on feasible, high-intent opportunities.

When these signals are visible to a single cross-functional "Revenue S&OP" team, the business finally runs a single playbook—a synchronized system where marketing plans are conditioned by supply reality, and supply is steered toward marketing's next move. As shown in Figure 8.2, the framework is circular: signals from forecasting, NPD, segmentation, and lead scoring continuously inform each other, tightening the loop with every cycle.

Future Trends: Agents, Simulations, and Continuous Planning

In my opinion, two shifts will define the next 24 months of AI in business planning—and they are coming faster than most executives realize.

First, **multi-agent systems**. We are moving from isolated models to agent-to-agent (A2A) ecosystems where pricing agents, inventory agents, and marketing agents negotiate with each other in real time. Think of them as digital colleagues reconciling constraints against goals continuously. Instead of static monthly plans and quarterly reviews, businesses will be orchestrated by thousands of micro-decisions each day: where to place stock, which segment to discount, and which lead to prioritize. In this future, the role of the human strategist is not to micromanage but to act as the orchestrator—setting the goals, guardrails, and trade-off rules the agents work within.

Second, **generative simulation**. For decades, planning meant "set a budget and hope for stability." That world is gone. Teams will increasingly rehearse shocks before they happen. Tariff changes, weather anomalies, port congestion, or viral demand spikes can now be stress-tested in hours rather than weeks. Imagine a world where your campaign calendar is a living simulation, not a static plan: widen purchase windows, re-phase launches, and shift budget between regions—all rehearsed digitally before reality forces your hand.



FIGURE 8.2 Resilient growth framework
Source: Designed by the author.

And what ties this together is not more dashboards, but large language models (LLMs) acting as connective tissue. They will not decide inventory flows on their own, but they will explain why an agent suppressed a promotion in Germany, summarize risks across marketing, sales, and supply chain, and even draft the memo for the weekly S&OP. This "narration layer" is crucial. Without it, executives see only black-box decisions. With it, they see a coherent strategy they can endorse—or override.

The frontier, in my view, is not AI replacing planners but planners supervising networks of agents that keep plans aligned with reality. Strategic thinkers will become even more valuable, not less. In a world of A2A negotiations and endless simulations, someone has to decide what winning looks like. The danger is not that AI will act without us, but that organizations without clear orchestrators will drown in micro-decisions that pull in different directions.

This is why I believe the next great competitive advantage will not come from "who has the best model," but from who has the *best orchestrators*—leaders who understand both the technology and the business, and who can set strategy in a way machines can operationalize.

Conclusion

AI is not replacing marketers; it is sharpening them. In a volatile market, *availability is the new creative, timing is the new targeting*, and *margin protection is the new growth*. Predictive inventory ensures campaigns have stock. Next-order prediction ensures outreach arrives when intent peaks. Segmentation and uplift modeling ensure discounts land where they change behavior, not where they leak. Predictive lead scoring ensures sales chase opportunities that are both likely and deliverable.

Put together, these tools create resilient growth—growth that compounds because it is synchronized, sustainable, and self-improving. That is how brands spend less to sell more, protect margin while building loyalty, and deliver on the promise they advertise.

At the same time, let's be clear: AI today is both overrated and underrated. Overrated, because many executives report no significant ROI despite the hype—AI has become a buzzword, added to products and pitches whether or not it solves a real problem. Yet, it remains deeply underrated, as we are only just scratching the surface of its potential in pragmatic business use cases. This is not about humanoid robots or artificial general intelligence (AGI) fantasies, but about the real, near-term capabilities that can transform demand forecasting, bundling, personalization, and lead scoring.

In my opinion, the winners will be those who resist the temptation of "AI for AI's sake." Sometimes a simpler form of automation is the right answer—and that's okay. In fact, it may be more efficient, especially when we consider the energy and computing cost of large-scale AI. The real skill isn't in chasing the most fashionable algorithm, but in understanding the business problem first, and then choosing the right technology stack.

This is why human orchestrators remain indispensable. The next era will reward leaders with strategic, omnichannel, cross-departmental mindsets—those who can connect supply chain, marketing, and sales into a coherent system. Asking the right questions is half the problem solved. AI will provide sharper answers, but only to leaders who know what to ask. Real advantage will come not from adding AI everywhere, but from orchestrating it where it truly moves the business needle.

Author's AI Assistance Disclosure

This work was supported by two large language models (LLMs)—ChatGPT and Gemini—which I have personally trained with my own writing and professional experience. I leveraged their capabilities to assist with a few preliminary tasks, specifically outlining the initial structure and collecting recent source material. The text itself, including the key insights and analysis, was developed and refined by me, ensuring that my voice and perspective are at the forefront of the final manuscript. I am accountable for all the content presented in this chapter.

Technical Notes

TECHNICAL NOTE A: ARIMA/LSTM BASELINES

What they are:

- **ARIMA (AutoRegressive Integrated Moving Average):** A classic statistical model for time-series forecasting that captures linear autocorrelation and seasonality by differencing the series ("I"), regressing on its own lags ("AR"), and smoothing residuals with a moving average ("MA").
- **LSTM (Long Short-Term Memory):** A recurrent neural network architecture that uses gates (input/forget/output) to retain or discard information over long sequences, enabling modeling of longer-range, nonlinear dependencies.

Why they're baselines:

- They are widely understood, fast to train, and provide solid performance on many structured demand series—so new models (e.g., Transformers) are benchmarked against them to demonstrate incremental value.
- In production, teams may keep an ARIMA/LSTM "champion-challenger" setup to detect drift: when the frontier model underperforms, the baseline is a safety net.

When they struggle:

- ARIMA assumes linearity and stationarity; sudden regime shifts, promotions, or viral spikes break those assumptions.
- LSTM can model nonlinearity but may underperform on very sparse, long-horizon retail series and can be harder to interpret for business users.

TECHNICAL NOTE B: XGBOOST–MPG HYBRID FOR NPD]

Concept:

Combine **XGBoost** (tree-boosting on structured features) with **Modified-Poisson Gamma (MPG)** repeat-purchase modeling. MPG estimates **segment-level cadence** (how often cohorts return), while XGBoost focuses on **individual-level lift this week** given price, seasonality, and context.

Calibration workflow:

1. Use MPG to produce a **base cadence prior** per segment (e.g., weekly, monthly, seasonal).
2. Feed cadence, recency, price, promo depth, season, channel exposure, and exogenous signals into XGBoost to predict **week-level reorder probability**.

3. **Calibrate** probabilities (Platt/Isotonic) and set **trigger thresholds** (e.g., P(reorder) ≥ 0.42 with stock ≥ safety band) to launch "due soon" journeys.
4. Monitor **lift vs. baseline** and **suppression savings**; re-tune thresholds quarterly.

Why it works:

- MPG yields interpretable cadence ("how often").
- XGBoost captures nonlinear interactions ("who this week") with minimal feature engineering.
- The hybrid balances **accuracy** + **explainability**, improving trust and operational adoption.

TECHNICAL NOTE C: DYNAMIC PRICING TRANSPARENCY

Fairness bands:

- Define allowed price ranges by product/region (e.g., ±8% from long-run median), with explicit exceptions (e.g., expedited logistics).

Communication framing:

- Label adjustments clearly ("express delivery fee," "seasonal sourcing cost") and keep messaging consistent across channels.

Suppression rules under constraint:

- If inventory < threshold or forecast error spikes, **suppress deep discounting** and prefer bundles/value messaging.

Governance:

- Log reason codes for changes; enable manual override; audit weekly in Revenue S&OP.

TECHNICAL NOTE D: TRANSFORMERS VS. STACKING

Transformers (self-attention):

- Excel at **sequential forecasting** and long-range dependencies (e.g., next-order timing from months of baskets).

- Strengths: capture context across time; scalable; strong on irregular patterns when trained well.
- Caveats: data-hungry; heavier compute; less interpretable without tools.

Stacking (model ensembles):

- Not an architecture; a **strategy**: combine diverse models (e.g., logistic regression, XGBoost, shallow nets) and train a **meta-learner** to weight them.
- Strengths: robust on **structured decision problems** (lead scoring, segmentation); handles heterogeneous signals; often lighter compute.
- Caveats: requires careful validation to avoid leakage; interpretability depends on base learners.

Pragmatic guidance:

- Use **Transformers** for sequence-heavy tasks (NPD, basket flows).
- Use **Stacking** where different algorithms capture complementary aspects (lead scoring, uplift).
- Many teams do both: a Transformer for timing and a stacked ensemble for who/what actions.

References

Amazon Staff. (2025, June 11). Amazon announces AI-powered innovations in delivery, forecasting and robotics. Amazon. https://www.aboutamazon.com/news/operations/amazon-ai-innovations-delivery-forecasting-robotics.

Bendeich, M. (2025, August 13). Just in time? Manufacturers turn to AI to weather tariff storm. *Reuters*. https://www.reuters.com/business/just-time-manufacturers-turn-ai-weather-tariff-storm-2025-08-13/.

Bertini, M., & Koenigsberg, O. (2024, May). Dynamic pricing doesn't have to alienate your customers. *Harvard Business Review*. https://hbr.org/2024/05/dynamic-pricing-doesnt-have-to-alienate-your-customers.

Capirci, C. (2025, May). 3 hospital supply chain directors explain how AI is helping them manage critical inventory. *Business Insider*.

Carrington, D. (2025, May 27). AI tool trial could save equivalent of 1.5m meals in food waste. *The Guardian*. https://www.theguardian.com/environment/2025/may/27/food-companies-edible-waste-charities-ai-tool-trial.

De, T. S., et al. (2024). A machine learning and empirical Bayesian approach for predictive buying in B2B e-commerce. Proceedings of ICMLSC 2024.

Gartner. (2024). Market guide for retail unified price, promotion and markdown optimization. Gartner Research.

González-Flores, L., et al. (2025). A B2B lead scoring model based on machine learning. *Frontiers in Artificial Intelligence*, 8, 1554325. doi:10.3389/frai.2025.1554325.

Grigoraş, A., et al. (2023). Transformer-based model for predicting customers' next purchase day. *Computers*, 11(11), 210. doi:10.3390/computers11110210.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). The bullwhip effect in supply chains. *Sloan Management Review, 38*(3), 93–102.

McKinsey & Company. (2021). The value of getting personalization right—or wrong—is multiplying. https://www.mckinsey.com/business-functions/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying.

Persana AI. (2025). 8 AI sales case studies driving growth in 2025. https://www.persana.ai/blog/8-ai-sales-case-studies-driving-growth-in-2025.

SuperAGI. (2025). Case studies in AI lead scoring: Real-world examples of how businesses are boosting conversions in 2025. https://superagi.com/from-cold-to-sold-mastering-the-art-of-ai-driven-sales-conversations-for-maximum-conversion-rates/.

9

AI INNOVATION IN SUSTAINABLE SUPPLY CHAIN MANAGEMENT

A PRISMA Systematic Review

Melisa Petra Benchis and Elizabeth Edgal

Introduction

Somewhere between Brazil's soybean plantations and Berlin's supermarket shelves, decisions are being taken, quietly, anonymously, and at scale. A contract is signed. A shipment is rerouted. A warning flashes in a warehouse management system. These discrete steps might seem dull, but combined they form a global process network that needs managing. The supply chain encompasses all activities associated with the flow and transformation of goods from raw materials through to the end user, as well as the associated information flows. Supply chain management (SCM) is the integration of these activities through improved supply chain relationships to achieve sustainability (Handfield and Nichols, 1999). And, quickly explained, sustainable supply chain management (SSCM) integrates environmentally and socially responsible practices throughout the supply chain life cycle, from raw material sourcing to product delivery and disposal with the aim of minimizing negative environmental impacts, enhance social well-being, and ensure economic viability (Seuring and Müller, 2008). In this sense, this chapter opens with the question: *how does AI influence supply chain sustainability?*

Sustainable supply chain management (SSCM) combines environmental, social, and economic goals along the supply chain life cycle, from raw material to final delivery. But in practice, the process is not linear. The complexity of modern supply chains—passing borders, time zones, and multiple organizational stakeholders—has created a system hard to map and even harder to modify. Pressure to solve the puzzle is increasing. Climate change, exploitation of workers, and consumption of resources are no longer intellectual hypotheses. Today they are very real pressures to business practices, public trust, and overall well-being of the planet. Policy efforts like the United Nations Sustainable Development Goals (e.g., SDG 12 in regards to sustainable production and SDG 13 in regards to taking action on climate) have heightened expectations on transparency and accountability (Raman et al., 2023). In this frame, private and public actors are being compelled to reduce emissions, stop waste, and uphold

DOI: 10.4324/9781003632429-13
This chapter has been made available under a CC-BY-NC-ND 4.0 license

moral practices in their supply chains (Chauhan et al., 2022; United Nations Global Compact, 2023).

As a response to these pressures, artificial intelligence (AI) has emerged as an important technology in recent years, especially after the COVID pandemic with the boom generated by generative large language models. AI can learn, identify patterns, find causalities, and thus make potential predictions which offers solutions that appear custom-made to overcome the operation and ethical obstacles of supply chains. Companies are now employing AI to automate inventory management, anticipate interruptions, conduct regulatory screening, and detect abnormalities through real-time logistics data streams (Qu and Kim, 2024). In the post-COVID-19 world, interest in AI has only intensified. In this light AI has become a foundation stone for building more responsive, resilient, and sustainable supply chain. So far, the evidence looks promising. A growing body of research suggests that AI can reduce operational waste, improve forecasting accuracy, and support cost-efficient decision-making in SSCM (Brintrup et al., 2023; Chauhan et al., 2022; Farooq and Yuen, 2024; Frimpong, 2025; Sanders et al., 2019; Vann Yaroson et al., 2025).

However, while the environmental and economic impacts of AI in SSCM have been addressed relatively extensively, there remains a research gap in social sustainability and inter-organizational dynamics. Many studies focus on how AI improves environmental performance (e.g., reducing waste or optimizing resource use) and economic performance (e.g., cost efficiency, improved forecasting). However, less attention have been paid to how AI technologies can be leveraged specifically to strengthen the social dimensions of SSCM, such as enhancing the quality of inter-organizational relationships, improving labour conditions, or promoting equity among supply chain partners (Benzidia et al., 2021; Qu and Kim, 2024). Also, limited research has delved into the mechanisms through which AI-based applications can improve inter-organizational coordination, knowledge sharing, and trust-building (Naz et al., 2022). This gap suggests a promising area for future research, focusing on the 'human–AI collaboration' aspect within supply chain ecosystems. Questions like: Is AI capable of establishing advanced collaboration? Can it construct mutual power sharing and shared meaning in global networks? Or might it intensify existing differences even more through black box algorithms and automation?

Also, much of the existing literature is not industry specific. Deployments of AI in agriculture, fashion, food logistics, or humanitarian supply chains are necessarily different, not only technically, but also socially and in terms of regulatory context. Yet many reviews lump these deployments together under an umbrella without isolating out the unique challenges each sector presents (Dubey et al., 2020; Kanyepe et al., 2025). At the same time, although there are numerous theoretical arguments regarding the potential of AI to aid in sustainability, evidence is limited. Case studies emphasize large, multinational corporations, companies that can afford to experiment with AI and have successful applications in practice. This is a distortion of small and medium enterprises (SMEs), especially from the Global South. It is also unknown the extent to which the benefits proposed by AI are pervasive or scalable (Badghish and Soomro, 2024; Hernandez et al., 2023).

Also, SSCM AI is often not interdisciplinary. The operational research has its eyes on algorithm performance, the management science looks at organizational

transformation, and ethicists is focused on the governance aspect, but none of these are typically integrated into an integrative model (Brintrup et al., 2023; Pennekamp et al., 2023). Finally, while previous reviews and commentaries have been able to identify some of the topics of immediate concern in AI and supply chain management; there is still a need for structured, evidence-based synthesis aimed at sustainability issues within this field. Our evidence indicates that while academic interest in AI for sustainable supply chains is rapidly growing, literature remains fragmented by disciplines, methods, and conceptual foundations. This fragmentation not only hides dominant themes and methodological tendencies but also obscures our understanding of industry-specific uses, theoretical bases, and normative implications.

Thus, gaps within the current literature on AI and sustainable supply chain management arise with a few more relevant, although somewhat disconnected, factors. First, while many articles explore the environmental and economic dimensions of sustainability, few acknowledge the social dimension (equity concerns, workforce conditions, trust, governance, etc.). Second, most of the literature review articles focus on multinational enterprises, meaning little is known about the SMEs efforts at AI-infused sustainability, particularly in the Global South. Third, few connect the literature reviewed across industries (agricultural textiles and food manufacturing, for example), meaning findings in the real world are not as strong because cross-industrial differences are not typically acknowledged. Finally, research into AI and sustainable supply chain management is broadly interdisciplinary and fragmented, calling for a distinction between technical versus managerial versus ethical differences instead of a unifying approach. This gap is bridged through systematic research to provide a reconciliatory review that would bring together various ESG perspectives of AI as a viable agent for sustainable supply chain management efforts to combine technological and organizational perspectives for a holistic approach.

In seeking to fill that gap, this review uses the PRISMA framework in a systematic way across a wide range of peer-reviewed literature (more details about the methodological process in Annex 1). The process involves not merely examining what technologies are highlighted and their impact on supply chain performance, but also how values like transparency, ethics, inclusivity, and resilience are being translated, or disregarded, into practice and research. This study summarizes four main thematic streams where the field is evolving: (1) AI as a driver of performance and optimization; (2) ethics-enabled AI resilience and risk management; (3) AI and sustainability and circular economy shifts; and (4) the emergence of ethical, explainable, and human-centric AI. The goal was not simply to highlight the theoretical summary of AI in sustainable supply chains, but to offer a roadmap for the future directions of AI use in SSCM and interorganisational relationships.

A PRISMA Literature Review of Trends in the Literature and Future Directions

As this chapter has shown previously, the path towards sustainable and collaborative supply chains driven by AI is as much one of architecture and ethics, as it is of algorithms. Creating systems that are fair, transparent, and inclusive requires more than isolated technical solutions, it requires a shared vision supported by structures of

governance and carefully made design choices. But while these needs are increasingly recognized in practice, the question remains: how are they identified and dealt with within the academic research that shapes future innovation? To gain a better understanding of the current state of knowledge, and to see where research is headed, this chapter now changes to a systematic literature review.

PRISMA Methodology

The research design employed to conduct this study was consistent with the PRISMA 2020 guidelines, aiming for an open and reproducible systematic review of literature. PRISMA, or Preferred Reporting Items for Systematic Reviews and Meta-Analyses, was initially developed to improve reporting quality of systematic and meta-analytic reviews. The development of PRISMA reflects growing demand for accurate, thorough reporting of scholarly research. Following these guidelines, our aim was to enable readers to understand in depth the process involved in seeking out, screening, and choosing the application of AI to supply chain management literature (Haddaway et al., 2022; Meline, 2006; Moher et al., 2009; Page et al., 2021). Review of literature included scholarly journals between 2023 and 2025, all of which were English for easy understanding by the authors.

To get a representative and inclusive sample of the existing academic literature on the topic, we searched several leading databases (11 databases) as follows: SSRN, ERIC, IEEE Xplore, Science Direct, MDPI, JSTOR, Web of Science, Google Scholar, Emerald, Scopus, and Springer. This list includes the most important sources of peer-reviewed articles around AI, supply chain management, and allied subjects. Our keywords were constructed around three phrases primarily: 'AI in supply chain management', 'AI in sustainable supply chain management' and 'AI in supply chain management in interorganizational relationships'. Our queries in the chosen databases resulted in a total of 4728 initial articles/entries. To limit it to peer-reviewed journal articles only, we first excluded books, conference papers, reports, editorials, and any documents that were not classified as an academic journal article. This left us with 4586 articles to screen.

To eliminate duplicate entries, the results that were imported into Zotero reference management software. With Zotero's built-in automatic duplicate detection tool, we identified and removed 477 duplicate entries, leaving us with a list of 4109 unique articles. The second step involved a manual screening of the title mainly and if necessary for clarity we also read the abstract. We excluded articles who's abstract or title had no substantial relevance to the study of AI in light of supply chain management. After filtering through the articles, we were left with 237 articles (a lot of the excluded articles approached the context of technology in Education and Medical Supply so they were excluded due to the business orientation of AI use in SSCM).

The full text (pdf) could be retrieved for only 183 articles from among these due to open access or private access perspective. The full-text screening was then carried out to determine if each article aligned with the purpose of the study. Three exclusion criteria were applied at this level. First, 48 articles were ruled out because they focused on AI, but not particularly in supply chain context. Next, 18 articles that referred to AI without making any tangible link to sustainability in supply chain or sustainable

supply chain management were excluded. Third, four articles were excluded because of the weak focus on AI as a technology between technologies e.g., blockchain or Internet of Things (IoT), which gave AI no overall attention.

After this iterative process of screening, 113 articles were chosen to be used in the final review. Inclusion and exclusion criteria were checked at every stage of selection by collaborative author discussion, reducing bias and enhancing methodological rigor. The whole search, screening, and choice process was mapped in the PRISMA model, granting a clear model that can be reproduced by future research. A flowchart of the phases of the individual selection process can be found in Figure 9.1 and the articles are presented in Table 9.1.

Findings

Theme 1: AI as a Catalyst for Supply Chain Performance and Optimization

One of the clearest trends emerging from the remaining entries is the placement of AI as a central driver of supply chain disruption. Across a wide range of research, AI is consistently framed as a catalyst for shifting supply chains from disconnected (reactive) systems to anticipatory, integrated systems; capable of managing complexity with precision and speed. This transformation is most obvious across performance areas like forecasting, inventory planning, logistics planning, and supplier evaluation. All are being reimagined through the prism of machine intelligence.

In this collection of works, there is a unifying emphasis on the potential of machine learning (ML), deep learning (DL), and mixed computational models to surpass traditional techniques across different supply chain functions. Techniques such as long short-term memory (LSTM), convolutional neural networks (CNNs), genetic algorithms, and reinforcement learning form the backbone of such transformations, delivering not just quicker predictions but also more context-aware, adaptive responses. Such outcomes suggest that for most researchers, AI is a shift in conception that pursues supply chain performance.

Forecasting Demand with Precision and Depth

AI acts as a facilitator for efficiencies made in nearly all SCM functions from forecasting to logistics to inventory and production planning. Researchers across industries note how AI SCM optimization provides developments that are quicker, more efficient and more adaptive. From a winery standpoint, an industry where research has been conducted on AI within an SCM context (Jauhar et al., 2025), AI customer segmentation and predictive analytics can help producers realize their production efforts based on anticipated demand. This means less overproduction efforts, less energy spent on refrigerated powered efforts, less waste via products going bad that should never have been created in the first place. Thus, AI customer segmentation is not just a marketing effort, but a sustainability effort born of better forecasting and the impact of resources on degradation.

For instance, AI encourages inventory predictive analytics and warehouse management systems that create a predetermined organization of resources that are neither

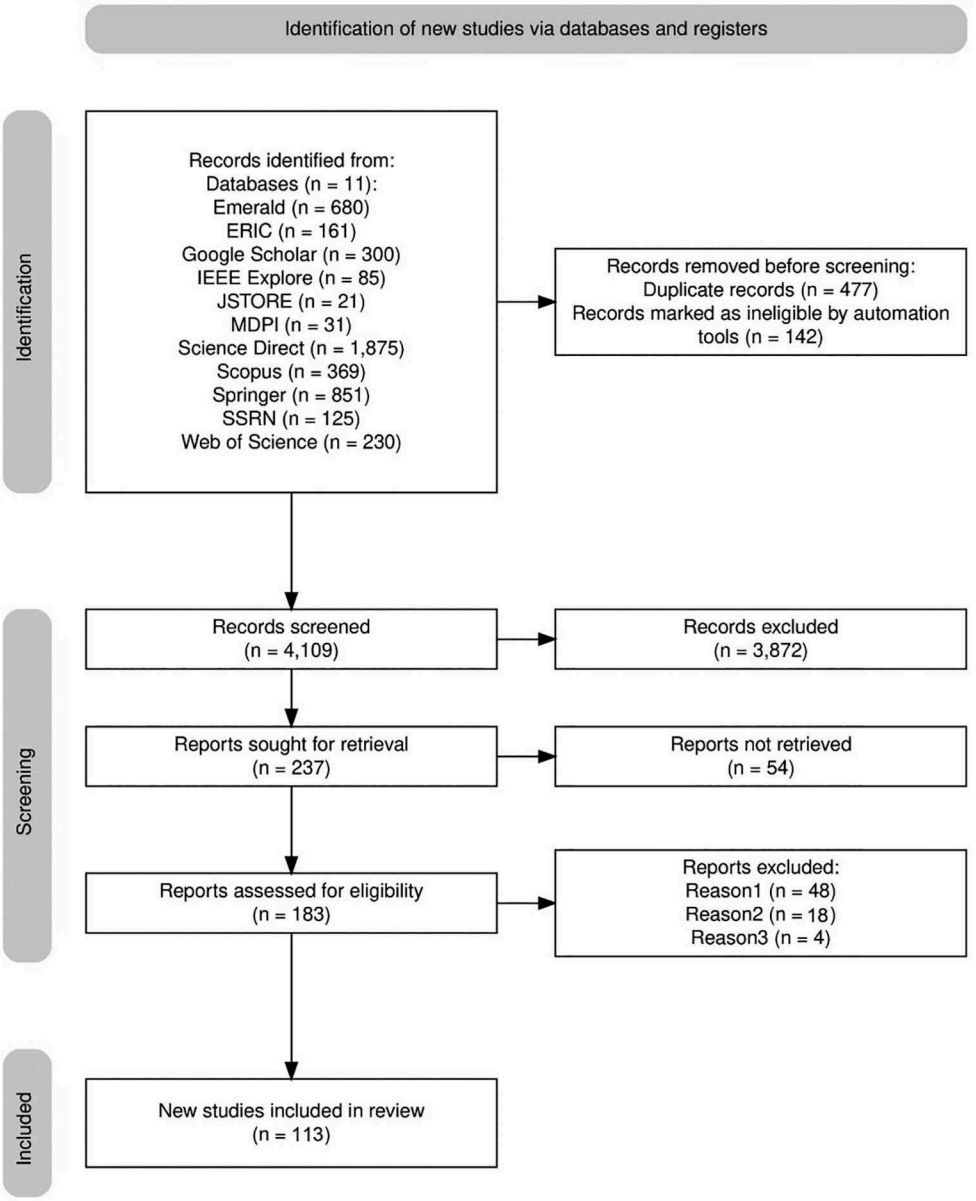


FIGURE 9.1 PRISMA flow diagram
Source: author's research and Haddaway et al., 2022 creation tool

too much stock (inefficient and unsustainable) nor stock outs (equally inefficient and unsustainable). Either extreme is inherently unsustainable as it forces expensive repositioning, additional powered warehouse systems, and degradation.

Furthermore, when AI is used as a performance enabler, it broadens the very definition of performance to include new sustainability-oriented values. For instance, AI's predictive capabilities can reduce uncertainty, prevent over-production, and avoid

TABLE 9.1 Included papers and thematic coupling

Aggregate Dimensions	*Themes*	*Papers*
Theme 1: AI as a Catalyst for Supply Chain Performance & Optimization	Forecasting Demand with Precision & Depth	Elmir et al. (2023); Danach et al. (2024); Ghouri et al. (2023); Mohammed et al. (2025); Ismaeil & Lalla (2024); Khadem et al. (2023); Logožar (2025); Mañay et al. (2024); Mohsen (2023); Paraschos (2025); Walter, A., Ahsan, K., Rahman, S. (2025); Zhu et al. (2023); Jauhar et al. (2025)
	Reinventing Inventory Control & Warehouse Management	Ahmed et al. (2024); Riad, Naimi, & Okar (2024); Aylak (2025)
	Optimizing Logistics & Last-Mile Delivery	Alheadary (2024); Eyo-Udo (2024); Roldán, Maqueira & Moyano (2023)
	Supplier Evaluation & Predictive Procurement	Kaarlela (2024); Rakholia et al. (2024)
	Digital Twins, Automation & ERP-AI Integration	Amir & Gupta (2024); Cannas et al. (2024); Georgiev et al. (2024); Hendriksen (2023); Jackson et al. (2024); Khan & Abdul (2024); Kulkarni & Bansal (2024); Nitsche et al. (2024); Panket, Wannapiroon, & Nilsook (2024); Pattnaik et al. (2024); Raab, Fischer & Brandtner (2025); Rana & Daultani (2023); Khadem et al. (2023); Thakur et al. (2023); Walter, S. (2023); Wang, Weizhong et al. (2024); Younesse et al. (2025)
Theme 2: AI for Supply Chain Resilience & Risk Management	Predictive & Proactive Risk Management	Attah et al. (2024); Bhattacharya et al. (2024); Effah et al. (2023); Enz et al. (2024); Hirsch et al. (2024); Kashem et al. (2023); Li X. et al. (2024); Mehmood S. et al. (2025); Masoomi et al. (2025); Naimi (2024); Grover, N. (2025); Olufunke (2024); Ordibazar et al. (2025); Mittal & Panchal (2023); Yıldırım Özmutlu & Arun (2025); Zamani et al. (2023)
	Theoretical Frameworks & Empirical Proofs	Belhadi et al. (2024); Dai et al. (2025); Dey et al. (2024); Gupta et al. (2024); Jia Fu et al. (2024); Kassa et al. (2023); V. Kumar, S. Kumar, Chatterjee, Mariani (2024); Le Thanh Tiep & Behl (2024); Singh, R.K. et al. (2024); Virmani et al. (2024)
	Generative AI & Human-in-the-Loop Models	Agrawal et al. (2024); Dubey et al. (2024); Khlie et al. (2024); Li L.et al. (2024); Li L.et al. (2024); Meafa, Benabdellah & Zekhnini (2025); Kosasih & Brintrup (2025); Samuels (2025)
	Ethical & Responsible Risk Management	Brintrup et al. (2024); Sadeghi et al. (2024)

Aggregate Dimensions	*Themes*	*Papers*
Theme 3: AI, Sustainability & the Circular Economy	Smart Optimization for Sustainability	Huang & Mao (2024); Panigrahi et al. (2025); Podrecca et al. (2024); Sahoo et al. (2024); Singh, A. et al. (2023); Wu et al. (2024)
	Circular Economy Enablement	Bhattacharya et al. (2024); Genc (2024); Khan et al. (2025); Mirzaee, H. & Ashtab, S. (2024); Salehi et al. (2024)
	AI & Green Supply Chain Metrics	Amin, Chakraborty & Baldacci (2025); Belu and Marinoiu (2025); El Saadani Gazar et al. (2024); Kurniawati & Cakravastia (2023); Yadav et al. (2024)
	AI-Blockchain Convergence for Circularity	Abdelhamid, Sliman and Ben Djemaa (2024); Adamashvili, Zhizhilashvili & Tricase (2024); Almuqren et al., 2023; Elsayed (2024); Hong & Xiao (2024); Kusnadi et al. (2024); Moetez et al. (2024); Ostroumov anf Wycislak (2025); Saidu et al. (2025); Singh, A.K. et al. (2023); Ara et al. (2023); Tara et al. (2023)
Theme 4: Ethical, Explainable & Human-Centric AI in Supply Chains	Explainability & Trust	Sharabati et al. (2024); Wang, Wenyi, et al. (2024)
	The Ethics of AI Labour & Impact Sourcing	Gonzalez-Cabello et al. (2024); Muldoon et al. (2023)
	Reputation, Risk & Transparency	van Hoek and Richey Jr (2025)
	Human-Centric AI for Empowerment & Inclusivity	Al Harrasi et al. (2023); Bas et al. (2023); Deghedi (2023); Gaikwad (2024); Raman et al. (2024); Weisz, E. et al. (2025)
	Designing for Ethical AI	Hao & Demir (2024); Robertson et al. (2025); Robertson et al. (2025); Xinyue H. & Emrah D. (2024)

Source: author's work

excess storage—all of which translate into lower waste and lower cost. In this sense, performance is understood as any effort that enhances the value of sustainability itself.

Reinventing Inventory Control and Warehouse Management

AI is also delivering measurable gains in inventory optimization. Reinforcement learning and expert systems now dynamically trade off stock levels against lead times, perishability, customer behaviour, and storage costs. Studies have reported stockout reductions, Riad et al. in particularly offering percentages, where stockouts reduced from 9% to 2%, delivery times cut from 8.5 to 4.2 days, production efficiency increased from 74% to 91%, and customer satisfaction rose from 70% to 88%, freeing up working capital and reducing waste (Riad et al., 2024).

AI makes warehouses more efficient through predictive restocking, robot automation, and layout optimization. Algorithms try out different configurations to figure out the most efficient picking routes and spatial usage. AI-powered warehouse systems improved delivery efficiency by 27.2% and reduced overstocking by more than 21.8%

in a real-world application over the span of 12 months (Aylak, 2025). Such improvements are especially critical for businesses handling perishables or fluctuating demand. Here, AI combines environmental and behavioural data to deliver products where they should be, when they should be there, minimizing loss and maximizing responsiveness.

Optimizing Logistics and Last-Mile Delivery

Artificial intelligence has also revolutionized transportation and logistics planning, particularly in the 'last mile' industry, which is the operationally most challenging and most costly segment of the delivery business. AI route platforms use real-time traffic, fuel-efficient information, weather, and delivery limitations to provide the optimum routes. This is done by reducing idling time, fuel saving, and on-time delivery. One study demonstrated that fuel consumption dropped by 22.8% and carbon emissions were reduced by 30.3% due to AI-based routing (Aylak, 2025). Additionally, AI-based systems utilizing NLP (Natural Language Processing) and computer vision are now being used in vehicle diagnostics and logistics monitoring systems to support predictive maintenance and prevent delay-inducing breakdowns. These systems also facilitate customer satisfaction through improved delivery time estimates, dynamic rerouting in case of disruption, and automated delivery status communication, all important in the present e-commerce era.

Supplier Evaluation and Predictive Procurement

AI capabilities extend far beyond internal functions. Its role in supplier assessment and procurement is also coming into the spotlight. Algorithms now sift through unstructured and structured data like performance records and financials, geopolitical trends, and social compliance audits, to rank and assess suppliers on multi-criteria decision-making.

This future-proof procurement allows businesses to anticipate supplier risk, ensure ethical sourcing, and drive more effective negotiations. In AI-driven platforms, procurement managers are no longer restricted to mere price and delivery performance because they have real-time dashboards evaluating supplier resilience, compliance with sustainability, and innovation value.

Digital Twins, Automation, and ERP–AI Integration

Among the most visible advancements in AI implementation is the creation of digital twins and their integration within enterprise resource planning (ERP) systems. Digital twins are virtual replicas of supply chain physical processes, and it allows for real-time simulation, disruption testing, and performance analysis. Coupled with AI, these systems become predictive engines with the capability to not only identify inefficiencies but also suggest adaptive solutions.

A good example is the SustAI–SCM model that integrates transformer-based AI with real-time procurement, logistics, and sustainability indicator data streams. The model achieved 28.4% overall operational cost dropped, along with fuel consumption, emissions, and delivery time reductions. Even in halal logistics (HSC), AI was

integrated within ERP systems to provide halal compliance through computer vision and rule-based expert systems, demonstrating both performance and cultural fit. Hybrids of ERP and AI are transforming enterprise planning through decision-making alignment across functions and time horizons.

Technically, aside from performance benefits, AI even pays strategic dividends: real-time responsiveness, greater visibility, and operational alignment to sustainability and ethics imperatives. But the value of AI is not just in substituting human decision-makers. Instead, its real potential is in enhancing human judgement, facilitating collective and contextual decision-making that is quicker, better-informed, and more visionary. With digital twins, automations, and AI–ERP integration driving future changes, the supply chain is being transformed from a linear activity to a living system, smart, responsive, and resilient.

Theme 2: AI for Supply Chain Resilience and Risk Management

AI is increasingly seen as a supply chain resilience facilitator that promotes predictions, responses, and recovery of disruptions. Yet AI is more than a supply chain efficiency facilitator; AI allows firms to predict potential disruptions, moving from a reactive focus on crisis mitigation to data-driven crisis anticipation. Thus, the integration of AI capabilities and their connection to dynamic capabilities view (DCV) and the resource-based view (RBV) provide a uniquely differentiated contribution of AI to firm-specific and network-oriented adaptability capabilities.

Thus, this study views collaborative resilience as the shared adaptive capacity of interdependent actors to predict, absorb, and recover from disruptions through information sharing, decision alignment, and trust-enhancing support. Resilience, in general, is the recovery capacity of any actor. Yet, collaborative resilience represents network learning and collaborative adaptation, how a collection of interconnected partners can adjust to avoid challenges in the first place, supported by AI-leveraging information collection and aligned preparation.

These perspectives are supported by evidence from a survey study of Vietnamese SMEs (n = 477), which shows that AI-enabled collaborative facilitation functions as an antecedent of resilience understood as a relational and systemic capability, rather than something developed by individual firms in isolation. Using structural equation modelling (SEM), the study examined the links between AI and resilience from competing theoretical viewpoints. The results indicate that the relationship is indirect: the strongest mediators are learning orientation and collaborative flexibility.

Practically, AI tools such as predictive analytics and digital twins enhance supply chain visibility by detecting disruptions early and enabling more effective communication among partners, thereby accelerating recovery responses. This reinforces the conclusion that AI does not independently make resilience-enhancing decisions; instead, it strengthens collective awareness, shared learning, and coordinated adaptation. When firms have AI systems they can apply in real time and use collaboratively, their situational awareness, and, consequently, the awareness of the entire supply network improves in an adaptive manner.

Predictive and Proactive Risk Management from Insight to Foresight

At the centre of resilience lies pre-emption and disruption anticipation. Multiple studies highlighted the way in which AI-based technologies like natural language processing (NLP), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) enable real-time monitoring of risks on both structured and unstructured sources of data. The models learn on multiple streams like sensor streams, weather conditions, news streams, and even sentiment on social media, to recognize warning signals for potential supply chain trouble.

For instance, with the case of weather extremes caused by climate change in cocoa value chains, AI-powered models anticipated transport disruption, loss at farm levels, and disruption at midstream levels with jarringly high precision. AI-powered dashboards and cognitive mapping graphically visualized cascading risk across nodes of supply, allowing decision-makers to act ahead of failure spillover.

The capability of AI to model 'what if' scenarios, particularly in tandem with digital twins, was reiterated repeatedly as well. These digital replicas of actual supply chain circumstances, along with AI scenario modelling, allowed firms to simulate the conditions of hypothetical disruptions (e.g., port shutdown or supply disruption) and pre-pose response strategies. In facilitating such a shift from reactive to proactive planning, AI recontextualized the notion of resilience: no longer as a protective buffer against risk, but as a smart flexibility embedded in the system itself.

Theoretical Frameworks and Empirical Proofs

To investigate how precisely AI crystallizes resilience, various studies built conceptual and empirical frameworks, typically grounded in dynamic capabilities view (DCV) and resource-based view (RBV). Both theoretical perspectives place AI not only as a tool but as a strategic resource that enhances a company's capability to sense, seize, and reconfigure in the face of change. An empirical investigation using structural equation modelling (SEM) tested 477 Vietnamese SMEs and showed that AI's strongest influence on resilience was indirect, mediated through dynamic capabilities such as organizational learning and process reconfiguration. The study also showed that open innovation (OI) significantly complemented this effect where *companies that collaborated in conjunction with external partners had higher agility and responsiveness gains.*

Similarly, adaptive capabilities were the key moderators of other studies. In Pakistani SMEs, for example, AI effects on sustainable business performance were strongest when companies had high organizational adaptability, meaning that the benefits of AI are achieved not by technology but by the structural and cultural flexibility of the adopting firm. But the most significant aspect in these models is that they displayed collaborative resilience, *were AI enhanced coordination between supply chain stakeholders, reinforcing interorganizational relationships through data exchange, joint planning, and real-time responsiveness.* AI-powered supplier evaluation, data exchange in real time, and automation of communication *were found to enhance visibility and responsiveness at the network level*, converting individual organizational resilience into ecosystem-level resilience.

Generative AI and Human-in-the-Loop Models

Generative AI (GenAI) also entered resilience strategy territory, specifically in risk modelling and disruption simulation, with much pomp. Leverage off massive amounts of data and live inputs, GenAI-based systems were able to forecast supply chain bottlenecks, formulate other sourcing approaches, and even suggest risk-optimized stock policies. A comparison study between GenAI transformers like GPT, BERT, and RoBERTa found that GPT variants were highly rated for adaptability and real-time risk handling, skills important for supply chains handling high-stress sectors like food and healthcare. Most importantly, these GenAI models were built on human-in-the-loop architectures that integrated computational prediction and managerial judgement. This intersection, in addition to improving operating visibility, translated AI systems into strategic thinking, allowing for more contextually dense decisions. It was an alternative template for crisis decision-making: augmented intelligence, not artificial.

Subtheme: Ethical and Responsible Risk Management

Besides technical resilience, the research highlighted another critical component: reputational and ethical resilience. In open digital supply chains, where all the stages can be tracked and audited, reputational risk is no longer siloed; it is shared. Labour exploitation by a remote supplier, once easily hidden, can now instantaneously engage global brands through traceable AI and blockchain systems. Several articles addressed this 'paradox of transparency'—while AI increases visibility and compliance monitoring, it simultaneously puts companies under public and regulatory inquiry. Businesses must address then ethics, fairness, and explainability not as an afterthought, but as design requirements in their AI systems.

Explainable AI (XAI) emerged as the solution of choice for this problem. By providing explainable logic paths for AI-generated decisions, whether in supplier selection, fraud detection, or disruption management; XAI software restored managerial confidence and stakeholder trust. For example, for perishable goods supply chains, segmentation-based demand forecasting models used Shapley values to rationalize customer behaviour segments so that companies could tailor strategy and communication without prejudice. In high-stakes contexts like humanitarian logistics, ethics are more loosely defined. There, AI was used to track population movement, quantify fairness in aid distribution, and reduce fraud, usually in time-limited and resource-constrained situations. Studies emphasized that AI must remain sensitive to human dignity and power disparities, particularly when used on vulnerable groups. Equally, in the fashion and food industries, AI and blockchain provided traceability and ethical sourcing regimes but remained plagued by data integrity issues, model bias, and equitable access. Time and again, the literature recommended strong regulatory climates, innovation cultures that were inclusive, and AI governance mechanisms that prioritized above all transparency, privacy, and fairness.

Theme 3: AI, Sustainability, and the Circular Economy

Leveraging its position in resilience, AI is currently helping supply chains graduate from survival to sustained regeneration. Just as the literature suggests, resilience is not recovery, it is about recovering forward. The same philosophy supports research that has been investigating how AI enables sustainability and circularity. Here, AI is not a systems optimizer but strategy for driving ecological intelligence, one that helps companies design supply chains that are not only resilient but sustainable.

Across the discussed research, one trend stands out is that AI is increasingly being used to balance operational efficiency with ecological stewardship. In combination with technologies like IoT and blockchain, AI makes supply chains traceable, adaptive, and resource constrained.

Smart Optimization for Sustainability

The most direct AI contribution to sustainability is its ability for resource flow optimization-reducing waste, emissions, and unwanted production. Several research works outlined how AI optimization aided inventory management, route planning, and energy usage with tangible environmental benefits. For instance, the SustAI–SCM framework employed transformer-based agentic AI to reduce 22.8% of fuel and over 30% of carbon emissions and reduce overstock and delivery waste. In real time, AI can reconfigure logistics plans to avoid high-emission routes, reroute deliveries around climate disruptions, and predict equipment failures before they result in wasteful downtime. In addition, ML and DL models have been used in demand forecasting to ensure production is suitably aligned with customers' needs. This prevents over-production, which is one of the primary reasons for environmental wastage in fashion and fast-moving consumer goods sectors. As one study illustrated, employment of LSTM-based forecasting models reduced wastage of food and holding cost of inventory simultaneously, particularly in perishable supply chains like wine and cocoa.

Circular Economy Enablement

Whereas AI makes sustainable production easier, it is also instrumental in enabling circular economy models (circuits in which products, materials, and resources stay in circulation as long as they are valuable). AI helps here in three ways: design, recovery, and feedback.

In product design, generative AI technologies are used to simulate materials and production methods that extend product life cycles or simplify disassembly. In recovery, AI facilitates better reverse logistics using smart diagnostics, predictive sorting, and waste classification systems, especially when coupled with computer vision. For example, in electronic waste recycling, AI can separate reusable components from scrap so that recovery is more efficient, and landfill pressure is reduced. Within food supply chains and the halal supply chain, AI was utilized to classify and verify products on the basis of religious and sustainability standards—ethical compliance as well as wastage reduction. Platforms based on blockchain with AI functionality like ABISChain[1] further amplify traceability and authenticity within such chains, such that circular behaviours become end-to-end auditable and reliable.

Feedback loops are at the centre of every circular economy, and AI makes it possible to learn in real time from every activity across the supply chain. With predictive analytics and reinforcement learning, businesses can continue to optimize sourcing, manufacturing, and distribution strategies to reduce environmental footprint. Digital twins also make this function possible, enabling businesses to simulate different sustainability scenarios and apply circular design principles with increased confidence.

AI and Green Supply Chain Metrics

Several studies developed quantitative models that connect AI capabilities to sustainability key performance indicators (KPIs). In a case study on electronics manufacturing, the deployment of AI yielded measurable improvement on several dimensions: delivery time was halved, stockouts fell by 78%, and production efficiency rose by 17%. Environmental performance was not sacrificed but improved simultaneously with efficiency, validating that economic and ecological gains are not working against each other when AI is applied correctly.

One of the primary conclusions drawn relates to the discrepancy between internal sustainable achievements and consumer understanding. Kurniawati and Cakravastia (2023) highlight that sustainability improvements within organizations are not always effectively transmitted to end consumers, particularly in complex supply chains where transparency remains limited. Similarly, Yadav et al. (2024) emphasize that without clear communication and visibility, sustainability efforts are unlikely to influence consumer awareness or behaviour. As a result, organizations may achieve substantial sustainability outcomes, such as reduced carbon footprints or improved energy efficiency, yet fail to realize reputational benefits due to this lack of accessibility. However, when sustainability performance is communicated effectively, internal improvements become externally visible and valued, for example through transparent reporting, eco-labels, and increased consumer awareness.

Moreover, although the literature often highlights the electronic and food industries, AI-driven sustainable performance is also emerging in sectors such as agriculture, textiles, and automotive manufacturing. For instance, precision agriculture utilizes AI to avoid excess fertilizer and decreases agricultural water use. Textiles involve predictive analytics through a better understanding of labour and resources for recycling effort. Finally, the automotive industry relies on machine learning to understand recycling capabilities for materials taken from end-of-life vehicles. This is important because AI-generated understanding operates like a web across industries to determine sustainable possible.

Ultimately, AI not only provides more environmentally friendly performance, but it also cultivates a feeling of circularity across systems, whereby digital intelligence connects production and consumption and repurposing efforts in a way that best presents itself for sustainability timeliness.

AI–Blockchain Convergence for Transparency and Circularity

Some of the most hopeful applications of green SCM have their foundation in putting together AI with blockchain. By joining up they provide real-time visibility, as well as

decentralized and secure access to international supply chains. AI forecasted and inquired on the information while blockchain protects it for authenticity and tracking purposes. Maersk's TradeLens, IBM's Food Trust, and Indonesia's halal logistics networks provided case studies illustrating how such interlinked systems enabled automated compliance, fraud discovery, and real-time verification necessary for both environmental monitoring and ethical sourcing. Smart contracts were written to halt the movement of goods if non-halal ingredients were detected by AI classifiers in one application case, offering end-to-end religious and safety compliance.

Along with product tracing, AI-driven blockchains facilitate resource-sharing models like collaborative sharing of inventory or leasing out of assets among competitors. Such models illustrate the role of interorganizational relationships in enabling adaptive and low-waste systems, supported by AI and blockchain infrastructure. And also, they reduce overproduction, reduce capital expenditure, and allow for adaptive, low-waste systems, particularly in high-emitting industries like agriculture, cement, and energy. Lastly, the literature discussed here creates a rich vision of AI not as a maximization of current systems but as a transformational force, capable of redefining supply chains to reflect the values of a circular, just, and sustainable economy. But this transformation is not spontaneous. It requires design intentionality, governance inclusiveness, and a readiness to rethink what value in business means. AI incorporation in the sustainable supply chains then must be offered, not only as a technical success but as a strategic and moral choice. Companies that take this route, making investments in clever, open, and transparent AI solutions, will thrive not only throughout Industry 5.0, but will set the trail for an era of industrial renewal where sustainability and intelligence go hand in hand.

Theme 4: Ethical, Explainable, and Human-Centric AI in Supply Chains

As the literature increasingly situates AI as an enabler of sustainable and regenerative supply chains, a deeper level of responsibility is visible. Circular by design, and with consequences thought through over the long term, it isn't just system optimizing, it's valuing technology. The shift toward smart, adaptive supply chains requires an essential question: how to enable sustainability through AI without losing fairness, transparency, or human dignity?

Explainability and Trust

One of the most prominent problems identified in a number of these studies was explainability of AI: an impediment which not only hinders organizational adoption but also weakens trust with end consumers and decision-makers. In highly regulated industries such as food, healthcare, and humanitarian logistics, explainability is imperative for compliance, auditing, and accountability.

Several studies explored the use of Explainable AI (XAI) methods to break open the AI 'black box'.[2] One good example was in the wine industry, where customer segments were modelled using long short-term memory (LSTM) algorithms and Shapley values. Through model behaviour explanation by customer features, e.g., income, wine expenditure, and frequency of purchase, businesses were able to tailor demand forecast

with knowledge of the rationale behind predictions. This level of openness not only increased the reliability of predictions but also allowed managers to make rational but ethically sound decisions.

Another notable contribution was a neuromyotonic reasoning framework proposed for link prediction in supply chain knowledge graphs. By inserting logical reasoning paths into graph neural networks (GNNs), the model achieved high predictive performance as well as ex-post explainability. It was able to demonstrate, for example, that a supplier had been identified as high risk because of common connections with lower-rated companies or spatial patterns, allowing human practitioners to check and trust the logic of the AI.

The Ethics of AI Labour and Impact Sourcing

While AI promises automation and efficiency, its fundamental system is built on low-wage, outsourced labour, especially in data annotation, a function essential to training machine learning models. Ethnographic research on Sama, a major AI annotation firm that operates in Kenya and Uganda, exposed this paradox. Despite Sama's self-marketing as an 'ethical AI' firm, field research revealed unsettling facts: low wages, surveillance-based management practices, gendered exploitation, and precarious contracts. This study gave a sobering reminder that supply chains of AI is also engaged with ethical risk, even if they are meant to optimize or augment other supply chains. It forces the concept of 'impact sourcing' necessarily virtue and instead requires strong labour rights, regulatory accountability, and transparency at every level of AI design. Interestingly, in response to public exposure and pressure from organizations like Fairwork, Sama changed, raising wages and providing longer contracts. This outcome shows that sustained pressure, public exposure, and worker advocacy can create genuine change in the ethics of AI work. It also reinforces a central thesis of the literature: ethical AI needs to be implemented structurally, not applied selectively.

Before investigating the transparency paradox, it is important for the reader to understand the framing of ethical governance as a necessary structural component of AI in supply chains. Digitally transformed sustainable innovation is ethically driven, yet while AI-driven developments create performance efficiencies, they also compel concern for fairness, transparency, and accountability as major international supply chain directives. In the Sama case study, AI outsourcing has social consequences and not all benefits of technology globalization are equitable. In a more theoretical framework, this dynamic will be linked to Institutional and Legitimacy Theory, which positions organizations for legitimacy as per transparency, reduced external pressures and social demands. At the same time, transparency is paradoxical; the more transparent one becomes about their AI production and implementation, the more monitored and accountable they are for their dubious deeds, putting them under publicized reputational scrutiny. Thus, the paradox of transparency is the challenge it creates between insight and exposure, which states that the more transparent one is technologically, the more they'll be held accountable for their actions.

To further support practitioners in navigating such paradoxes, Table 9.2 introduces an Ethical AI Implementation Matrix, which takes the four significant ethical dimensions of AI (accountability, fairness, privacy and explainability) and translates them

TABLE 9.2 Ethical AI implementation matrix

Ethical Dimension	*AI Implementation Stage*
Accountability	Planning & Design
Fairness	Data Collection
Privacy	Model Training & Deployment
Explainability	Operation & Monitoring

Source: Author's synthesis based on Institutional and Ethical-AI literature (2022–2024)

into performance requirements based on their development cycles. This makes transparency assessable as it ethically translates recognizable action based on development stages.

This integration positions ethical AI not as a technical afterthought but as a continuous alignment process between technological potential, organizational responsibility, and societal expectations.

Reputation, Risk, and the Paradox of Transparency

With the rising dependence on AI and blockchain to increase visibility and traceability in supply chains, a new reputational risk has come into being. The research by van Hoek and Richey revealed the dual role of AI in reducing and increasing reputational risks. AI can identify warning signs of ethical violations like labour abuses, unsustainable practices, or social media outcry, enabling firms to intervene proactively. But on the other hand, such equipment increase a company's liability by making the actions of tier-2 and tier-3 suppliers traceable, potentially exposing the company to legal and reputational risks. This creates a paradox: the more open and AI-based a supply chain becomes, the greater the reputational risk of the firms will be. Counterintuitive to this, the authors suggest proactive and collaborative management of reputational risk through shared responsibility, cross-tier governance, and ethical use of AI, through shared responsibility, cross-tier governance, and ethical use of AI, all of which depend on resilient interorganizational relationships.

Human-Centric AI for Empowerment and Inclusivity

Besides technical transparency and risk management, the human-driven AI emphasis includes empowering employees, decision-makers, and customers. Different articles urged the embedding of AI systems that augment instead of replacing human intelligence, particularly in the emerging economies and small and medium enterprises (SMEs). One such instance came out of a study of AI chatbot uptake in Indian SMEs, where the results showed that although AI did not significantly contribute directly to sustainable supply chain performance, it increased substantially supply chain visibility and innovation capability, both of which were mediators of performance gain. The research upheld that AI performs well when it augments existing human capabilities, as opposed to being an impenetrable, self-contained entity. Similarly, research on AI in humanitarian supply chains emphasized the imperative of ethical prioritization,

especially when dealing with populations under stress. Predictive analytics via AI for drone delivery, aid prioritization, and real-time logistics have to include not just efficiency, but also dignity, equity, and respect for culture.

Designing for Ethical AI: Governance, Standards, and Education

Experts, in all the applications, concluded that ethical AI outcomes are not going to be achieved based on technical design alone. There is a multi-level solution with design ethics, stakeholder education, policy alignment, and industry-wide norms.

Some authors offered ethical AI frameworks with the following principles:

- Algorithmic fairness and detecting bias
- Data privacy and informed consent
- Human-in-the-loop systems
- Design auditability and accountability

In this sense an AI system should be able to allow human intervention, validation, or override real-time decisions in the process. It is very critical for such industries dealing with perishable products, pharmaceutical products, or supply chains related to humanitarian goods, where estimations by an AI need calibration with the reality on ground and ground intelligence. Sectoral ethical norms, particularly for fast fashion, mining, and agriculture, are also necessary for AI application in supplier screening and sustainability reporting. Blockchain-enabled smart contracts or AI-sustainability ratings systems need to be preceded by transparent metrics and assured third-party audits to both serve performance and ethics.

Embedding AI within supply chain management is not an engineering innovation, but it is a cultural and ethical shift that needs to be taken into consideration. As this review illustrates, the ability of AI to optimize, automate, and forecast is massive. But its worth, in the long term, will be in its capability to honour human dignity, foster trust, and mirror common values within intricate, worldwide chains. Ethical, explainable, and human-centred AI is no luxury. It is the building block for AI systems that last, learn, and elevate. To accomplish this, organizations will have to commit not just to newer models but to the moral pillars that reinforce equitable AI deployment. The integration and application of AI into the business model needs to be seen not as a replacement of human talent, but as its sustainer and enhancement.

Notes

1 A novel AI-enhanced blockchain platform designed to solve long-standing challenges in IoT-based supply chains.
2 'Black box' phenomenon is a term often used to describe systems, particularly AI and machine learning programs, that produce outputs or decisions without visibly demonstrating how and why they arrived at the outcomes.

References

Ahi, P., Searcy, C., 2015. Assessing sustainability in the supply chain: A triple bottom line approach. *Applied Mathematical Modelling* 39, 2882–2896. doi:10.1016/j.apm.2014.10.055.

Aylak, B. L. (2025). SustAI-SCM: Intelligent Supply Chain Process Automation with Agentic AI for Sustainability and Cost Efficiency. *Sustainability*, 17(6), 2453. doi:10.3390/su17062453.

Badghish, S., Soomro, Y. A., 2024. Artificial intelligence adoption by SMEs to achieve sustainable business performance: application of technology–organization–environment Framework. *Sustainability* 16, 1864. doi:10.3390/su16051864.

Bag, S., Wood, L. C., Xu, L., Dhamija, P., Kayikci, Y., 2020. Big data analytics as an operational excellence approach to enhance sustainable supply chain performance. *Resources, Conservation and Recycling* 153, 104559.

Benzidia, S., Makaoui, N., Bentahar, O., 2021. The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. *Technological Forecasting and Social Change 165*, 120557. doi:10.1016/j.techfore.2020.120557.

Brintrup, A., Baryannis, G., Tiwari, A., Ratchev, S., Martinez-Arellano, G., Singh, J., 2023. Trustworthy, responsible, ethical AI in manufacturing and supply chains: synthesis and emerging research questions. doi:10.48550/arXiv.2305.11581.

Chauhan, C., Kaur, P., Arrawatia, R., Ractham, P., Dhir, A., 2022. Supply chain collaboration and sustainable development goals (SDGs). Teamwork makes achieving SDGs dream work. *Journal of Business Research 147*, 290–307. doi:10.1016/j.jbusres.2022.03.044.

Cheong, B. C., 2024. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics. 6*. doi:10.3389/fhumd.2024.1421273.

Cobbe, J., Veale, M., Singh, J., 2023. Understanding accountability in algorithmic supply chains. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. pp. 1186–1197. doi:10.1145/3593013.3594073.

Dubey, R., Gunasekaran, A., Bryde, D. J., Dwivedi, Y. K., Papadopoulos, T., 2020. Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting. *International Journal of Production Research*, 58(11), 3381–3398. doi:10.1080/00207543.2020.1722860.

Farooq, M., Yuen, Y., 2024. Artificial intelligence in supply chain management: a comprehensive review and framework for resilience and sustainability. doi:10.21203/rs.3.rs-3878218/v1.

Frimpong, V., 2025. The sustainability paradox of artificial intelligence: how AI both saves and challenges resource management efforts. doi:10.2139/ssrn.5176930.

Haddaway, N. R., Page, M. J., Pritchard, C. C., McGuinness, L. A., 2022. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews* 18, e1230. doi:10.1002/cl2.1230.

Handfield, R. B., Nichols (Jr.), E. L., 1999. *Introduction to Supply Chain Management*. Prentice Hall.

Hartmann, D., de Pereira, J. R. L., Streitbörger, C., Berendt, B., 2024. Addressing the regulatory gap: moving towards an EU AI audit ecosystem beyond the AI Act by including civil society. *AI Ethics*. doi:10.1007/s43681-024-00595-3.

Hernandez, A., Caballero, A., Balmes, I., Niguidula, J., 2023. Artificial intelligence for sustainability: evidence from select small and medium enterprises in the Philippines. doi:10.1109/ICBIR57571.2023.10147579.

Jauhar, S.K., Harinath, S., Krishnaswamy, V. et al., 2025. Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics. *Annals of Operations Research* 354, 103–142. doi:10.1007/s10479-024-06348-z.

Jum'a, L., Zimon, D., Sroufe, R., Tyan, J., 2024. Sustainable supply chain management's impact on triple bottom line performance: does the firm size matter? *Corporate Social Responsibility and Environmental Managemen*t 31, 4673–4693. doi:10.1002/csr.2826.

Kanyepe, J., Chibaro, M., Morima, M., Moeti-Lysson, J., 2025. AI-powered agricultural supply chains: applications, challenges, and opportunities. pp. 33–64. doi:10.4018/979-8-3693-6468-0.ch002.

Kassa, A., Kitaw, D., Stache, U., Beshah, B., Degefu, G., 2023. Artificial intelligence techniques for enhancing supply chain resilience: a systematic literature review, holistic framework, and future research. *Computers & Industrial Engineering 186*, 109714. doi:10.1016/j.cie.2023.109714.

Kurniawati, D., & Cakravastia, A. (2023). Sustainability performance and transparency in supply chains: A review of challenges and opportunities. *Journal of Cleaner Production*, 382, 135257. doi:10.1016/j.jclepro.2022.135257.

Lara, C.L., Wassick, J., 2023. Future of supply chain: challenges, trends, and prospects. doi:10.48550/arXiv.2301.13174.

Lopes de Sousa Jabbour, A.B., Laguir, I., Stekelorum, R., Gupta, S., 2025. The nexus of artificial intelligence and sustainability performance: Unveiling the impact of supply chain transparency and customer pressure on ethical conduct. *Journal of Environmental Management 379*, 124847. doi:10.1016/j.jenvman.2025.124847.

Meline, T., 2006. Selecting studies for systemic review: inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders* 33, 21–27. doi:10.1044/cicsd_33_S_21.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med 6*, e1000097. doi:10.1371/journal.pmed.1000097.

Naz, F., Agrawal, R., Kumar, A., Gunasekaran, A., Majumdar, A., Luthra, S., 2022. Reviewing the applications of artificial intelligence in sustainable supply chains: exploring research propositions for future directions. *Business Strategy and the Environment 31*, 2400–2423. doi:10.1002/bse.3034.

Nogueira, E., Gomes, S., Lopes, J. M., 2025. Unveiling triple bottom line's influence on business performance. *Discover Sustainability 6*, 43. doi:10.1007/s43621-025-00804-x.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews* 10, 89. doi:10.1186/s13643-021-01626-4.

Pennekamp, J., Matzutt, R., Klinkmüller, C., Bader, L., Serror, M., Wagner, E., Malik, S., Spiß, M., Rahn, J., Gürpinar, T., Vlad, E., Leemans, S. J. J., Kanhere, S. S., Stich, V., Wehrle, K., 2023. An interdisciplinary survey on information flows in supply chains. *ACM Computing Surveys* 56(2), 1–38. doi:10.1145/3606693.

Qu, C., Kim, E., 2024. Reviewing the roles of AI-integrated technologies in sustainable supply chain management: research propositions and a framework for future directions. *Sustainability* 16, 6186. doi:10.3390/su16146186.

Raman, R., Sreenivasan, A., Ma, S., Patwardhan, A., Nedungadi, P., 2023. Green supply chain management research trends and linkages to UN Sustainable Development Goals. *Sustainability* 15, 15848. doi:10.3390/su152215848.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), 2024.

Riad, M., Naimi, M., & Okar, C. (2024). Enhancing Supply Chain Resilience through Artificial Intelligence: Developing a Comprehensive Conceptual Framework for AI Implementation and Supply Chain Optimization. Logistics, 8(4), 111. doi:10.3390/logistics8040111.

Riahi, Y., Saikouk, T., Gunasekaran, A., Badraoui, I., 2021. Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. *Expert systems with applications* 173, 114702.

Samuels, A., 2025. Examining the integration of artificial intelligence in supply chain management from Industry 4.0 to 6.0: a systematic literature review. *Frontiers in Artificial Intelligence*, 7, 1477044. doi:10.3389/frai.2024.1477044.

Sanders, N., Boone, T., Ganeshan, R., Wood, J., 2019. Sustainable supply chains in the age of AI and digitization: research challenges and opportunities. *Journal of Business Logistics* 40. doi:10.1111/jbl.12224.

Seuring, S., Müller, M., 2008. From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production, Sustainability and Supply Chain Management* 16, 1699–1710. doi:10.1016/j.jclepro.2008.04.020.

Sodhi, M. S., Tang, C. S., 2021. Supply chain management for extreme conditions: research opportunities. *Journal of Supply Chain Management* 57, 7–16.

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., Fischl, M., 2021. Artificial intelligence in supply chain management: a systematic literature review. *Journal of Business Research* 122, 502–517. doi:10.1016/j.jbusres.2020.09.009.

Ullah, S., Nobanee, H., Iftikhar, H., 2023. Global financial integration, governance-by-technology, and green growth. *International Review of Financial Analysis* 90, 102838. doi:10.1016/j.irfa.2023.102838.

United Nations Global Compact, 2023. Supply Chains Key to Unlocking Net Zero Emissions | Accenture. United Nations Global Compact.

van Wynsberghe, A., 2021. Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* 1, 213–218. doi:10.1007/s43681-021-00043-6.

Vann Yaroson, E., Abadie, A., Roux, M., 2025. Human–artificial intelligence collaboration in supply chain outcomes: the mediating role of responsible artificial intelligence. *Annals of Operations Research*. doi:10.1007/s10479-025-06534-7.

Wright, D., Igel, C., Samuel, G., Selvan, R., 2025. Efficiency is not enough: a critical perspective of environmentally sustainable AI. doi:10.1145/3724500.

Yadav, R., Kumar, D., & Singh, S. (2024). Consumer awareness and behavioral responses to sustainable practices: The role of communication and transparency. *Sustainability*, 16(3), 1125. doi:10.3390/su16031125.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART IV

The Human Dimension of AI Transformation: Managing Change, Ethics, and Trust



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

10

EFFECTIVE COMMUNICATION AND AI VALUE FOR POTENTIAL CLIENTS

Silvia Elena Iacob and Nicolae Moroianu

In today's fast-evolving business world, artificial intelligence (AI) has quietly become the invisible hand shaping the way companies interact with their clients, turning ordinary exchanges into intelligent, data-driven experiences that feel increasingly seamless and personalised. Imagine a customer reaching out with a question: the response is immediate, guided by chatbots or virtual assistants, with predictive analytics anticipating needs and offering solutions almost before the question is fully asked. Behind this efficiency, however, lies a challenge no algorithm can yet overcome – the subtle warmth of human empathy, the trust built through understanding, and the nuanced interactions that sustain long-term relationships.

This chapter tells the story of AI as more than a mere tool for automation. It explores how communication specialists can assess AI applications responsibly, ensuring that these technologies enhance customer engagement without replacing the irreplaceable human touch. By investigating AI-driven tools across service delivery, marketing, and customer interaction, the chapter reveals the ways in which these technologies optimise responses, personalise experiences, and streamline processes, while still depending on human oversight to nurture trust and satisfaction.

Through an exploratory review of scientific literature, international reports, and institutional documents from Europe and the United States, the narrative illustrates how AI contributes to understanding customers' perspectives, demonstrating value through tangible outcomes, simplifying technical complexity, and continuously self-improving, as depicted in Figure 10.1.

Hence, the chapter puts emphasis on the fact that true value of AI emerges in hybrid communication models, where efficiency meets empathy, allowing businesses to enhance client satisfaction and loyalty while maintaining authentic human connection. The potential AI value for its clients include understanding customers' perspective and providing value demonstration through outcomes, while simplifying technical complexity and self-improving continuously, as Figure 10.1 clearly illustrates.

Looking forward, AI holds the potential to develop adaptive emotional intelligence, recognise individual voices, retain prior interactions, and deliver personalised solutions

DOI: 10.4324/9781003632429-15

This chapter has been made available under a CC-BY-NC-ND 4.0 license

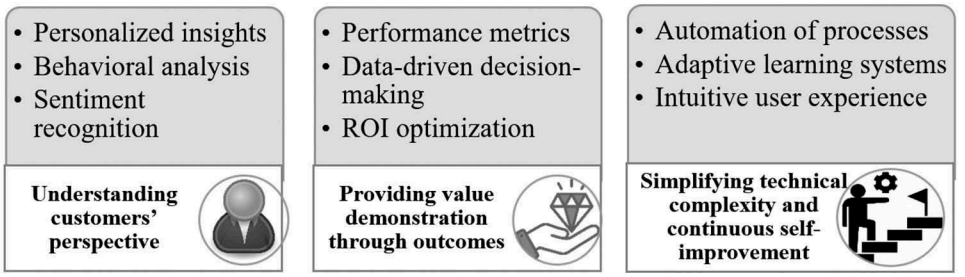


FIGURE 10.1 AI value for potential clients
Source: Authors' contribution

that eliminate repetition and friction in customer journeys. In this way, the story of AI in corporate communication becomes one of partnership, where machines augment human capabilities, and together they create a richer, more responsive, and sustainably engaging experience for clients.

Overall, integrating AI technologies with human communication strategies has become essential for businesses that aim to enhance client value. Due to humans' particular quality of empathy-driven interactions, businesses should implement hybrid communications models maintaining human involvement to balance automation and emotional intelligence. Hence, our chapter contributes to the literature by examining the potential value of AI mass-adoption within the corporate sector but also by revealing the numerous strategies involved in reaching AI readiness from a multifaceted perspective that allows potential clients to gain a sustainable growth within the AI landscape. Future research directions could explore the evolving emotional intelligence capabilities of artificial intelligence systems and the development of real-time adaptive learning models designed to refine, enhance, and personalise client interactions. Such systems possess the potential to retain the entirety of previous communications, recognise individual voices, and thus eliminate the need for repeated identification or for clients to reiterate their issues to multiple operators.

1 Introduction: AI and the Reconfiguration of Client-Oriented Communication

In the current climate of accelerated digital transformation, artificial intelligence (AI) has become a defining force in reshaping organisational communication practices. Thus, businesses are increasingly integrating AI into their operations not merely to automate routine processes, but to strategically enhance their engagement with clients. Meanwhile, due to the ability of AI systems to analyse vast volumes of data, detect patterns in customer behaviour, and provide real-time tailored responses, the nature of client interaction has greatly been transformed. Hence, this shift has signalled a profound evolution in communication dynamics with customer-centric AI applications becoming central to value creation and service personalisation. Additionally, the deployment of AI in communication, particularly in business-to-consumer (B2C) contexts, reflects a growing reliance on algorithmic systems to mediate, facilitate and

sometimes even lead human interactions. For instance, tools such as chatbots, virtual assistants, and predictive analytics engines now function as intermediaries in client journeys, while shaping everything from initial contact to post-sale follow-up. Likewise, these systems are celebrated for their capacity to increase efficiency, improve consistency, and lower operational costs. Yet, the benefits of AI-driven communication are not limitless in spite of their strategic ability to streamline transactional aspects of communication through automation; in fact, AI applications frequently lack the emotional intelligence, empathy, and contextual sensitivity that are essential in building trust-based enduring customer relationships.

Not so long ago, organisational communication was still largely shaped by familiar human rhythms: conversations unfolded at a manageable pace, relationships grew through repeated contact, and trust was built gradually, often through small, seemingly unremarkable exchanges that accumulated meaning over time. Today, however, that landscape is being quietly but decisively redrawn by the rapid advance of AI, which has slipped into the everyday workings of businesses not as a futuristic novelty, but as a powerful, almost invisible presence that increasingly mediates how organisations speak to, listen to, and ultimately understand their clients. What began as a practical solution for handling routine tasks has evolved into a strategic instrument for engaging customers more efficiently, more consistently, and on a far greater scale than was ever previously possible, largely because AI systems can sift through vast quantities of data, recognise subtle behavioural patterns, and respond in real time with messages that feel surprisingly tailored to individual needs and preferences.

As a result, the very nature of client interaction has been transformed, with communication no longer unfolding solely between people, but often passing through algorithmic intermediaries that shape the journey from first contact to long after the sale has been completed. Chatbots answer questions at all hours of the day and night, virtual assistants guide users through complex choices, and predictive systems quietly anticipate problems before they are even voiced, all of which has been widely praised for improving efficiency, ensuring consistency, and keeping costs under control. Yet beneath this smooth and highly optimised surface lies a more uncomfortable realisation: however swift and sophisticated these systems may be, they frequently struggle with the emotional nuance, situational awareness, and genuine empathy that underpin trust-based, long-term relationships, leaving open the question of whether streamlined communication always translates into meaningful connection.

Much of what we currently know about AI in communication reflects this tension. Researchers have devoted considerable attention to the technical brilliance of automated services and to the ways in which AI supports digital marketing and customer engagement, often drawing on well-established ideas about why people accept or reject new technologies, such as their perceived usefulness, ease of use, and overall satisfaction. These perspectives have undoubtedly deepened our understanding of adoption and efficiency, yet they have also tended to view value through the lens of organisational performance, quietly sidelining the client's lived experience. In focusing so heavily on internal metrics and operational gains, existing discussions have too often overlooked a more fundamental question: whether AI genuinely enriches communication from the client's point of view, or merely accelerates interactions without deepening them.

Against this backdrop, the present chapter sets out to shift the spotlight away from automation for its own sake and towards the conditions under which AI can meaningfully contribute to client value, particularly when its use is guided by thoughtful human oversight rather than blind technological enthusiasm. The argument unfolds around the idea that AI becomes truly effective not when it replaces human communicators, but when it works alongside them, amplifying their capacity to be responsive, attentive, and emotionally attuned. Seen in this light, hybrid communication, in which machines handle scale and speed while humans safeguard authenticity and relational depth, emerges not as a compromise, but as a strategic necessity for organisations seeking sustainable relationships rather than fleeting transactions.

To explore this terrain, the chapter draws on a wide-ranging body of recent research, institutional analyses, and policy reports, weaving these strands together into an integrated perspective on how AI reshapes client-facing communication. Rather than offering a purely technical account, it traces how different AI applications are woven into everyday contact points, from initial outreach and service delivery to ongoing relationship management and feedback loops, revealing how personalisation, immediacy, and predictive insight can reshape the client journey into something far more fluid and continuously evolving. At the same time, it confronts the limits of these systems, acknowledging their difficulty in grasping context, expressing genuine concern, or responding sensitively to emotionally charged situations, and showing why human presence remains indispensable precisely where communication matters most.

As AI becomes more deeply embedded in organisational life, further questions inevitably surface, particularly around responsibility and ethics. When automated systems make mistakes, offer misleading information, or reinforce hidden biases, accountability becomes blurred, prompting difficult reflections on transparency, fairness, and trust. Addressing these concerns requires more than technical fixes; it calls for organisations to cultivate digital fluency, embed responsible innovation into their culture, and align AI deployment with broader communicative values rather than short-term efficiency gains. Evidence from early adopters suggests that those who approach AI in this way not only streamline operations, but also redefine what meaningful engagement looks like in a digital economy increasingly shaped by algorithmic mediation.

Ultimately, this chapter looks ahead to a future in which the most successful uses of AI in communication are likely to be those that treat technology not as a replacement for human interaction, but as a relational collaborator that supports clearer understanding, reduces cognitive effort, and enables more adaptive, responsive exchanges. The central message is a simple yet demanding one: integration, not substitution, offers the most promising path forward. Only by embedding AI within ethically grounded, emotionally aware, and strategically coherent communication practices can organisations fully harness its potential while preserving the profoundly human need for connection, trust, and mutual understanding in an increasingly digital world.

2 Theoretical Framework: From Communication Theories to AI and the Construction of Client Value in the Digital Age

In the context of digitally transformed business environments, communication emerges as both an infrastructural foundation and a strategic enabler of value creation. Since it is no longer confined to linear, transactional exchanges, effective communication is increasingly defined by its responsiveness, contextual relevance, multimodality, and emotional resonance. Thus, within such digitally mediated ecosystems, the convergence of technological intelligence and relational intent shapes the communicative paradigms through which organisations engage clients, foster trust, and cultivate enduring loyalty [1–3]. At the heart of this transformation lies the integration of AI technologies, which have fundamentally reconfigured the communicative landscape through applications such as natural language processing (NLP), machine learning (ML), predictive analytics, and chatbot interfaces. These technologies operate not only as operational tools but as semi-autonomous communicative agents, capable of parsing user sentiment, anticipating preferences, and adapting message delivery in line with inferred behavioural patterns. Consequently, AI enables real-time responsiveness, automates information flows, and facilitates personalisation at scale, thus reshaping the architecture of interaction, and transforming communication from a reactive process into a proactive, anticipatory system of engagement.

To formally conceptualise these dynamics, several theoretical frameworks provide analytical grounding (Figure 10.2). First, the Technology Acceptance Model (TAM) [4] remains a foundational lens for understanding user engagement with AI interfaces, specifically because it posits that perceived usefulness and ease of use are primary determinants of technology adoption, principles that continue to inform the design and deployment of AI-enabled communicative systems. Hence, when clients perceive AI interactions as both intuitive and value-enhancing, their receptiveness to algorithmic mediation increases, thereby reinforcing engagement continuity. Complementing TAM, the Customer Experience Theory (CET) offers a more detailed comprehension of how communication contributes to perceived client value and emphasises the emotional, cognitive, and sensorial dimensions of the client journey, recognising each interaction as a potential satisfaction touchpoint. Since within AI-mediated environments these touchpoints are both digitised and increasingly curated, being shaped by algorithms that seek to optimise the affective resonance of each communicative moment, therefore the challenge is to ensure informational accuracy, but also to preserve emotional authenticity within synthetic interactions. In assessing the quality of such digitally mediated services, the E-S-QUAL model [5] provides a targeted evaluative framework centred of dimensions such as efficiency, system availability, privacy, and fulfilment as key indicators of digital service performance (Figure 10.2). Originally developed to measure service quality in online environments, once applied to communicative AI, these dimensions put emphasis on the necessity of designing systems that are not only fast and available, but also transparent, secure, and aligned with user expectations. When AI meets these criteria, the likeliness to be perceived as trustworthy and competent increases, thus reinforcing relational value.

Additionally, the concept of client value itself must be problematised to fully grasp AI's communicative implications, starting from the traditional perspective where value

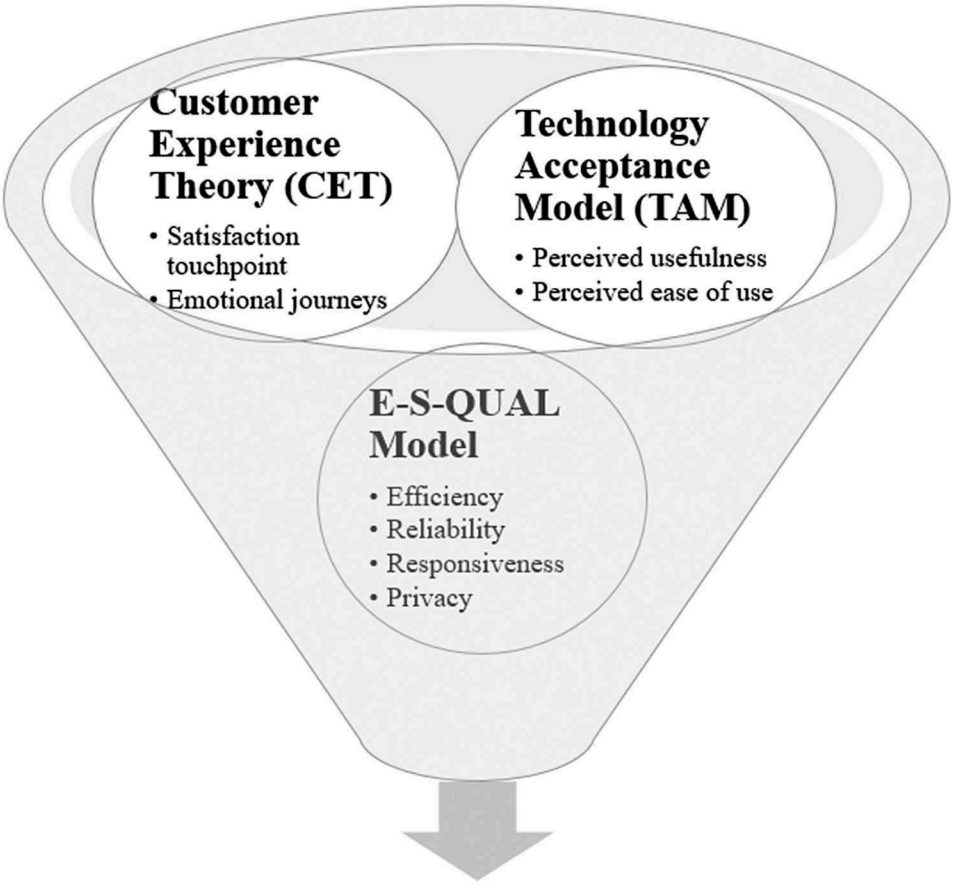


FIGURE 10.2 Interconnected theoretical foundations driving AI-generated client value
Source: Authors' contribution

has been conceived through a transactional lens, primarily centred on the exchange of goods and services. However, in the context of AI-mediated communication, value assumes a relational character and includes emotional, functional, and symbolic dimensions. For instance, emotional value pertains to feelings of trust, satisfaction, or appreciation generated during the interaction; functional value related to the efficiency, relevance, and cognitive clarity of the exchange; while symbolic value is embedded in the client's perception of the brand or institution as technologically sophisticated, forward-looking, or client-centric. Essentially, communication acts as the vector through which these value dimensions are transmitted, amplified, or diminished. For example, an AI system that efficiently delivers personalised responses contributes to functional value, but it fails to recognise distress signals or misinterprets cultural nuance, it may compromise emotional value. Likewise, the very presence of advanced AI may confer symbolic prestige, positioning the organisation as innovative, yet also invites heightened scrutiny regarding ethical transparency and user agency.

Can an entity whose chatbot recommends that a customer not choose the colour black be accused of racism or xenophobia?

Consequently, this theoretical synthesis sets the foundation for subsequent sections which will operationalise these concepts through empirical evidence and applied frameworks, in order to advance a multi-dimensional comprehension of how AI reshapes communicative value and redefines the communicative conditions under which value is perceived, negotiated, and co-constructed.

3 Taxonomy of AI Applications in Customer Communication

The proliferation of AI within the domain of customer communication has led to a profound recalibration of the ways organisations engage with their clients across digital platforms. For this reason, this section proposes a systematic examination of the typologies, functionalities, and limitations of AI-based tools currently deployed to enhance corporate communication strategies.

On one hand, AI communication systems can be broadly categorised into three functional domains: conversational agents, automated marketing infrastructures, and behavioural analytics engines. Firstly, conversational agents encompass chatbots and virtual assistants such as Zendesk's Answer Bot, Amazon's Alexa, and OpenAI's ChatGPT and simulate human dialogue through natural language processing (NLP) and machine learning (ML), facilitating real-time engagement and query resolution. All these AI applications have demonstrated substantial efficacy in addressing routine queries, reducing waiting times, and supporting 24/7 availability, thereby improving perceived service responsiveness [6].

There is an increasing need to develop a specific and standardised language for communication with AI, allowing systems to exchange consistent and interpretable messages between users and AI interfaces – similar to the way in which vehicles communicate through signalling, not for their own awareness but for the benefit of surrounding drivers.

Secondly, automated marketing infrastructures integrate AI-driven systems to personalise email marketing campaigns, segment customer profiles, and optimise content timing through predictive modelling. Then, real-time behavioural analytics further enhance customer communication by continuously tracking user interactions, purchasing patterns, and emotional valence through sentiment analysis. For example, the Sephora Virtual Artist employs computer vision and ML to customise product recommendations based on facial recognition and individual beauty preferences, while ING Bank's INGE virtual assistant illustrates the potential of AI to deliver multilingual, round-the-clock banking support through contextual learning and automated decision-making protocols.

Moreover, a key contribution of AI in this landscape is its role in advancing conversational marketing, a domain where the focus shifts from transactional broadcasting to dialogue-based, context-aware engagement. A particular interesting feature of AI tools is their capability of modulating tone, adapting responses, and anticipating user intent, thereby reducing friction within the customer journey and enhancing the perceived personalisation of the brand [7]. Hence, data-informed communication practices leverage vast datasets to identify explicit preferences but also latent

behavioural patterns, which allows organisations to engage clients proactively, predict future purchases, and prevent churn.

On the other hand, despite these advances, current AI tools exhibit significant constraints such as the overreliance on rule-based architectures that fail to accommodate diverse emotional expression or contextual fluidity, leading to interactions that are efficient but impersonal. For instance, even though machine learning enables continuous improvement, most systems still operate within narrow domains and lack the ability to generalise knowledge across communicative contexts [8]. To elucidate these developments, Table 10.1 presents a taxonomy of AI applications in customer communication with illustrative cases and theoretical implications.

Taking into consideration the evolving nature of AI tools' technological substrate, the communicative affordances embedded within customer interaction systems are also undergoing rapid transformation. Specifically, the shift from unimodal text-based interfaces toward multimodal context-aware agents represents a paradigmatic advancement in digital communication systems, one which enables a more naturalistic and emotionally resonant dialogue between brands and customers, as newer models demonstrate a marked improvement in long-form coherence, problem-solving capabilities, and user-specific contextualisation. Thus, these innovations reconfigure the boundaries of AI-mediated communication, permitting firms to deploy increasingly sophisticated and scalable customer-facing tools that align with relational marketing logics. To better illustrate this accelerated technological evolution, Table 10.2 presents a comparative overview of general-purpose generative AI models developed by leading frontier labs between 2022 and early 2025. Additionally, Table 10.2 captures the transition from limited text-only models with constrained contextual processing to multimodal systems capable of advanced reasoning and real-time data integration. For example, tools such as OpenAI's GPT-4o and Google's Gemini 2.0 Flash are capable of interpreting and generating text, audio, and images, but

TABLE 10.1 Taxonomy of AI applications

AI Category	*Functionality*	*Representative Tools*	*Theoretical Implications*
1. Conversational Agents	Simulate human interaction via NLP; provide real-time assistance	ChatGPT, Alexa, Zendesk Bot; ING's INGE	Enhances service responsiveness; supports TAM's "perceived ease of use"
2. Automated Marketing Systems	Personalise content; automate campaign delivery; segment customer data	Mailchimp AI, Salesforce Einstein	Reinforces CET's emotional journey via targeted touchpoints
3. Behavioural Analytics	Monitor user behaviour in real time; predict preferences and purchases	Google Analytics, Adobe Sensei	Supports E-S-QUAL by enabling responsiveness and personalisation
4. Hybrid Interfaces	Combine multiple AI capabilities into integrated communication ecosystems	Sephora Virtual Artist; AI-powered CRM dashboards	Facilitates relational value creation through symbolic brand association

Source: Authors' contribution

TABLE 10.2 Communicative functions dynamics of AI tools from 2022 to 2025

AI tool	*2022–23*	*January 2025*
Anthropic	**Claude** ⟶	**Claude 3.5**
	• Not multimodal (text only) • Limited contextual understanding (difficulty with complex conversations) • No tool usage	• Multimodal (text, audio, and images) • Enhanced contextual understanding and coherence during long interactions • Experimental computer usage capability for some users
Google Gemini	**Google Bard** ⟶	**Gemini 2.0 Flash**
	• Not multimodal (text only) • Fair reasoning • Limited contextual understanding (difficulty with complex conversations) • Limited real-time data integration • Low personalisation (limited adaptability)	• Multimodal (text, audio, and images) • Advanced reasoning (capable of multistep problem-solving and nuanced analysis) • Enhanced contextual understanding (maintains coherence in long dialogues) • Real-time data integration (from Google Search) • Advanced personalisation (user context)
Meta	**Llama 1** ⟶	**Llama 3.3**
	• Not multimodal (text only) • Fair reasoning • Limited contextual understanding (difficulty with complex conversations) • No API access	• Text-based (earlier versions were multimodal, Llama 3.2) • Advanced reasoning (capable of multistep problem-solving and nuanced analysis) • Enhanced contextual understanding (maintains coherence in long dialogues) • API access (tools for model and agent development)
Microsoft	**Phi-1** ⟶	**Phi-4**
	• Not multimodal (text only) • Fair reasoning • Focused training (smaller, coding-focused data set)	• Multimodal (text, audio, and images) • Advanced reasoning (capable of multistep problem-solving and nuanced analysis) • Comprehensive training (diverse data)

AI tool	*2022–23*	*January 2025*
OpenAI	**GPT-3.5** ⟶	**OpenAI o1**
	• Not multimodal (text only) • Fair reasoning ability (e.g. scored high on SAT, but bottom 10% on bar examination) • Limited contextual understanding (difficulty with complex conversations) • Standard API access (for text generation)	• Multimodal (text, audio, and images) • Advanced reasoning (e.g. top 10% on bar examination) • Enhanced contextual understanding (maintains coherence in long dialogues) • Advanced API access (supports multimodal inputs)

Source: Authors' contribution based on McKinsey [9]

also exhibit enhanced personalisation through dynamic adaptation to user history and intent. Hence, this advancement directly contributes to a new communicative paradigm that supports emotional nuance, predictive dialogue, and culturally sensitive response generation across geographies and industries.

As data from Table 10.2 shows, the evolution of AI-enabled communicative tools between 2022 and 2025 reveals a discernible transition from static, rule-based interaction models to increasingly multimodal and context-aware systems. Moreover, a comparative reading of the data presented in Table 10.2 indicates that most platforms have expanded their capabilities beyond text-only interfaces to include visual and auditory processing. Thus, this transition from unimodality to multimodality signifies more than a technical augmentation [10] and reflects a paradigmatic redefinition of human–machine interaction, whereby artificial communicators begin to approximate the perceptual multiplicity characteristic of human cognition. In communicative terms, this ability enables richer meaning construction through the convergence of linguistic and non-linguistic semiotic resources.

Parallel to this development, the evolution of contextual reasoning and discourse coherence emerge as a strategic axis of innovation from the 2022 overview, when systems like GPT-3.5, Bard, or Llama 1 were limited in their ability to sustain long-form interactions and failed to preserve thematic continuity or to respond adaptively to user intent. In contrast, the 2025 iterations exhibit markedly improved abilities in multistep reasoning and pragmatic coherence, indicating a move away from shallow, stimulus–response architectures toward models capable of deploying contextually contingent communicative strategies. More specifically, from the standpoint of dialogical theory such tools increasingly fulfil the criteria of interlocutors rather than mere information dispensers, thereby recalibrating their function within the communicative ecology of organisations (Table 10.2). In addition, this development aligns with Grice's [11] cooperative principle, notably in how AI models increasingly conform to conversational maxims (relevance, quantity, and manner), thereby enhancing their pragmatic legitimacy as interlocutors, being capable of constructing dialogue in alignment with implicit communicative norms.

A further significant development concerns the extent to which AI tools incorporate mechanism for real-time personalisation and data-informed adaptability. First, the

integration of user context, particularly evident in Google's Gemini 2.0 and Anthropic's Claude 3.5, permits dynamic responsiveness that was previously unattainable in general-purpose language models. In particular, Gemini 2.0's context-aware architecture and Claude 3.5's experimental computing functionality suggest a move toward what Luhmann [12] conceptualised as self-referential systems, defined as entities capable of observing their own operations and modulating their communicative outputs accordingly. Secondly, the integration of live data streams and enhanced user modelling allows these systems to recalibrate their output in relation to emergent communicative cues, marking the emergence of a new communicative regime, wherein interaction is co-constructed across temporal, informational, and experiential vectors. Additionally, these AI tools capabilities reshape the epistemological assumptions supporting AI-mediated communication, foregrounding fluidity, subjectivity, and adaptation (Table 10.2).

Finally, the expansion of application programming interface (API) access and tool orchestration capacities across platforms represents a structural deepening of AI's role in the communicative apparatus of contemporary organisations. In contrast to their earlier versions (Table 10.2), the 2025 models can now perform external task execution, engage in multimodal input processing, and support integration with other digital agents or systems. Consequently, this endows them with the capacity to function not only as isolated conversational partners but as nodal communicators embedded in broader digital infrastructures, emphasising the considerable communicative implications of AI systems which now act as distributed agents capable of mediating complex flows of information and meaning across institutional, temporal, and technological boundaries.

To sum up, the integration of frontier generative AI capabilities has enabled customer communication platforms to evolve beyond rule-based automation and adopt dialogic, adaptive interaction models. As a result, these systems increasingly meet the definitional thresholds of intelligent agents, characterised by their capacity for iterative learning and the dynamic adjustment of communicative strategies in response to evolving contextual cues. Consequently, the implications for customer experience design are substantial, as AI is no longer confined to its role as a tool for operational optimisation, but is now positioned as an active co-creator in the construction of brand-customer meaning and relational value. As the chapter progresses, subsequent sections will explore how these AI capabilities translate into value co-creation mechanisms and ethical considerations, setting the foundation for a critical understanding of AI's evolving communicative role.

4 AI Value Creation: From Automation to Strategic Engagement

In contemporary digital ecosystems, AI has transcended its original role as a facilitator of automation to emerge as a strategic instrument of value creation across organisational communication functions. At a fundamental level, AI systems reduce operational inefficiencies by minimising response time, enhancing availability, and streamlining standardised processes such as customer support and client onboarding [13]. Besides these tools' capability to enable 24/7 assistance, AI-enabled applications reinforce message consistency across communication touchpoints, which represents a

decisive factor in maintaining brand identity and consumer trust in omnichannel environments.

Additionally, beyond efficiency gains, AI-driven systems serve as architects of personalisation and foster tailored communication pathways that resonate with individual preferences and behaviours. Due to their incorporated data-driven feedback loops, these technologies can curate hyper-personalised content, ranging from dynamic webpage recommendations to adaptive email marketing sequences, thus reframing personalisation as a strategic rather than purely operational tool [6]. Likewise, AI systems are capable of interpreting large-scale behavioural patterns, thereby refining its targeting capabilities iteratively, enhancing both message relevance and the probability of conversion. This kind of mechanisms exemplify the fusion of machine learning and strategic communication planning wherein content delivery is not only reactive but continuously optimised through recursive learning.

Moreover, value creation also stems from AI's predictive capacities whereby anticipatory algorithms forecast consumer intent before explicit expression. In this vein, platforms such as Netflix and Amazon operationalise this capability by offering content and product suggestions aligned with inferred needs, thereby shaping demand and deepening user engagement [14]. Therefore, this form of communicative pre-emption indicates how AI tools can restructure the customer journey through anticipating informational or transactional requirements, thus reducing friction and increasing perceived brand responsiveness. Another distinguished feature of advanced AI systems lies in their capacity for self-improvement, since each interaction enables the system to enhance its semantic precision, contextual coherence, and affective calibration thereby generating communicative value not only through output but through cognitive evolution [15], and marking down the transition from static algorithmic instruction to dynamic learning architecture.

Altogether, these developments put emphasis on a paradigmatic transition from AI as a technical enabler of automation to AI as a strategic actor in design, delivery, and refinement of communication practices, one that should represent a focal point for organisations aiming to effectively harness these evolving capabilities in order to gain both in operational efficiency, and in relational capital and long-term stakeholder engagement. In order to visualise this transition, the following figure provides a conceptual schema mapping the AI-enabled value creation chain, from basic automation functionalities to complex strategic engagement.

As Figure 10.3 indicates, each phase of this continuum is associated with distinct communicative functions, system capabilities, and value outcomes, illustrating how AI tools progressively move from executing predefined tasks to co-producing personalised, emotionally resonant interactions. This schematic representation is informed both by theoretical developments in the fields of human–computer interaction and strategic communication [7; 8], and by empirical insights extracted from contemporary business applications, providing a foundation for further discussion on how AI technologies are reconfiguring the communicative environment of digitally mature organisations.

It must be emphasised that Figure 10.3 illustrates the gradual transition from traditional communication methods to an AI-driven communicative environment, thus highlighting the numerous ways AI reshapes interactions across various platforms. In

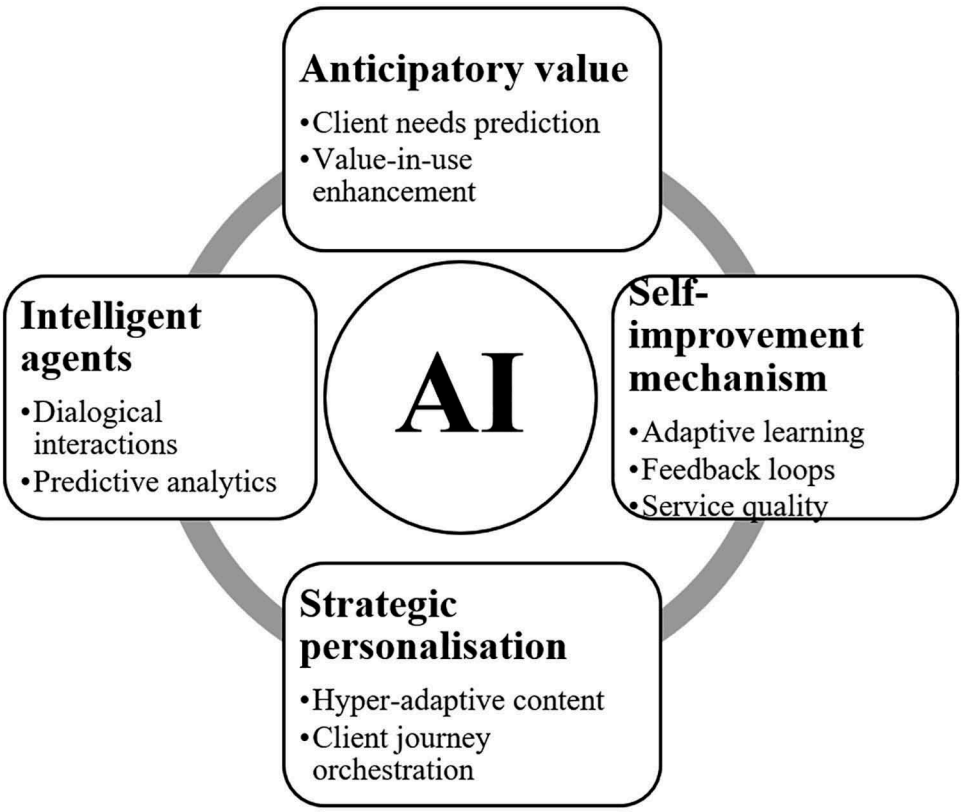


FIGURE 10.3 Transition to an AI-driven communicative environment
Source: Authors' contribution

fact, this transition marks a shift from manual, human-centred communication processes toward more automated, data-driven exchanges, enabling faster, more personalised and context-aware messaging. In addition, Figure 10.3 emphasises key technological functions such as anticipatory value, self-improvement mechanisms, strategic personalisation, and intelligent agents, which collectively empower AI systems to interpret, generate, and optimise communication in real time.

Last but not least, as organisations and individuals increasingly adopt AI communication tools, the environment evolves to integrate diverse data sources, including social media feeds, customer feedback, and real-time analytics, resulting in enhanced decision-making and more effective audience targeting for organisations. Accordingly, Figure 10.3 also captures the role of AI in enabling predictive communication strategies, where algorithms anticipate user needs and preferences aiming to deliver relevant content proactively. The result is a dynamic ecosystem where human and machine collaboration drives efficiency, accuracy, and engagement at unprecedented levels. On the other hand, the transition to an AI-driven communicative environment also presents challenges such as ensuring data privacy, maintaining ethical standards, and addressing the potential loss of human empathy in digital interactions. For those

reasons, Figure 10.3 implicitly suggests the need for balanced approaches that combine AI capabilities with human oversight to deliver relevant content proactively.

5 Human Oversight and the Limits of AI in Communication

As the saying goes, not everything that glitters is gold, and this holds true for AI-driven communication when stripped of human oversight. Particularly, despite its capacity to process information at remarkable speed and scale, AI remains inherently devoid of emotional intelligence, lacking the capacity to interpret, respond to, and evoke human feelings, mainly because the subtleties of empathy, remorse, or reassurance cannot be distilled into algorithms. Studies in therapeutic domains affirm that AI may replicate certain patterns of empathy, but users perceive human-authored emotional responses as significantly more supportive and authentic, even when the AI content is identical [16]. Therefore, the absence of these affective dimensions in AI-only interactions reduces communication to a transactional exercise, bereft of the relational depth that characterises human discourse.

That being said, the pitfalls extend beyond mere emotional absence, whereby the misinterpretation of intent and failure to respond appropriately in crises loom large as tangible risks. Since AI relies on pattern recognition, it falters when confronted with ambiguity or nuance, rendering it ill-equipped to navigate the unpredictable terrain of human sentiment. For instance, in moments demanding sensitivity, such as during brand crises or customer dissatisfaction, automatised responses risk exacerbating alienation rather than mitigating it. Evidence from business communications shows empathic chatbots enhance customer satisfaction only in certain contexts, while under time constraints, such sophistication can backfire and degrade customer experience [17]. Moreover, virtual communication environments lack paralinguistic cues, leading to frequent misreading of emotional intent and unintended alienation. Hence, these deficiencies jeopardise immediate relationships but also threatens long-term trust, which remains the bedrock of effective communication.

Nevertheless, the answer does not lie in the wholesale rejection of AI but in a synthesis of machine precision and human judgement. For example, hybrid communication models, which couple the analytical rigor of AI with the ethical discernment and cultural attunement of professionals, offer a path forward. In such frameworks, human communicators serve as stewards of nuance, calibrating messages with an appreciation for context, tone, and morality, qualities beyond AI's reach. Thus, this collaboration ensures that efficiency need not come at the expense of authenticity, creating a communicative ecosystem that respects both scale and sensitivity.

Moreover, it is worth recalling that one of the earliest and most influential communication models, Shannon and Weaver's Mathematical Theory of Communication [18], was designed to improve technical transmission by reducing noise and increasing the fidelity of information transfer between sender and receiver. On the other hand, their model was intentionally mechanistic and treated communication as a linear signal-based process, while it did not and could not account for interpretive variables such as empathy, cultural framing, or ethical responsiveness (Figure 10.4).

As communication theorists from Jakobson [19] to Habermas [20] later argued, human meaning-making requires more than accurate transmission, mainly because it

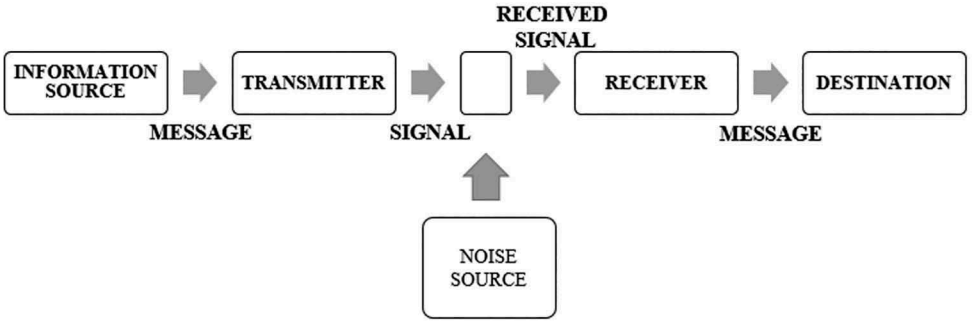


FIGURE 10.4 The classic communication model by Shannon and Weaver (1949)
Source: Authors' contribution based on Shannon and Weaver [18]

depends significantly on context, shared understanding, and the social function of language. In this light, the recent proliferation of AI in communicative contexts evokes a return to the Shannon-Weaver paradigm, the only difference being the fact that the machines are now positioned as both message encoders and decoders. Despite AI-driven systems' capability of emulating linguistic structures with increasing sophistication, they are structurally incapable of navigating what Gregory Bateson [21] termed as the *meta-communication* layer, which refers to the messages behind the message. Thus, this absence of second-order understanding renders AI systems inadequate in communicative situations where ethical discernment, emotional interpretation, or culturally specific nuance is required. In fact, using AI-driven applications in such particular contexts produces a communicative act stripped of its human texture, practically representing a sterile exchange of words.

Consequently, the development of hybrid communication models emerges as a pragmatic and theoretically sound response, due to these models' capacity to preserve AI's value in high-volume, data-rich environments while reintroducing human agency at critical junctures, particularly where interpretive judgement, relational sensitivity, or reputational risk is involved. In this hybrid communication model, the communication professionals' role shifts from transmitter of information to curators of meaning, in order to assess tone, recalibrate content for diverse publics, and ensure that messages resonate within specific socio-cultural contexts, because the communicative process is one of high complexity [22] that no algorithm can yet fully replicate.

If theory is the architect of practice, then communication models must evolve in response to the technological and epistemic shifts of the present. Thus, the hybrid communicative model presented in Figure 10.5 constitutes such an evolution from the classical paradigms of Lasswell [23] and Berlo's SMCR (source-message-channel-receiver) model [24], or newer frameworks like media richness theory [25]. First of all, Laswell's enduring formulation, "Who says what in which channel to whom with what effect", remains conceptually valid but it now operates within an ecosystem where machine learning co-authors the "what", and digital infrastructures co-select the "channel". Therefore, rather than replacing the foundational models, a more strategic move would be their rational augmentation to accommodate non-human actors and data-driven modalities. To that end, the rational hybrid model reinterprets Berlo's

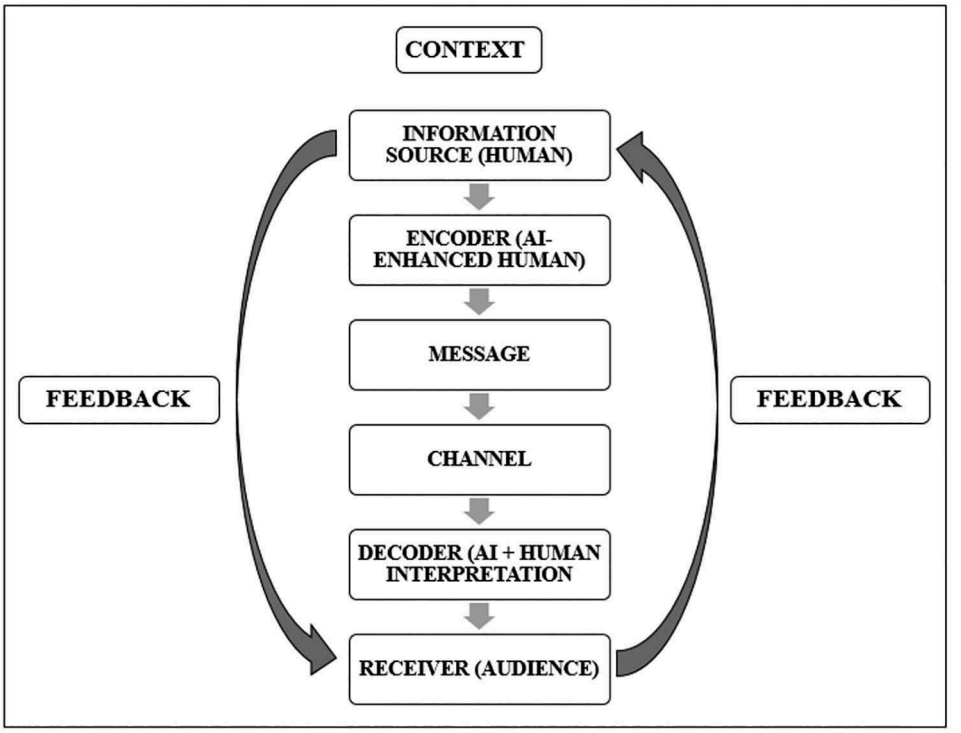


FIGURE 10.5 The new hybrid communicative model AI-driven
Source: Authors' contribution

SMCR schema [24] by reconfiguring the roles of source and encoder as shared between human communicators and algorithmic support systems. Even though Berlo privileged the psychological disposition and communication skills of the human sender, the contemporary landscape introduces a new dynamic driven by AI's capacity to structure, enhance, and even simulate elements of message production. Yet, simulation is not synonym for understanding, thereby the model presented in Figure 10.5 insists on the continuing primacy of human intention and discretion, notably when communicative aims transcend the transactional and move into reals of persuasion, reassurance, or ethical positioning. Finally, the model integrates key premises of media richness theory [25], which argues that communication effectiveness depends on the degree to which a medium can convey nuanced cues and facilitate feedback. For instance, in contexts marked by ambiguity or emotional complexity such as crisis communication or intercultural negotiation, channels of high richness (e.g. synchronous dialogue, video conferencing, in-person engagement) prove to be indispensable. Thus, the hybrid model honours this premise and embeds channel selection within a decision logic that considers not only technical efficiency, but also relational depth and symbolic resonance.

Within this hybrid communicative model illustrated in Figure 10.5, human communicators play the role of both senders and receivers but their objective transcend simply decoding information, in fact they negotiate meaning through a shared

interpretive process influenced by culture, experience, and social context, dimensions no algorithm can fully internalise. Taken together, the model positions communication professionals as strategic interpreters that develop a collaborative relationship with AI-enabled systems to extend reach, increase clarity, and refine message architecture. Hence, humans' role is both technical and epistemological, as they have to assess what counts as meaning, which channels preserve or distort it, and how interpretation unfolds across diverse publics, in order to ensure that communicative acts remain effective, ethically grounded, and humanly intelligible.

Although not without a degree of theoretical resistance, it must be acknowledged that the mechanisation of communication in its present AI-enhanced form has rekindled the very tensions that classical theorists of meaning and discourse sought to resolve, namely between the transmissibility of information and the intelligibility of meaning. As we have endeavoured to show, no amount of computational refinement and no algorithm however elegant in design could replicate the ethical intentionality, contextual elasticity, and emotional precision upon which the most consequential forms of human communication depend. Thus, the hybrid model emerges a dialectical synthesis in which the calculative rationality of artificial systems is deliberately tempered by the interpretive labour of human agents. However, it would be a grave error to presume that such synthesis arises organically or inevitably from the co-existence of machine and mind, taking into account that it must be constructed and sustained by institutions aiming to integrate such technologies without abdicating human judgement. Consequently, this structural and procedural question of how readiness for communicative AI is to be theorised, operationalised, and ethically governed now commands our attention.

6 Organisational AI Readiness for Client-Oriented Communication

If it is true, as historical materialists and institutional economists alike have contended that the tools of production fundamentally transform the relations within which they are deployed, then it follows that the adoption of AI in the communicative domain cannot be confined to a technocratic exercise. Rather, it must be situated within a broader reconfiguration of organisational form and function, wherein questions of infrastructure, knowledge, labour, and legitimacy are jointly at stake. To be more specific, in this more rigorous conception AI readiness is not a static end-state to be achieved through checklist compliance or episodic investment, but a dynamic condition of systemic alignment, one in which abundant data is curated, where infrastructures are not simply digitised but interoperable, and where staff are both trained in usage and cultivated in judgement [26]. Moreover, such readiness must be temporalised and structured to proceed in phases, through cycles of audit, experimentation, internal deliberation, and iterative recalibration [27]. Often neglected in technocratic discourse, the internal dimension demands particular attention, as no AI system can succeed communicatively if the intentions behind its deployment remain opaque to those within the organisation tasked with its implementation. Likewise, beyond a matter of branding or message control, external readiness must represent a principled engagement with clients whose autonomy must be respected through clarity of purpose, transparency of process, and the maintenance of opt-out mechanisms that preserve the ethical primacy of human agency in all communicative acts.

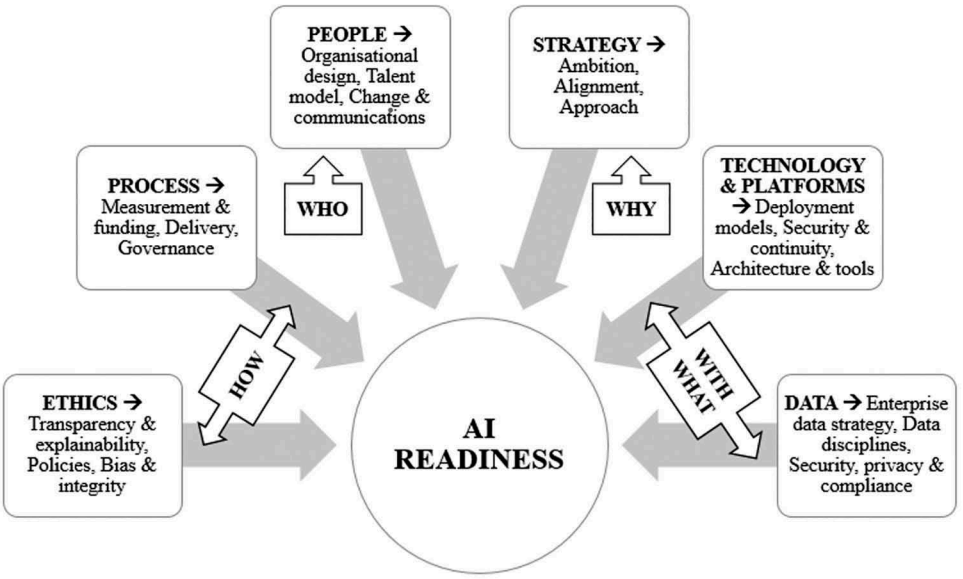


FIGURE 10.6 AI readiness operating model proposed by Deloitte
Source: Authors' contribution based on Deloitte (2020) [28]

To this end, Figure 10.6 proposes a synthetic model for assessing AI readiness in organisational communication and integrates six central pillars: People, Strategy, Technology and Platforms, Data, Ethics, and Process. In fact, these interdependent domains represent the scaffolding upon which a coherent AI implementation must be constructed, reflecting the complexity of organisational transformation in the digital age, where the alignment of human capital, strategic ambition, ethical integrity, and data governance is indispensable. First of all, the inclusion of *People* refers to the necessity of designing agile talent models and fostering change-resilient communication cultures, but also represent the answer for the question of who must be prepared in terms of AI readiness. Secondly, the reason for AI readiness implementation resides in the urgent need for an AI-aligned *Strategy* by all organisations aiming to remain relevant in the digital economy. Thirdly, the resources these institutions must be equipped with are *Data* which addresses the foundational role of quality, governance, and privacy, as well as *Technology & Platforms* that relate to the robustness, scalability, and security of the technological backbone. To answer how AI readiness should be incorporated, this schema proposes the *Ethics* dimension based on non-negotiable principles of transparency, bias mitigation, and policy integrity, whereas *Process* captures the need for ongoing evaluation, delivery structures, and governance. Collectively, these six vectors form a holistic framework through which organisations can assess and augment their communicative capacity in the AI-driven age.

On the other hand, Figure 10.6 also outlines the core challenges associated with deploying AI in client-facing roles, including organisational rigidity, talent constraints, and ethical ambiguities. In contrast, the transition from recognising these challenges to resolving them requires a systemic approach that leverages technology, strategy, leadership,

and cultural readiness [29]. In this context, organisations that demonstrate high digital maturity tend to mobilise specific organisational capabilities that mitigate the risks shown in Figure 10.6. For instance, a fragmented operating model or lack of skilled talent could be addressed through well-articulated transformation frameworks that integrate leadership vision, cross-functional collaboration, and targeted upskilling initiatives [30]. Similarly, the data quality and governance issues raised in Figure 10.6 can only be resolved when organisations adopt structured and scalable AI deployment models grounded in long-term value creation rather than short-term automation gains [31]. As organisations seek to operationalise AI within their client-facing functions, success hinges on a constellation of strategic enablers, besides technological procurement.

Thus, Figure 10.7 presents six enduring success factors identified across digital transformation literature and empirical studies, which collectively place emphasis on long-term effectiveness and adaptability in AI-enhanced business environments. Ranging from the articulation of a business-led digital roadmap to the activation and scaling of AI solutions, these factors constitute a comprehensive framework through which organisations can align their strategic aspirations with their operational realities. In comparison with the diagnostic lens of Figure 10.6, the illustration in Figure 10.7 acts as a prescriptive model that translates abstract information goals into actionable capabilities, nurturing adaptability and sustained performance in AI-enhanced environments.

Therefore, the successful integration of AI into client-oriented communication frameworks demands more than digital technologies or procedural upgrades, it requires a paradigmatic shift in organisational cognition and behaviour. In this vein, previous empirical studies have demonstrated that the effectiveness of AI-driven communication strategies is closely associated with the maturity of internal governance, the clarity

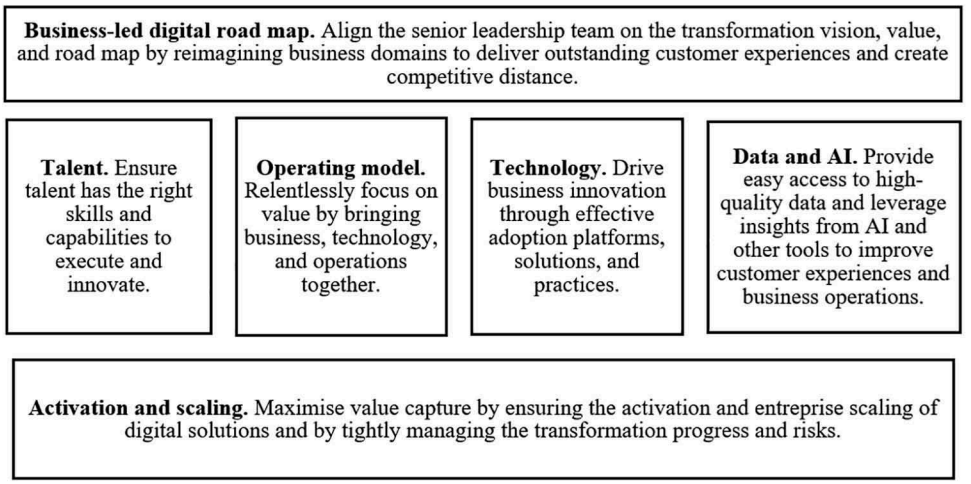


FIGURE 10.7 Framework for the coordinated execution of value creation based on the six enduring success factors enabling tech-business transformations
Source: Authors' contribution based on Mayer et al. [30]

of AI objectives, and the active engagement of cross-functional teams [32]. Additionally, organisations that embed these capabilities tend to transition more smoothly from exploratory AI use to scalable impact, while they also ensure that AI systems are continuously adjusted based on client feedback and evolving market conditions. Furthermore, the readiness of an organisation to deploy AI in external-facing roles, such as communication with clients, is grounded in its ability to balance transparency, ethical foresight, and strategic agility. Specifically, recent scholars put emphasis on the provision of opt-out mechanism and human-in-the-loop protocols aiming not only to enhance client trust but also to align AI adoption with emerging regulatory frameworks such as the EU AI Act [33; 34; 35]. Therefore, rather than associating readiness with speed, it must be understood as the capacity to enact and sustain AI systems that are operationally effective, as well as socially responsible.

Overall, organisational AI readiness is best conceptualised as a layered construct encompassing digital infrastructure, data governance, talent development, and adaptive communication, four dimensions that must operate in synchrony to enable a strategic transition from fragmented experimentation to holistic transformation. In this digital era where AI continues to reshape the terrain of client engagement, firms that proactively build institutional readiness anchored in ethical design, transparent communication, and continuous learning will be better positioned to realise both operational gains and durable client relationships [36].

7 Empirical Synthesis: Sectoral Diffusion, Economic Structure and Implementation Frictions

This section of our chapter integrates empirical data derived from the international and interdisciplinary institutions such as the OECD [37], European Commission, McKinsey, Accenture, and Deloitte. In addition, the data examined spans the quantitative proliferation of AI-driven models, sectoral dependencies, organisational development hierarchies, and economic barriers to implementation, with the objective of achieving a diagnostic interrogation of patterns, frictions, and structural asymmetries that characterise the adoption of AI within client-oriented communicative architectures.

On one hand, the progressive adoption of AI within communicative infrastructures is neither spontaneous nor exogenous, but rather the result of evolving trust dynamics between institutional actors and algorithmic systems. For instance, 78% of organisations reported the use of AI in 2024, representing a substantial increase from 55% one year before, a statistical shift that cannot be dismissed as a cyclical anomaly but must be interpreted as a systemic recalibration in managerial cognition and risk tolerance [37]. Similarly, the amplification of institutional trust is further evidenced by the US Food and Drug Administration's approval of 223 AI-driven medical devices in 2023, which is a monumental ascent from a mere six approvals in 2015. Moreover, the implications of these adoption trends are particularly acute within client-facing communication systems, where AI-enabled systems demonstrate palpable superiority in tasks requiring rapid decision cycles such as time-constrained programming. Therefore, the economic interpretation of this phenomenon mirrors the productivity gains observed during the general-purpose technologies in earlier industrial epochs. Yet, the

main difference resides in the expansion of metrics into tangible dimensions such as satisfaction, reliability, and perceived empathy in client relations [38].

Furthermore, in operational environments these transformations are increasingly embodied in autonomous systems whose real-world integration transcends experimental validation, mainly due to their capacity of generating novel communicative expectations such as consistency, predictability, and responsiveness. Moreover, these attributes, despite their operational nature, also serve as proxies for trust in AI interfaces, thus shifting the communicative burden from relational nuance to algorithmic assurance. Such is the case of Waymo, which executes approximately 150.000 autonomous rides per week, thereby exemplifying the tangible reconfiguration of client service modalities.

On the other hand, trust and satisfaction are not binary constructs but evolving states contingent on technical reliability, institutional endorsement, and user habituation. Therefore, Figure 10.8 provides an illustration of the AI-enabled systems evolution by sector from 2003 to 2024, which is marked by a sectoral proliferation within the industrial fields from 2015 up to present times. Yet, the most inspiring aspect about this exponential dynamic is not its stabilising presence of sectors that already demonstrated sustained absorptive capacity over time, but its adoption in areas where client interaction constitutes a strategic vector.

In comparison with Figure 10.8 that addresses volume of AI-enabled systems over the last two decades, Figure 10.9 refines the view by tracing proportional concentration within sectors and offering insight into structural dependencies. For example, the persistent overrepresentation of technology and financial services suggests the presence of network externalities and capital-intensive feedback loops, as well as path dependency and a possible misalignment between communicative inclusivity and innovation inertia. Additionally, data from Figure 10.9 indicate an exponential increase in industrial AI-enabled systems from 60% in 2023 to surprisingly 90% of total AI models in

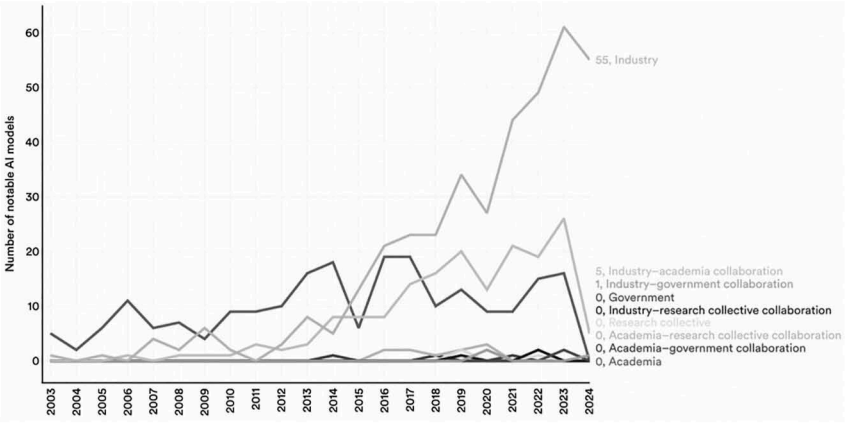


FIGURE 10.8 Number of notable AI models by sector between 2003 and 2024
Source: Maslej et al. [39]

2024, while governmental or academia-research have decreased severely over the same period of time.

Next, Figure 10.10 shifts the analytical unit to institutional output by putting emphasis on the disproportionate contribution of private actors in the development of AI-enabled tools comparing to their academic counterparts, such as MIT or UC Berkeley.

Moreover, data revealed in Figure 10.10 indicate a rather negligible academic contribution to AI-driven landscapes with only two models per each of the two universities, while the biggest winners seem to be Google and OpenAI, each with seven AI-models launched in 2024, followed closely by Alibaba with six models, then by Nvidia, Meta, and Apple, each with four AI models. Thus, the transition from sectoral distribution (Figure 10.8), through proportional intensity (Figure 10.9), to organisational concentration (Figure 10.10), provides a descriptive mapping of the expansion of AI-driven technologies within sectors, but also articulates an economic logic of asymmetric innovation caused by the impact of infrastructural and institutional consolidation on the diffusion of AI in client communication.

Another key factor in AI-enabled system diffusion across domains of activity is their inference price, which indicates the expenses of querying a trained model and is measured in USD per million tokens. Hence, Figure 10.11 represents a very illustrative graphic of the downward trend in inference cost across multiple performance benchmarks over the period from 2022 to 2024, dropping almost ten times for GPT-4o in merely seven months after its launch in May 2024. Besides technological optimisation, this decline also represents an increasing marginal efficiency of computational resources, thereby lowering the economic barriers to deploying AI at scale in communication contexts. Consequently, as inference prices fall, broader organisational participation in AI adoption becomes feasible, and cost-efficiency augments the plausibility of embedding AI interfaces into real-time client engagement protocols without prohibitive investment thresholds.

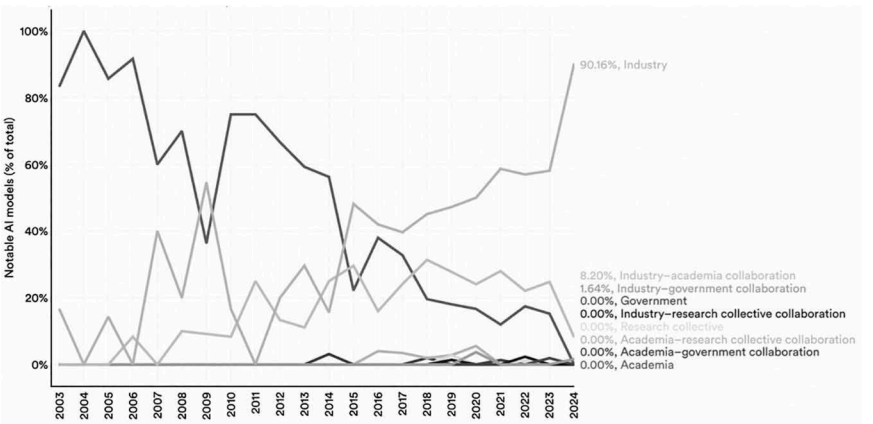


FIGURE 10.9 Number of notable AI models (% of total) by sector between 2003 and 2024
Source: Maslej et al. [39]

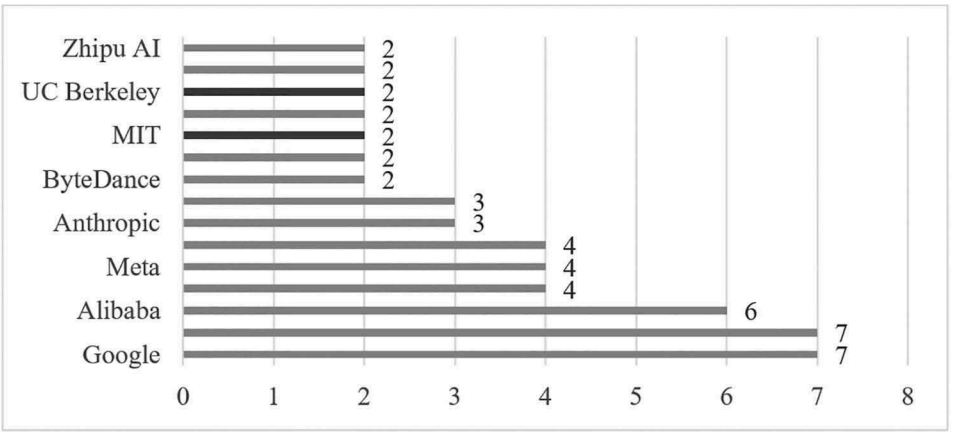


FIGURE 10.10 Number of notable AI models by organisation in 2024
Source: Authors' contribution based on Maslej et al. [39]
Note: UC Berkeley, MIT = academia; all the other organisations are industry

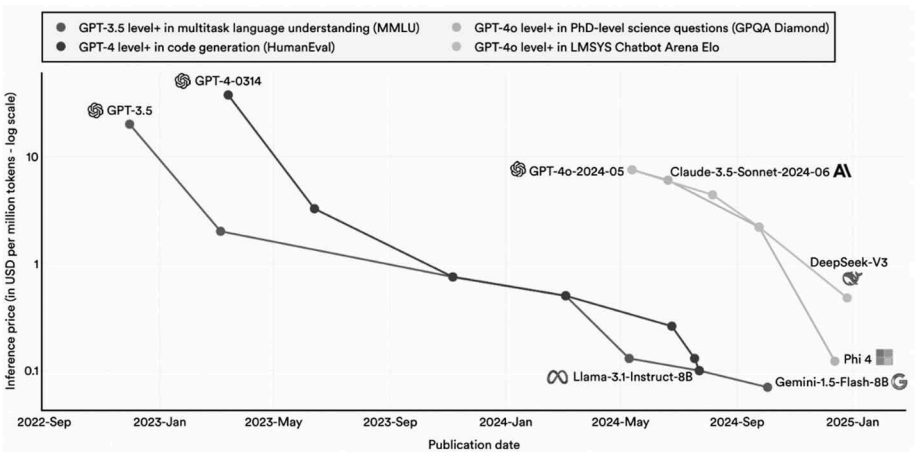


FIGURE 10.11 Inference price across select benchmarks between 2022 and 2024
Source: Maslej et al. [39]

In contrast to inference cost, Figure 10.12 captures the volatile and often escalating expense of training state-of-the-art AI models for the period between 2019 and 2024. As expected, data in Figure 10.12 indicates a rather upward escalating trend, marked by enormous amounts of funds invested into AI-driven systems training. For example, in 2020 the company OpenAI invested 4 million dollars into training GPT-3, while training GPT-4 was 20 times more expensive in 2023. Similarly, the estimated training cost for Llama-2–708 was only 3 million dollars in 2023, whereas in 2024 the training price has reached 170 million dollars, which is 56 times more expensive than the previous training cost.

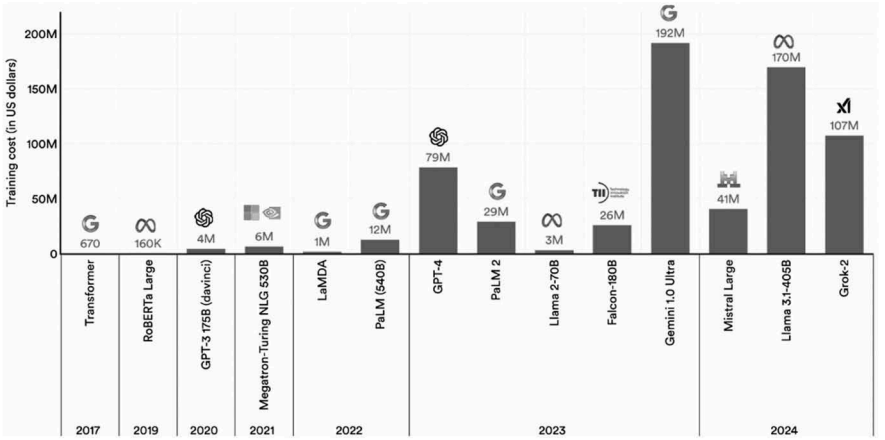


FIGURE 10.12 Estimated training cost of select AI models between 2019 and 2024
Source: Maslej et al. [39]

Furthermore, Figure 10.13 deepens the insight emphasised by Figure 10.12 by plotting training cost against computational scale for AI models developed in 2024, further indicating a steep correlation and an unsustainable resource intensiveness of cutting-edge models. Hence, such asymmetry raises questions about environmental externalities, organisational scalability, and the allocative efficiency of communicative AI investments.

Taken together, Figures 10.11 through 13 place emphasis on a bifurcated cost structure in the AI lifecycle, where inference becomes progressively accessible and cost-effective (Figure 10.11), and the capital constraints around training and computing (Figures 10.12 and 10.13) produce a centralised architecture of capability. In this light, communicative efficacy becomes a consequence of economic stratification in the production process, whereas the promise of AI-enhanced communication rests on the paradox of limiting the adaptability of AI systems to specific communicative contexts. In addition, Figure 10.14 introduces a necessary counterpoint to the otherwise linear narrative of AI integration by providing a taxonomy of incidents reported by organisations participant in McKinsey's survey between 2022 and 2024 [39], from adversarial attack (56%) and privacy violation (55%), to unintended decision making (51%), model bias (47%), or performance failure (46%).

Complementary to the empirical concerns reported in Figure 10.14, Figure 10.15 illustrates the organisational obstacled to integrating responsible AI (RAI) frameworks, such as: knowledge and training gaps (51%), resource or budget constraints (45%), regulatory uncertainty (40%), technical limitations (32%), organisational resistance (22%), and lack of executive support (16%). In contrast, without an embedded RAI architecture, AI systems deployed in client contexts risk amplifying asymmetries of understanding and exacerbating distrust.

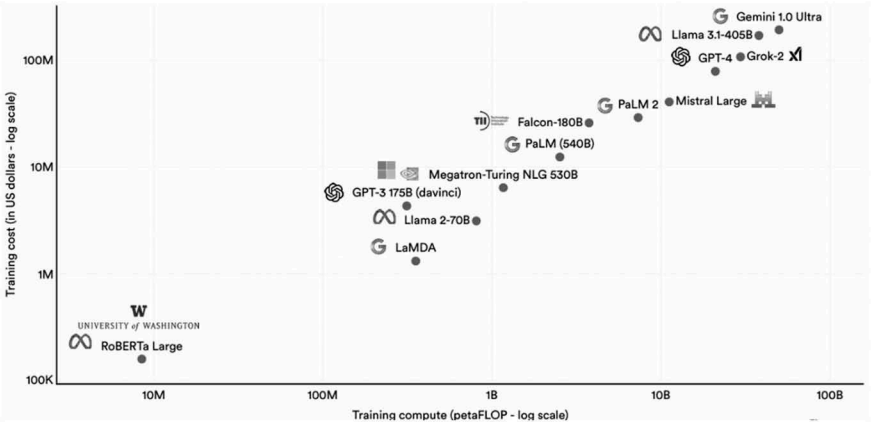


FIGURE 10.13 Estimated training cost and compute of select AI models in 2024
Source: Maslej et al. [39]

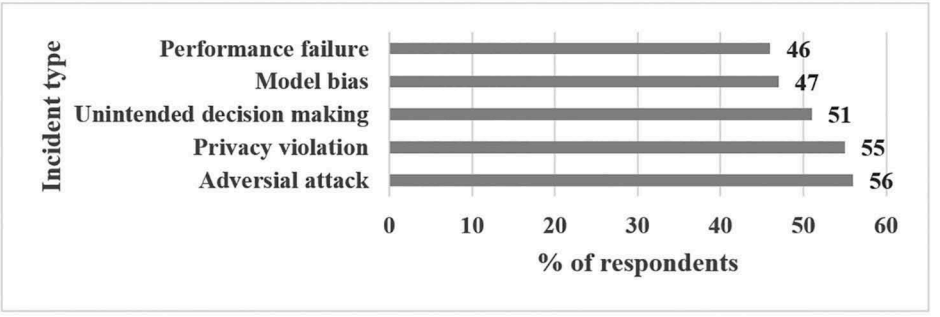


FIGURE 10.14 AI-related types of incidents reported by organisations between 2022 and 2024
Source: Authors' contribution based on Maslej et al. [39]

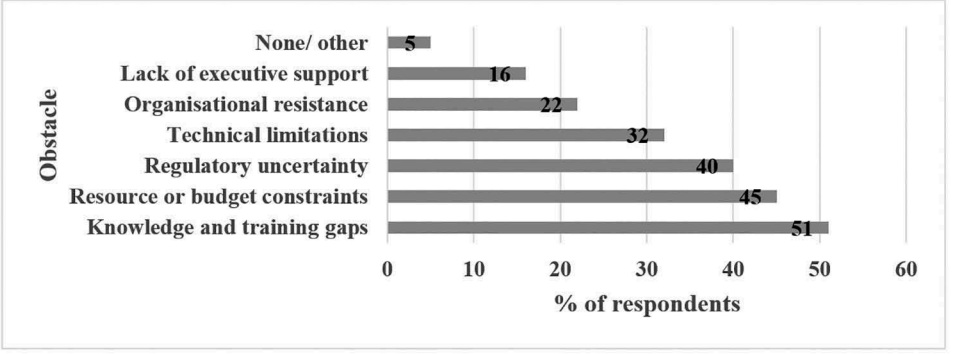


FIGURE 10.15 Main obstacles to the implementation of responsible AI (RAI) measures in 2024
Source: Authors' contribution based on Maslej et al. [39]

Another important benchmark of the RAI framework are considered to be the refusal rates of AI models, due to the fact that high refusal rates may be interpreted as evidence of embedded safeguards. As AI models have evolved, they have also marked increasingly higher refusal rates, as shown in Figure 10.16 where DeepSeek's refusal rate increased from v3's 0.41 to R1's refusal rate of 0.53, or OpenAI's refusal rate constantly ameliorating with 0.35 from o1-mini (0.45) to o1 (0.80), both launched in 2024 just three months one after the other. However, when refusal rates surpass a certain threshold, user experience risks devolving into frustration which erodes the very trust that responsible design it seeks to cultivate.

Collectively, Figures 10.14 through 16 represent a very important contribution in terms of reframing the discourse on AI efficacy by putting emphasis on its operational, ethical, and experiential liabilities. Additionally, these figures also illuminate the paradox of precaution because the more one encodes safety and responsibility, the more constrained and potentially ineffective the system becomes in real-world interaction, which represents a tension inviting a more nuanced theorisation of communicative robustness as a dynamic equilibrium negotiated across multiple axes of institutional capability, user expectation, and epistemic modesty.

In light of the preceding empirical analysis marked by a dialectic between systemic acceleration and normative constraints, the analytical arc returns to the foundational research question posed at the chapter's beginning which investigates in what manner might enterprises communicate the multidimensional value of AI to their clients. Rather than offering a simplistic reduction, Figure 10.1 furnishes a finely calibrated conceptual apparatus comprising three interdependent dimensions: the epistemological decoding of client perceptions, the articulation of value through quantifiable deliverables, and the orchestration of algorithmic intricacies into accessible and iterative frameworks. Thus, far from being a static model, this tripartite framework represents a performative scaffold upon which dynamic communication strategies may be staged,

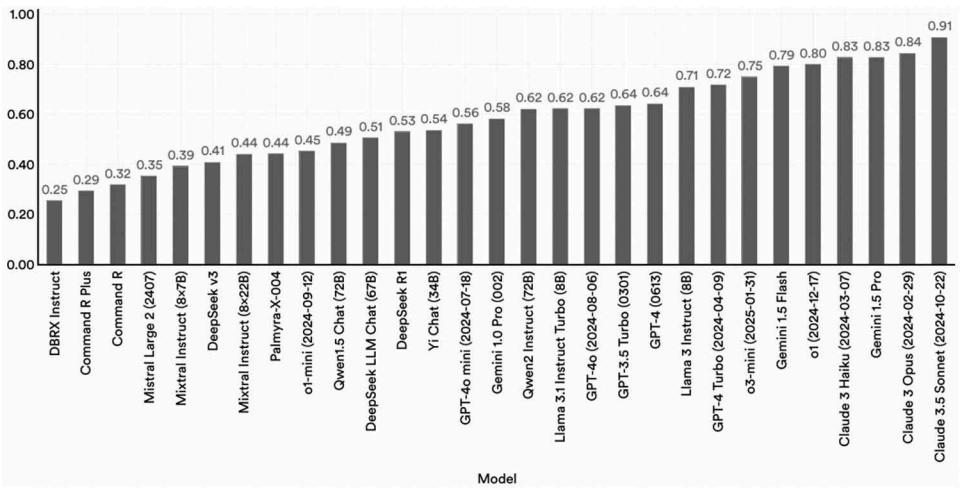


FIGURE 10.16 AIR-bench expressed by their refusal rate in 2024
Source: Maslej et al. [39]

refined, and operationalised. Firstly, the initial pillar interpreting the customer's vantage point transcends the transactional logic of market segmentation and enters the realm of affective computation, because it mandates the deployment of advanced behavioural diagnostics, sentiment parsing engines, and predictive preference models which, in concert, transmute the abstract consumer into a data-rich, dialogic subject. This epistemic elevation of the client from statistical artefact to semiotic interlocutor is observable in real-world ecosystems such as Amazon's algorithmic recommendation matrices or Netflix's adaptive content curations, all these AI-driven systems being recalibrated continuously in terms of communicative relevance and intuitive anticipation. Secondly, the next dimension of Figure 10.1 is represented by rendering value intelligible via outcome-oriented metrics such as return-on-investment indices, real-time decision analytics, and strategic benchmarking. This reconfiguration is already exemplified by corporations such as Salesforce and SAP which embed analytics into interface design so that client communication is easily measured, modulated, and authenticated not from the boardroom but from dashboards. The final pillar encompasses engineering communicative simplicity amidst computational density, but also entails the synthesis of interface design, user experience (UX), and adaptive learning logics into a coherent epistemic encounter. For example, platforms like Grammarly and Duolingo instantiate a recursive feedback loop wherein users receive guidance but iteratively evolve in tandem with the system's own learning architecture. Thus, the communicative function suffers great transformations from the declarative interaction to the tacit co-evolution.

Aiming to operationalise the principles outlined in Figure 10.1, we propose in Figure 10.17 a communication model based on three interdependent vectors: interpretive analytics, performative outcomes, and adaptive design. Each component is configured to function as an iterative communicative system that evolves in response to client feedback, data refinement, and system learning. First, the interpretive vector mobilises granular data extraction and contextual modelling to construct high-resolution client profiles, which represent dynamic and continuously updated cognitive maps guiding message personalisation, tone calibration, and expectation alignment. For this vector, the methodological core is based on natural language processing (NLP), affective computing, and behaviour-based segmentation, transforming communication in a real-time interpretive act in which AI systems function as both interlocutors and semiotic analyst.

Second, the performative vector illustrated in Figure 10.17 translates operational outcomes into communicative value, requiring embedding key performance indicators (KPIs) within communication streams, transforming client-facing messages into evidentiary artefacts, and promoting dashboards, conversational agents, and real-time analytics visualisations as the new mediums of expression. It must be noted that this vector insists on traceability, transparency, and empirical rigour, thus value is not narrated but evidenced in order to enable trust to be transferred from human brand ambassadors to AI-driven communicative platforms. Lastly, the adaptive vector reconfigures the user-system interface into a pedagogical space, wherein communication becomes an evolving interaction where AI systems modulate complexity, tempo, and content structure based on user behaviour and feedback loops. For this vector, the main design principles rely on explainable AI (XAI), interface minimalism, and

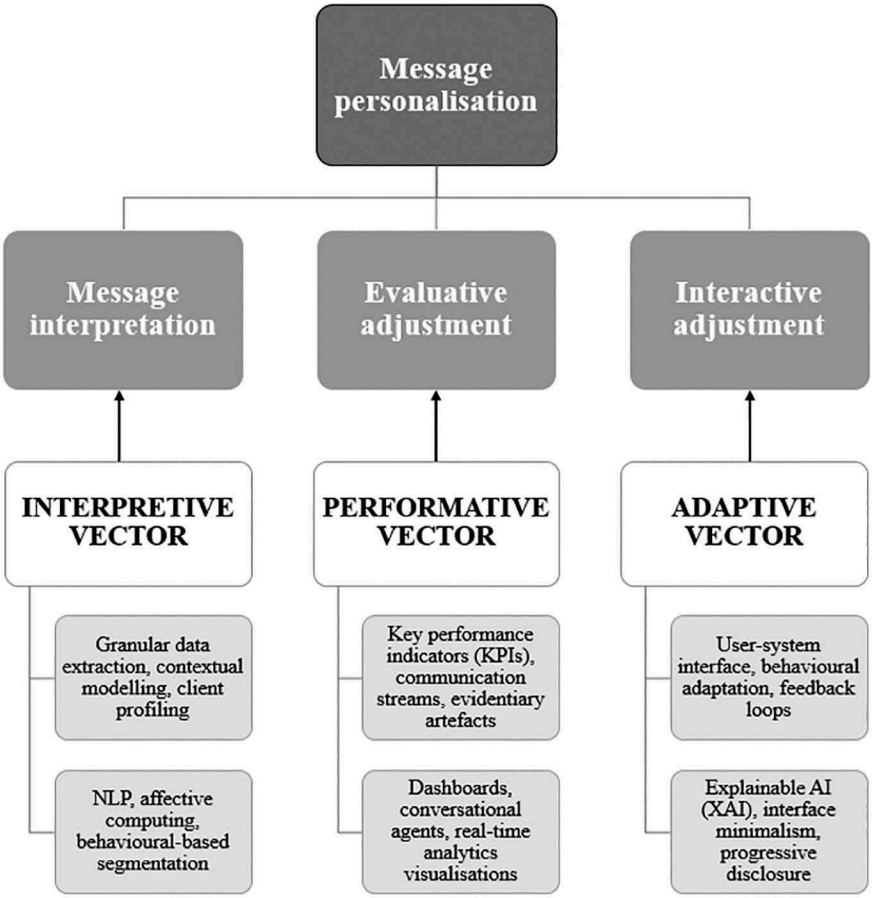


FIGURE 10.17 The triadic vectors of AI-driven client communication
Source: Authors' contribution

progressive disclosure, the main aim being to foster cognitive comfort and procedural familiarity thereby reinforcing the user's sense of agency, competence, and trust.

Consequently, these three vectors form a triadic communication model that mirrors the epistemic logic of Figure 10.1, but also recognises communication as a recursive, data-anchored, and user-responsive practice, one capable of capturing the full spectrum of value that AI can deliver to potential clients.

8 Challenges and Ethical Imperatives in AI-Mediated Communication

The diffusion of algorithmic communicative agents into client-oriented domains requires a reappraisal of longstanding ethical categories, mainly because the ethical contemporary landscape is characterised by two intersecting dynamics: a surge in deployment and entrenchment of AI across sectors, and a proliferation of normative instruments and guidelines that seek to govern such deployment [39; 40]. Moreover,

the ethical stakes associated with AI-enabled agents in communication cannot be reduced to single failure modes, since they are rather distributed across epistemic (explainability), distributive (bias and fairness), procedural (consent and governance), and systemic (accountability and institutional design) dimensions. In fact, each dimension interacts with communicative aims such as trust, clarity, and responsiveness, and with organisational incentives (efficiency, scale, brand reputation), thereby producing a typology of risk and opportunity that must be managed through coherent policy and design choices [39].

On one side, one very important concern in client-oriented systems is the propensity for algorithmic processes to reproduce or amplify pre-existing social asymmetries. Specifically, machine learning systems possess the capacity to optimise objective functions on historically produced data, and if those data encode structural inequities, the resulting outputs will characteristically replicate those inequities [41]. Additionally, scholars have drawn attention to how ostensibly neutral optimisation criteria yield disparate impacts in domains ranging from credit scoring to targeted advertising, whereas the consequence for communication is that certain audiences may be misrepresented, excluded, or patronised by automated interlocutors, thereby undermining trust and exacerbating reputational risk. Therefore, addressing distributive fairness demands interventions at multiple points in the pipeline: data collection and curation practices, model selection and validation with counterfactual or subgroup analyses, ongoing monitoring for distributional drift, and governance structures that enable redress. In contrast, formal fairness metrics such as demographic parity, equal opportunity, and calibration are useful analytical instruments but remain unable to substitute for participatory governance that identifies which fairness criteria are legitimate within particular communicative contexts and which trade-offs are socially acceptable [42].

Secondly, there is quite often for clients confronted with decisions or recommendations emerging from black-box systems to reasonably demand for explanations, which in fact are components of epistemic legitimacy. To this matter, the literature on explainable AI (XAI) has expanded significantly over the last five years, thereby offering a numerous diversity of local and global explanation methods as well as systematised surveys of methods and their limitations [43]. Nevertheless, the translation of model-centric explanations into communicative artefacts that non-expert clients can consume remains problematic, since explanations that are technically faithful might still be unintelligible. Conversely, simplifications that render black boxes comprehensible risk misrepresenting causal structure or producing misleading assurance. Also, organisations face great regulatory pressure pushing them towards greater transparency, often associated with design guidance due to the particularity of practical integration of XAI into client interactions requiring iterative human-centred evaluation: A/B testing with real users, suability assessment, and domain specific interpretability validation [43].

Next, privacy concerns in communicative AI are twofold: the direct handling of personal data and the indirect inferences derived from behavioural traces. On one hand, the contemporary regulatory regimes such as the EU General Data Protection Regulation (GDPR) provide a legal framework that predicates lawful processing on demonstrable principles like purpose limitation, data minimisation, and the right to transparency [44]. On the other hand, these principles are straightforward in

formulation but demanding in application, with conversational agents collecting richly granular interactional data that admit sensitive inferences, and informed consent in online systems often degenerating into procedural formality. Therefore, the literature recommends privacy-by-design praxis and concrete procedural mechanics (data protection impact assessments, meaningful granular consent interfaces) so that systems which mediate client relationships are engineered to reduce the surface area of invasive inference and to preserve autonomy where possible [45]. From an organisational perspective, it must be noted that beyond legal compliance firms must recognise that consent alone remains unable to restore trust if core practices (profiling, automated scoring, retention policies) rest opaque or unconstrained.

Moreover, contemporary proposals for algorithmic accountability intersect jurisprudence and organisational design because the answer to the question of identifying the responsible entity when a model misfires points out to institutional fault lines [35; 46]. This is mainly due to firms' responsibility to assemble governance chains that ensure traceability through versioned model registries, reproducible pipelines, internal audit functions, and cross-functional risk committees, whereas accountability cannot be delegated exclusively to technocratic artefacts. Instead, promoting accountability implies empowered human roles such as audit officers, ethics councils, or named responsible individuals, but also external channels for remedy [46].

On the other side, it must be acknowledged that communicative AI-driven platforms are exposed to many technical hazards such as adversarial manipulations, data poisoning, model extraction, and systemic failure under distributional shift. For instance, scholars have demonstrated through a diverse palette of modalities ranging from images to text and audio how adversarial examples and related attacks can pose distinct risks when systems are used in client interactions, resulting in manipulated inputs that produce erroneous or harmful advice [47]. Likewise, other studies in the realm of malicious use have argued that vulnerabilities in widely deployed models can be weaponised, thus raising the imperative for robust testing, red-team exercises and hardened production deployments [47]. In that vein, operational resilience must be designed into architectures aiming to preserve safety and protect the normative commitments implicit in client communication: graceful degradation modes, human supervision thresholds, anomaly detection and robust fallback communications that escalate to human operatives when confidence scores or anomaly detectors cross pre-defined boundaries [43]. Although precautionary design choices such as conservative refusal policies are an important mechanism for preventing harm, they also create a tension due to the fact that excessive refusal rates undermines usability and erodes the very trust such mechanisms are designed to protect. To this end, empirical studies of refusal behaviour generated by AI-enabled communicative systems indicate that premium models have evolved to increase refusal thresholds, yet users may interpret frequent refusal as functional brittleness [38]. Therefore, communicative systems must balance protective filtering with informative fallback, specifically a refusal should represent a communicative act that explains why the system is unable to answer, what the client might expect, and how human support could be engaged, thereby treating refusals as opportunities for transparent reciprocity rather than as a blunt safety silencer.

Furthermore, the transition from rhetoric to practice in ethical terms requires organisations to implement concrete procedural instruments, such as: impact

assessments (privacy, fairness, and safety), routine audits, and external certification mechanisms. For instance, a Data Protection Impact Assessment (DPIA) mandated under GDPR for high-risk processing represents a paradigmatic example of ethical principles integration because it forces systematic documentation or risks, mitigation strategies and residual harms. Similarly, fairness audits and red-team evaluations can provide empirical scrutiny of model behaviour across subgroups and use-cases, while external validation by independent third parties can help legitimate organisational claims of responsible practice [40]. From the perspective of communication professionals, the arrival of AI-enabled systems requires an expansion of domain expertise that integrates technical literacy with ethical discernment, so that practitioners become fluent in what models can and cannot do, capable to interpret evaluation metrics, and skilled at translating technical constraints into client-facing explanations and options. Thereby ethical judgement in communicative settings will often rest on contextualised decisions like choosing modes of disclosure that balance transparency with usability, designing opt-out pathways, or deciding when to escalate to human engagement, so that training programmes and cross-functional collaboration frameworks become indispensable. Thus, the concept of human-in-the-loop must be operationalised as a set of role descriptions, escalation criteria, and performance metrics that ensure timely and informed human intervention [40].

Accordingly, a minimal ethical toolkit for communicative AI should include six measures complementary for coupling XAI with participatory design practices: purposive data governance, XAI-driven explanation interfaces calibrated to user context, procedural DPIAs and fairness audits, human-in-the-loop escalation and restorative channels, robust security engineering, and governance architectures that allocate responsibility with named owners and independent review [35].

9 Conclusions and Future Research Directions

The assimilation of AI into client-oriented communication places emphasis on a massive reconfiguration of the ways in which organisations conceive, enact, and preserve value in their stakeholder relationships. Thus, this chapter has carried out an interrogation on the multiple dimensions through which AI is recalibrating communicative practice, putting emphasis on its capacity to automate processes, tailor engagements, and refine interactions across an increasingly heterogeneous array of contact points. Nevertheless, our chapter has also foregrounded the structural constraints of such systems, specifically their inability to embody the subtleties for ethically contingent decisions, or to apprehend the full depth of socio-cultural context, which remain the preserve of human professional communicators whose interpretative and affective faculties continue to constitute an indispensable communicative asset.

As previously argued, at the core of this transformation resides a productive, yet unresolved tension between efficiency and empathy, between the procedural regularity of automation and the situational responsiveness of authenticity. Hence, the analysis advanced here positions integration rather than substitution as the most credible trajectory for communicative innovation, leading the path for a new hybrid communication model where AI is not cast as the usurper of human communicators, but as a disciplined partner whose operational precision can be orchestrated alongside human

capability and ethical judgement. Moreover, the chapter has sought to articulate a conceptual architecture within which future communicative models might evolve by advancing a hybrid paradigm in which human and machine agents operate in complementarity, each compensating for the other's structural limitations.

Furthermore, following the prior dissection of the structural tension between algorithmic efficiency and human empathy, the trajectory of AI from deterministic, rule-bound architectures towards adaptive, multimodal systems has reengineered the operational conditions under which organisations engage with their clients. Additionally, the emergence of conversational agents, predictive analytics environments, and virtual assistants as semi-autonomous communicative entities has introduced a qualitatively distinct mode of interaction, one in which the parsing of sentiment, the anticipation of latent needs, and the delivery of precisely calibrated informational content occur as integrated functions. Therefore, such systems have modified the communicative calculus itself by displacing reactive exchanges with architectures designed for anticipatory engagement and pre-emptive alignment of service to expectation. However, this reorganisation of communicative practice is accompanied by inherent structural limitations whose significance cannot be understated. For example, the incapacity of AI systems to replicate genuine emotional intelligence constitutes a decisive challenge notably where the communicative act must encompass empathy, cultural literacy, or the ethical navigation of ambiguous circumstances. Likewise, empirical studies in both crisis communication and intercultural negotiation have demonstrated that algorithmically generated responses tend to fracture under conditions of uncertainty, urgency, or emotional volatility, domains in which the interpretive dexterity and moral reasoning of human actors remain indispensable to the preservation of trust and legitimacy [1; 6; 39; 48]. Consequently, this recognition calls for a design of communicative systems that neither subordinate human agency to automation nor dilute the computational advantages of machine intelligence, particularly a model predicated on principled integration, wherein the algorithmic capacities for scale, speed, and data-driven precision are systematically interwoven with human faculties of ethical judgement, situational awareness, and adaptive modulation.

In order to render the preceding synthesis actionable within organisational practice, this chapter advances and AI-driven hybrid communication model in which algorithmic capacities are orchestrated in continuous interplay with human oversight. In addition, this innovative framework is conceived as a reconstitution of communicative architecture itself, wherein the generation, transmission, and interpretation of meaning are distributed between machine intelligence and human agency in a manner that is both complementary and reflexive. To this end, our model carries out a critical re-examination of foundational communication theories such as Shannon and Weaver's mathematical model of information transmission, Berlo's SMCR schema of source-message-channel-receiver relations, and the contingent parameters of media richness theory, with the aim of recasting them within the conditions of a digitally mediated environment. In such an ecosystem, communicative efficacy transcends the traditional linear transmission or channel optimisation to become a process of co-construction, in which AI systems execute high-velocity analytical and generative functions while human interlocutors embed these outputs within ethical, cultural, and contextual frames.

Hence, Table 10.3 below illustrates the principal components of this hybrid model and provides an analytical taxonomy and a repertoire of empirically grounded exemplars, so it can also serve as a strategic instrument for the design of AI–human communicative systems. Specifically, this hybrid model positions AI as a technical enabler and humans as epistemological stewards, whereby together they form a communicative ecosystem that balances scale with sensitivity, and automation with authenticity.

Despite the coherent and analytically rigorous framework for apprehending the role of AI within client-oriented communication, certain limitations must also be acknowledged as they extend beyond the realm of technological implementation only to encompass epistemological, ethical, and organisational dimensions. Thus, Table 10.4 provides an illustrative representation of the principal limitations currently encountered in the deployment of AI within client communication paradigms, accompanied by pertinent research directions for future scholarly inquiry. Although it is evident that AI-enabled applications present many opportunities for organisations aiming to increase their institutional efficacy, AI systems are yet to attain a level of emotional intelligence comparable to human interlocutors. Another challenges relates to the accurate interpretation of contextual cues, particularly those influenced by cultural variability, or even to the ethical domain where algorithmic transparency and inherent biases undermine user confidence and necessitate rigorous ethical oversight. Additionally, the effective integration of AI technologies is limited by the

TABLE 10.3 Operational dynamics of the AI–human interface in client-oriented communication

Component	*Description*	*AI role*	*Human role*	*Literature example*
Message generation	Crafting content based on user data and context	NLP-based personalisation	Ethical framing, tone calibration	Kietzmann et al. (2018)
Channel selection	Choosing appropriate medium for message delivery	Predictive analytics	Contextual judgement, cultural sensitivity	Daft and Lengel (1986), Cheng et al. (2022)
Feedback interpretation	Analysing client responses and adjusting strategy	Sentiment analysis, behavioural tracking	Emotional decoding, relational adaptation	Rubin et al. (2024)
Crisis communication	Managing sensitive or high-stakes interactions	Alert systems, escalation protocol	Empathy, reassurance, reputational management	Jaquelier et al. (2025)
Value co-creation	Engaging clients in personalised, iterative dialogue	Adaptive learning, real-time personalisation	Strategic engagement, trust cultivation	Cheng et al. (2022)
Ethical governance	Ensuring transparency, fairness, and user autonomy	Explainable AI, bias detection	Policy design, opt-out mechanisms	Veale and Borgesius (2021)

Source: Authors' contribution

TABLE 10.4 Limitations and research trajectories in AI-mediated communication

Limitation	*Description*	*Recommendation for future research*
Emotional intelligence deficit	AI systems lack genuine empathy and emotional nuance	Develop emotionally intelligent AI using affective computing frameworks
Contextual misinterpretation	AI struggles with ambiguity and cultural context	Explore context-aware interaction design and cross-cultural AI training
Ethical ambiguity	Algorithmic opacity and bias threaten user trust	Advance explainable AI and ethical auditing protocols
Organisational readiness gaps	Many forms lack infrastructure and talent for AI integration	Study phased AI adoption models and workforce upskilling strategies
Overreliance on automation	Excessive automation risks alienating clients	Investigate optimal human–AI balance in communication workflows
Limited longitudinal data	Short-term studies dominate current literature	Conduct longitudinal studies on AI's impact on client loyalty and trust
Fragmented theoretical integration	Communication theories are underutilised in AI research	Synthesise classical and contemporary models in AI communication design
Regulatory uncertainty	Emerging laws lack clarity on communicative AI	Analyse legal frameworks and propose policy guidelines for AI use

Source: Authors' contribution

heterogeneity of organisational preparedness and by the overdependence on automation, which may inadvertently estrange clients and erode relational dynamics. Likewise, the predominance of short-term studies has resulted in a paucity of longitudinal data to substantiate the sustained impact of AI-driven communication on client trust and loyalty, while the evolving regulatory landscape still remains ambiguous putting emphasis on the urgency for comprehensive legal analyses and the formulation of coherent policy frameworks.

Notwithstanding, the chapter concludes with a prospective reflection on the evolving role of AI in client communication and places emphasis on the true potential of AI to foster synergistic interactions between human cognition and machine intelligence, thereby delivering an innovative conceptualisation that repositions AI as a collaborative interlocutor that amplifies human competencies while safeguarding the interpretive depth inherent to interpersonal exchange. Within this framework, communication professionals assume the role of strategic interpreters tasked with curating meaning, modulating tone, and upholding ethical considerations, whereas AI-enabled tools function as a dynamic adjunct that enhance the breadth, precision, and responsiveness of communicative acts.

In conclusion, this chapter argues the proposition that organisational conceptualisations of communication require a fundamental shift from perceiving communicative act as a linear, unidirectional process towards recognising it as a co-constructed, ethically governed, and affectively sophisticated phenomenon. Hence, through the deliberate incorporation of AI within such hybrid frameworks organisations can optimise AI's functional potential, as well as safeguard the essential human dimensions of relational depth and meaningful engagement.

References

[1] Babatunde, S., Odejide, O., Edunjobi, T. & Ogundipe, D. (2024). The role of AI in marketing personalization: A theoretical exploration of consumer engagement strategies, *International Journal of Management & Entrepreneurship Research*, 6 (3), pp. 936–949. Available at: doi:10.51594/ijmer.v6i3.964.

[2] Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. doi:10.1177/1094670517752459.

[3] McLean, G., & Wilson, A. (2019). Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. *Computers in Human Behavior*, 101, 210–224. doi:10.1016/j.chb.2019.07.002.

[4] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://www.jstor.org/stable/249008.

[5] Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A multiple-item scale for assessing electronic service quality. Journal of Service Research, 7(3), 213–233. doi:10.1177/1094670504271156.

[6] Chen, J.-S., Le, T.-T.-Y. and Florence, D. (2021). Usability and responsiveness of artificial intelligence chatbot on online customer experience in e-retailing, *International Journal of Retail & Distribution Management*, 49(11), pp. 1512–1531. doi:10.1108/IJRDM-08-2020-0312.

[7] Kietzmann, J., Paschen, J. & Treen, E. R. (2018). Artificial intelligence in advertising: how marketers can leverage artificial intelligence along the consumer journey, *Journal of Advertising Research, 58*(3), pp. 263–267. Available at: doi:10.2501/JAR-2018-035.

[8] Verhoef, P. C., Kannan, P. K. & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: introduction to the special issue, *Journal of Retailing*, 91(2), pp. 174–181. Available at: doi:10.1016/j.jretai.2015.02.005.

[9] McKinsey & Company (2025). The state of AI: How organizations are rewiring to capture value. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.

[10] Bakhtin, M. M. (1981). *The Dialogic Imagination: Four Essays*. Edited by M. Holquist and C. Emerson. Austin: University of Texas Press. Available at: http://www.europhd.net/sites/europhd/files/images/onda_2/07/27th_lab/scientific_materials/jesuino/bakhtin_1981.pdf.

[11] Grice, H. P. (1975) Logic and Conversation. In Studies in the Way of Words. Cambridge, MA: Harvard University Press. Available at: https://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf.

[12] Luhmann, N. (1995) *Social Systems: Essays on Self-Organisation*. Stanford University Press. Referenced via synthesis in: Keenan, B. & Sokol, K. (2023) Mind the gap! Bridging explainable artificial intelligence and human understanding with Luhmann's functional theory of communication, *arXiv*. Available at: https://arxiv.org/abs/2302.03460.

[13] Davenport, T. H. & Ronanki, R. (2018). Artificial intelligence for the real world, *Harvard Business Review, 96*(1), pp. 108–116. Available at: https://hbr.org/2018/01/artificial-intelligence-for-the-real-world.

[14] Jarek, K. & Mazurek, G. (2019). Marketing and artificial intelligence, *Central European Business Review*, 8(2), pp. 46–55. Available at: doi:10.18267/j.cebr.213.

[15] Hardcastle, K., Vorster, L. and Brown, D. M. (2025). Understanding customer responses to ai-driven personalized journeys: Impacts on the customer experience, *Journal of Advertising*, 54(2), pp. 176–195. doi:10.1080/00913367.2025.2460985.

[16] Rubin, M., Arnon, H., Huppert, J.D. and Perry, A. (2024). Considering the role of human empathy in AI-driven therapy, *JMIR Mental Health*, 11, e56529. Available at: doi:10.2196/56529.

[17] Juquelier, A., Poncin, I. and Hazée, S. (2025). Empathic chatbots: A double-edged sword in customer experiences, Journal of Business Research, 188, 115074. Available at: doi:10.1016/j.jbusres.2024.115074.

[18] Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press. Available at: https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

[19] Jakobson, R. (1960). Closing statement: Linguistics and poetics. pp. 350–377. Available at: https://monoskop.org/images/archive/8/84/20141028191454!Jakobson_Roman_1960_Closing_statement_Linguistics_and_Poetics.pdf.

[20] Habermas, J. (1984). *The Theory of Communicative Action. Volume One: Reason and the Rationalization of Society*. Boston: Beacon Press. Available at: https://teddykw2.wordpress.com/wp-content/uploads/2012/07/jurgen-habermas-theory-of-communicative-action-volume-1.pdf.

[21] Bateson, G. (1972). *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Chicago: University of Chicago Press. Available at: https://raggeduniversity.co.uk/wp-content/uploads/2023/10/1_xtract_Gregory-Bateson-Steps-to-an-Ecology-of-Mind_-Collected-Essays-in-Anthropology-Psychiatry-Evolution-and-Epistemology-Jason-Aronson-Inc-1987.pdf.

[22] McQuail, D. (2010). *McQuail's Mass Communication Theory*. 6th ed. London: SAGE. Available at: https://nibmehub.com/opac-service/pdf/read/McQuail's%20Mass%20communication%20theory.pdf.

[23] Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (ed.) *The communication of ideas*. New York: Harper and Brothers.

[24] Berlo, D. K. (1960). *The Process of Communication: An Introduction to Theory and Practice*. New York: Holt, Rinehart and Winston. Available at: https://archive.org/details/dli.ernet.235577.

[25] Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements: Media richness and structural design. *Management Science, 32*, 554–571. Available at: http://www.jstor.org/stable/2631846.

[26] Butler, T., Espinoza-Limón, A. & Seppälä, S. (2023). Towards a Capability Assessment Model for the Comprehension and Adoption of AI in Organisations. *arXiv*. Available at: https://arxiv.org/abs/2305.15922.

[27] Übellacker, T. (2025). Making Sense of AI Limitations: How Individual Perceptions Shape Organizational Readiness for AI Adoption. *arXiv*. Available at: https://arxiv.org/abs/2502.15870.

[28] Deloitte Insights. (2020). *AI readiness in government: Six areas for assessing AI readiness*. Deloitte. Available at: https://www.deloitte.com/us/en/insights/industry/public-sector-services/ai-readiness-in-government.html

[29] Brynjolfsson, E. and McAfee, A. (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W. W. Norton & Company. Available at: https://wwnorton.com/books/9780393254297.

[30] Mayer, H., Yee, L., Chui, M. & Roberts, R. (2025). *Superagency in the Workplace: Empowering People to Unlock AI's Full Potential*. McKinsey & Company. Available at: https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/superagency%20in%20the%20workplace-empowering-people-to-unlock-ais-full-potential-v.

[31] Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books. Available at: https://www.profilebooks.com/work/the-age-of-surveillance-capitalism/.

[32] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Medaglia, R., Le Meunier-FitzHugh, K., Le Meunier-FitzHugh, L. C., Misra, S., Mogaji, E., Sharma, S. K., Singh, J. B., Raghavan, V., Raman, R., Rana, N. P., Samothrakis, S., Spencer, J., Tamilmani, K., Tubadji, A., Walton, P. & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. Available at: doi:10.1016/j.ijinfomgt.2019.08.002.

[33] Veale, M. and Borgesius, F. Z. (2021). Demystifying the draft EU Artificial Intelligence Act – analysing the good, the bad, and the unclear elements of the proposed approach, *Computer Law Review International*, 22(4), pp. 97–112. doi:10.9785/cri-2021-220402.

[34] European Commission (2024a). Shaping Europe's Digital Future: Artificial Intelligence in Europe. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.
[35] European Commission (2024b). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.
[36] Ratten, V. (2024). Artificial intelligence: Building a research agenda. *Entrepreneurial Business and Economics Review*, 12(1), 7–16. doi:10.15678/EBER.2024.120101.
[37] OECD (2022). The Impact of Artificial Intelligence on the Future of Work and Labour Markets. https://www.oecd.org/employment/ai-future-of-work.htm.
[38] Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A. and Oak, S., (2025). *Artificial Intelligence Index Report 2025*. Stanford University. Available at: https://hai.stanford.edu/ai-index/2025-ai-index-report.
[39] Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., Lotufo, J. B., Rome, A., Shi, A., Oak, S., 2025. *The AI Index 2025 Annual Report*. Stanford: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. Available at: https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf.
[40] Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. doi:10.1038/s42256-019-0088-2.
[41] Barocas, S. and Selbst, A. D. (2016). Big data's disparate impact, *California Law Review*, 104 (3), pp. 671–732. Available at: doi:10.15779/Z38BG31.
[42] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). doi:10.1177/2053951716679679.
[43] Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier', NAACL/ACL Proceedings (Demonstrations). PDF available at: https://aclanthology.org/N16-3020.pdf.
[44] European Commission (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.
[45] Wachter, S., Mittelstadt, B. and Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation, *International Data Privacy Law*, 7(2), pp. 76–99. Available at: doi:10.1093/idpl/ipx005.
[46] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G. and Yu, H. (2016). Accountable algorithms, *University of Pennsylvania Law Review*, 165. Fordham Law Legal Studies Research Paper No. 2765268. Available at: https://ssrn.com/abstract=2765268.
[47] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G.C., Steinhardt, J., Flynn, C., Ó Éigeartaigh, S., Beard, S. J., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R. and Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv*. Available at: doi:10.48550/arXiv.1802.07228.
[48] Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022). Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure, *Internet Research*, 32(2), pp. 496–517. Available at: doi:10.1108/INTR-08-2020-0460.

11

MANAGEMENT OF ORGANIZATIONAL CHANGES DURING AI ADOPTION

Amy KW Lee

Purpose of the Chapter

For a business to thrive in today's rapidly evolving global landscape, it must be adaptable and agile. However, in the era of the 4.0 digital revolution, digitalization and artificial intelligence (AI) have become transformative forces in shaping industries, redefining business models, and revolutionizing decision-making processes. The extent to which AI is integrated into an organization's operations can impact its productivity, efficiency, operational excellence, and overall competitive advantage. AI is no longer a futuristic concept; it is a reality.

Will AI replace a human role or job function? This common yet flawed perception is incorrect because it instils fear, uncertainty, anxiety, and resistance in employees. Instead, the management must design a strategy based on an integrated human-centric AI model that positions AI as a tool for augmenting human capabilities rather than replacing them. AI can drive productivity, reduce costs by streamlining operations, and enhance decision-making when implemented correctly. This balanced relationship between advanced technology and the workforce will help many businesses succeed in the 4.0 digital revolution.

Using a range of well-researched frameworks and real-world case studies, this chapter explores how businesses can leverage AI and educate their workforce to adopt it, driving long-term growth, innovation, and creativity, while challenging the status quo by developing blue ocean thinking. The key to successfully utilizing AI lies in its technical capabilities and how leaders communicate, design, implement, and align their industry 4.0 AI strategy with their organization's vision, mission, goals, and people.

Introduction

In the current era of digitalization, artificial intelligence (AI) plays a pivotal role in many businesses, particularly in enhancing operational efficiency and excellence. This

DOI: 10.4324/9781003632429-16

This chapter has been made available under a CC-BY-NC-ND 4.0 license

rapidly advancing technology is poised to revolutionize the global economy, with a forecast estimating a positive contribution of around $15.7 trillion by 2030 (Fenwick, Molnar, & Frangos, 2024). It has, and certainly will, transformed how business is conducted, from making informed decisions to utilizing AI as an effective tool to increase productivity and efficiency. Among its most critical applications, AI is increasingly shaping complex business functions such as supply chain management, where operational efficiency and customer value are directly linked.

By definition, AI is classified with various functions and applications such as speech recognition, AI-related consulting and system integration services, chatbots, machine and deep learning, intelligent applications, virtual personal assistant (VPA), computer vision, insight engines, data-analytics and labelling, automated machine learning (AutoML), augmented intelligence, field-programmable gate array (FPGA) accelerators, robotic process automation software, and graphic processing unit (GPU) accelerators (Sicular et al., 2019).

When an organization changes its operating model from the traditional delivery of products and services to one that is operated on a software-automated, algorithm-driven digital platform, it must put in new roles such as AI-business translator and data scientist and invest in setting up AI factories of data pipelines, algorithm development, experimental platforms, and related software architectures (–ngstr–m et. al., 2023). The roles of management may also change as they transition from decision-makers to portfolio curators of AI algorithms and data flows; hence, they must be trained to possess both AI-relevant competencies and domain-specific business skills.

One of the most significant areas where AI integration is reshaping organizational operations is the supply chain. For example, the supply chain in a manufacturing organization is a complex operation that involves production and distribution channels, spanning from procurement to suppliers, internal operations, and sales channels, ultimately creating value for consumers. The primary objective of an effective supply chain is to deliver solid customer value and increase profitability by managing the expectations of various stakeholders, such as enhancing responsiveness and agility. To achieve a high level of sophistication in supply chain solutions, the AI strategy must facilitate all major flows among the different functions of the supply chain, both within and between the organizations.

By leveraging digitalization and AI technologies, companies will gain greater visibility into their entire supply chain, enabling them to integrate different stakeholders and utilize assets to enhance their competitive edge (Helo & Hao, 2022). AI is transforming the supply chain in terms of speed of execution, scale of operations, and the challenges posed by increased complexity. It is not a cheap investment; few companies have received the expected returns. The failure rate of many AI projects is high, primarily due to human resistance and poor change management rather than the technology itself being advanced. Therefore, management and business decision-makers need to understand that organizational change surrounding AI adoption requires a solid integration of people, processes, and culture, rather than focusing solely on machines and AI models.

Adopting AI in the supply chain can be more complex than enhancing processes using traditional models such as Six Sigma and Lean operations. The nature of AI systems differs from traditional technology adoption, as they require organizations to

continually learn and adapt, creating challenges in integration and scalability (Cannas et al., 2024). Some of these challenges included technological complexity, data quality, and the need for skilled experts (Robertson et al., 2025). Reflecting these complexities, AI integration necessitates significant fundamental operational changes, affecting both the organization's social aspects (people and structure) and its technical aspects (tasks and technology).

Leavitt's Diamond Model (Leavitt, 1965) identifies the four interconnected dimensions (people, tasks, structure, and technology) that determine organizational change. For example, an IT manager will use their computer knowledge to implement a new accounting system that aligns with the company's strategic plan. This IT manager is accountable, based on the company's organizational structure, which governs stakeholder interactions through effective resource allocation and orchestration. The AI system presents more complex socio-technical challenges that cannot be solely addressed through organizational adjustments; therefore, management must explore additional enablers beyond the company, such as broader systemic influences and core functional elements (Robertson et al., 2025). There is a strong emphasis that companies should not only focus on the technical aspects of AI-driven technologies and machine learning initiatives, but also integrate strategic insights, operational improvements, ethics, governance, regulated policies, and organizational culture. The technologies used must align with the existing operations or facilitate operational changes that add value to the firm. For example, suppose a company wants to utilize AI to enhance its supply chain efficiency. In that case, the AI system must be compatible with the existing supply chain management software and logistical processes (Abonamah & Abdelhamid, 2024).

Aligning business goals and outcomes with the implementation of AI and machine learning initiatives that involve stakeholders from both technical and business units is crucial to the overall success of the organization. Once they are defined and established, they should be measured using specifically designed key performance indicators (KPIs) to evaluate the success of these technological initiatives. This approach will help ensure that the technologies used connect different parts of the organization, making the whole AI-driven implementation a seamless process (Abonamah & Abdelhamid, 2024). Adopting AI and integration requires understanding the company's internal and external factors by mapping socio-technical and organizational-level factors. The purpose is to demonstrate how a modification in one component triggers corresponding adjustments in another area. These dynamics set the stage for a deeper exploration of how AI reshapes inventory management, optimizes operations, and creates new challenges and opportunities for staff and organizational strategy.

Understanding AI-Digital Transformation

AI is transforming the way businesses operate, including the creation of systems that result in process efficiencies and advanced data analysis, ultimately contributing to the value creation process of the organization. The AI technologies differ from the traditional technologies because they are machines capable of performing cognitive tasks that humans would otherwise complete. It possesses a remarkable ability to learn from a large dataset, enabling it to produce informed decisions and has become a digital agency (Climent, Haftor, & Staniewski, 2024). Unlike conventional digital

technologies, which follow programmable instructions from a human programmer, AI-enabled systems are given a problem or task and an algorithm to identify the patterns in the data sets. The transformation of prior instructions written by a human programmer into pattern identification in the data has revolutionized digitalization to a new level (Climent, Haftor, & Staniewski, 2024).

The rapid rise of AI technologies, especially generative AI tools, has caused a significant impact on the way executives design and implement their business strategies. They are now required to strategize and execute simultaneously by integrating AI and internet technologies into their strategy processes and prioritizing value-added activities that deliver business results (Prasad Agarwal, 2024). Some senior executives have abandoned the traditional use of industry analysis and the pursuit of unique resources. Instead, they simultaneously design a strategy that involves both decision-making and action-taking, while still aligning with the company's vision, mission, and goals. They now challenge conventional industry and business boundaries, opening organizations to a broader range of stakeholders, user communities, and social media platforms (Ha, 2024). By employing a trial-and-error methodology, their new strategy design incorporates more data, advanced communication technologies, sophisticated control mechanisms, effective governance, and continuous improvement measures. This new approach, which utilizes deep learning models with a transformer architecture and pre-training techniques, has the potential to enhance generative AI in decision-making systems, leading to improved strategy design in rapidly changing technological environments (Prasad Agarwal, 2024).

Companies have been investing heavily in AI to achieve their primary goal of creating value for stakeholders, particularly customers. For example, Apple is utilizing AI to enhance mobile phone security by integrating it into face recognition for user login (Climent, Haftor, & Staniewski, 2024). Many other firms utilize AI to deliver business and commercial benefits, such as cost reduction through increased productivity and the provision of high-quality products and services. For example, Amazon has scaled back its AI-enabled recruitment process (Jeans, 2020) because it has struggled to create value from its AI investments.

New market entrants can be disruptive as they introduce new business models, operational processes, and innovative technologies. To survive and remain competitive, companies must become leaner in achieving operational excellence, be more agile in their business approaches, and innovate their value chain by using new technologies such as AI (Oosthuizen et al., 2021). The new age of digitalization has increased the complexity of the retail value chain, as customers are becoming increasingly connected. Their service expectations have increased due to the transformation from retail to omnichannel retailing. For example, entrants like Amazon and Alibaba have pressured traditional retailers, such as Toys "R" Us, to perform more productively with faster service and higher-quality deliveries. Additionally, AI has enabled these digital companies to expand into unrelated industries, such as the global shipping value chain (Verhoef, Broekhuizen, & Bart, 2019). The new digital technologies are transforming the retail value chain in a complex and evolving manner.

Companies using the traditional models are facing the following risks (Oosthuizen et al., 2021) shown in Table 11.1.

Data management is crucial in AI implementation and operation, and translating data into knowledge is key to an organization's success. Therefore, it is recommended

TABLE 11.1 Risk assessment of digital value chain with mitigation strategies

Areas of the Value Chain	*Risks Management*	*Mitigation Strategies*
Stakeholder Management	Each stage of the value chain adds value, and the increased complexity of digitalization also increases the number of stakeholders and structures involved. Example: A global automotive firm has suppliers across the Asia Pacific and Europe, each using different procurement platforms. This results in mis-alignment in delivery.	Establish a centralized governance framework by using digital collaboration platforms. Improve the stakeholders' management by appointing stakeholder relationship managers and having regular alignment meetings.
Customer Services	The complex value chain can confuse internal operations and segments, leading to delayed customer responses and reduced customer value. Example: miscommunication between internal teams, resulting in delays in customer responses.	Implement an effective CRM system (e.g. Salesforce) to manage the communication channel between the internal teams and customers. Provide regular training to the front-line staff, especially in the use of AI-driven chatbots.
System Integration and Data Management	There is a tendency for each stakeholder to use their own data platform and software, which creates risks in system integration and data management. Management needs to be able to analyze large volumes of data to gain a competitive advantage, as data-driven decisions are crucial to a company's success. Example: In healthcare, hospitals use different electronic medical record (EMR) systems, which prevent seamless data sharing and risk patient safety.	Establish a data governance committee to design and implement a master data management system by investing in a cloud-based integration platform.
Cost Management	The high cost of maintaining a highly complex value chain may be passed on to consumers through more expensive products and services, thereby decreasing customer value. Example: A global firm selling luxury products has incurred high costs by coordinating global suppliers, esulting in higher retail prices when compared to online disruptors.	Apply operational improvement frameworks such as Lean Six Sigma and Kaizen to streamline operations and implement zero-based budgeting for value chain activities.
Vulnerability	An overly complex value chain may decrease the company's competitiveness against the more agile digital disruptors who can leverage new technologies for scaling and cost-cutting. Example: Traditional banks face disruption by fintech firms, which scale faster and cheaper using digital-first models.	Invest in digitalization such as AI, block-chain, and cloud computing. Regular benchmarking against digital disruptors and forming strategic partnerships with fintech startups and tech vendors.

Source: Oosthuizen et al. (2021)

that the knowledge management framework be integrated with AI technologies within the value chain, enabling the system to manage, share, utilize, create, and process information effectively. These are the building blocks of AI, which have further developed AI into deep learning applications, intelligent applications, and the creation of insight engines. While many businesses recognize the value of data analysis in enhancing competitiveness, they often struggle to derive meaningful insights from the data. This is largely due to the limitations of using big data, which require powerful technologies, substantial computer processing power, specialized technical personnel, and effective "what-if" models to analyze. The data management system can be enhanced with advanced AI technologies, providing more insightful information, anticipating future customers' expectations and needs, and enabling organizations to develop products that enhance customer value. Hence, expanding the knowledge framework using AI technologies is a key motivator for investing more in this technology and its advancement (Huang, Sun, & Fan, 2025).

To effectively utilize AI technologies in their global value chain, responsible management must allocate a specific budget for integration initiatives and implementation processes, such as establishing a robust IT infrastructure to support AI applications and actively recruiting and empowering a pool of technical experts. Therefore, AI can optimize collaborative decision-making, innovate cooperative models, and enhance collaboration efficiency, thereby increasing global competitiveness. To enhance collaboration on a global scale, firms are encouraged to engage in strategic dialogues with their international cooperative networks and participate in joint research and development projects by creating a platform to develop an international talent pool and establish multinational think tanks. (Huang et al., 2025)



FIGURE 11.1 AI-enabled value chain
Source: Huang et al. (2025)

Customer Engagement

One of the primary goals of the supply chain is to create value for the consumers. According to Xu et al. (2020), the definition of AI in the context of customer service is "a technology-enabled system for evaluating real-time service scenarios using data collected from digital and/or physical sources to provide personalized recommendations, alternatives, and solutions to customers' enquiries or problems." To succeed, organizations must deliver value that exceeds customer expectations to thrive in a highly competitive market. AI technologies can enhance customer engagement by improving customer service, strengthening sales support functions, and fostering trustworthy relationships through personalized services. For example, utilizing personalized chatbots to address customers' queries is a common approach to reduce service costs and improve customer response times (Reddy, 2017). Another example is The North Face, a retailer of outerwear, which utilizes augmented intelligence to assist customers in finding the right clothing for specific weather conditions (Trotter, 2018).

Inventory Management

Inventory management is integral to the supply chain, as raw materials are transformed into finished products through a well-organized and collaborative process. Additionally, the retailer must understand the customers' needs so that they can purchase the right resources and offer saleable products. AI technologies, such as chatbots, insight engines, intelligent applications, machine learning, and virtual assistants, can help these retailers gain customers' insights by forecasting future demand and reconciling that with the firm's operational capabilities and inventory levels. It can help to reduce working capital by creating a localized assortment matrix.

AI technologies can improve the internal processes of the supply chain by working and supporting the workers and optimizing the following three business processes (Helo & Hao, 2022):

1. Monitoring – Real-time tracking is possible by connecting equipment, products and transportation (i.e., trucks) with Internet of Things (IoT) sensors.
2. Analyzing – Providing sophisticated information to management for better decision-making is achieved by collecting more data for advanced analysis.
3. Acting – Management can be more responsive to stakeholders' needs by having greater visibility into their supply chain and taking reasonable actions.

AI technologies can further enhance traditional inventory management by streamlining operations, such as increasing the accuracy of demand prediction, ensuring popular items are consistently stocked on shelves, and utilizing clustering technologies to meet future customer requirements. For example, utilizing machine learning, deep learning, and intelligent applications enables matching demand to supply across multiple data sources (Bughin et al., 2017).

Predictive inventory management will help drive forecast accuracy and optimize the inventory throughout the value chain, creating more profitable value for retailers (Petropoulos et al., 2018). For example, using predictive machine learning

applications, the e-commerce retailer Otto Group has reduced its out-of-stock rate by 80%. The value created was increased profitability, achieved by being more responsive to market shifts and staying competitive (Trotter, 2018).

Operations Optimization

One effective way to increase customer value is to have an intelligent and effective business response system in place. The improvement is not limited to logistics and the value chain but rather encompasses the entire business flow. Inefficient operation will create bottlenecks within the value chain and slow production. Optimizing operations can increase a company's value by enhancing operational efficiency, improving agility, and boosting productivity throughout the retail value chain (Helo & Hao, 2022). These improvements are the driving force for AI-related capital investments.

AI technologies such as deep learning, intelligent applications, machine learning, robotics, process automation, and virtual assistants can shorten the value chain by improving production speed and enhancing inventory management from procurement to customers. For example, JD.com, China's largest retailer, delivers 92% of its orders on the same day or the next through an AI-enabled operational process (Trotter, 2018). Nike implemented augmented intelligence to design personalized shoes and deliver them within two weeks (Chao et al., 2019).

Generative AI (Gen AI) is being adopted by organizations like Toyota, Volkswagen, and Amazon Web Services for customer demand forecasting, inventory management, logistics planning, and real-time supply chain tracking. For example, Amazon Web Services has utilized Gen AI to enhance customer engagement by empowering businesses to anticipate and plan for market fluctuations.

Gen AI can generate complex data, such as images, music, or text, without human intervention, as its functionality is based on a machine learning algorithm known as generative adversarial networks (GANs). This is revolutionary because it has transformed how organizations operate and is highly adaptable, as it can learn and evolve to suit different circumstances and tasks. As this Gen AI utilizes the power of deep learning, machine learning, and neural networks to analyze vast amounts of complex data with accuracy, it is becoming a powerful tool for various industries, including healthcare, education, finance, and entertainment. As cloud computing and open-source tools become more accessible, Gen AI is becoming both affordable and scalable (Dubey, Gunasekaran, & Papadopoulos, 2024).

Strategic Challenges for AI Adoption

AI technologies are widely adopted to stay competitive, which means a growing demand for skilled workers who can implement and manage AI systems. For example, an AI-enabled retail value chain relies on high-quality data and trained employees; if data is sub-optimal, biased outcomes and increased risks may follow. While high-skilled workers are in demand, low-skilled staff will find it challenging to compete with machines unless upskilled (Frey & Osborne, 2017). Organizations must carefully examine the technological and organizational platforms required to implement an AI-enabled value chain effectively.

The main challenges to adopting AI technologies in supply chains include high capital costs, change management, technical difficulties, staff acceptance, and capability gaps (Hangl, Behrens, & Krause, 2022). Additional barriers include transparency issues, security and privacy risks, and poor data management. While benefits include efficiency, accuracy, customer satisfaction, and cost savings, staff resistance remains significant (Hangl, Behrens, & Krause, 2022).

Staff Resistance and Humanizing AI

One major reason for staff resistance to implementing AI technologies is the fear of job displacement. Workers often believe machines will replace them, although research shows that human–robot collaboration is most effective. This means the workers will not be replaced entirely, as the process has been enhanced with information and capabilities that the workers cannot handle. Low-skilled roles may decline, but more high-skilled jobs will emerge. Companies must therefore invest in training and upskilling, shifting job activities from tasks like data entry to roles such as enterprise architecture.

A successful AI implementation requires not only a comprehensive AI strategy within the organization but also full support from top management and clear communication at all levels within the firm. Strong leadership is necessary to address any misunderstandings and mistrust between retail employees and customers regarding AI technologies, particularly in issues such as AI adoption readiness, perceived power asymmetry, and the fear of the unknown. The radical framework of the business re-engineering model, which starts from scratch, can be a suitable AI strategic approach, especially when the organization lacks a strong technological infrastructure for data collection (Helo & Hao, 2022).

The fear of using AI cannot be underestimated, especially in long-established companies, where entrenched processes and traditional mindsets create resistance. Leadership in such firms is often conservative and risk-averse, shaped by founders and senior members who have weathered multiple business cycles. This makes building a balanced AI-powered workforce even more challenging.

1 AI Recruitment and Retention

To sustain AI adoption and accept the changes driven by AI, companies need to create an informed, interested, and engaged workforce that is willing to work continuously with AI-based process improvements and solution development. Hence, management must not only invest in the AI system but also in their workforce, who can grow and remain as innovative as the technologies evolve and improve. There is a critical need to motivate staff to learn about data and AI by encouraging continuous learning, developing an understanding of data, and exploring how AI can drive improvements, thereby creating a growth mindset (–ngstr–m et. al., 2023).

Recruiting and retaining highly experienced talent in the use of AI will help bridge the knowledge gap, ensuring businesses have the expertise required to integrate AI effectively. For example, Microsoft (2019) launched an AI Business School to help managers navigate organizational changes and learn how to incorporate AI into their strategies and operations.

2 AI Assistant Tools for the Organization

The aim of creating AI assistant tools for the workforce (i.e., on a similar principle to using a calculator) is to support the staff in improving their efficiency rather than making their roles redundant. GE adopted AI to optimize its industrial operations using the Predix platform. This technology required a cultural shift and substantial investment in data analytics, as well as re-skilling. (Chui, Manyika, & Miremadi, 2018) Similarly, Netflix utilized AI to refine its recommendation algorithms, thereby enhancing customer retention and increasing viewing time. (Gomez-Uribe & Hunt, 2015)

3 AI Training, Upskilling, and Using Effective Communication

The aim of conducting AI training and utilizing AI for more effective communication is to enhance the quality of the company's workforce, making it more creative, innovative, and adaptable to business changes. Adaptive capacity, defined as an individual's ability to respond to threats and act in the social and natural environment, is crucial for an organization to build its resilience and endurance in the face of transition (Farrow, 2021). The establishment of trust using AI tools and the willingness to change, depending on their relationships with the new AI system, depends on the mindsets of these workers. In the process of AI adaptation, the staff needs to be "digitized" as the organization is required to (Fenwick, Molnar & Frangos, 2024):

1. provides ai training so they know what it does and how it works;
2. retrain the staff to increase their skills and capabilities in working with AI;
3. creates a culture that fosters change and innovation in new ways of doing things.

HRM and Human-Centric AI

Many companies have recognized the importance of having an effective HRM (human resource management) system to support AI implementation, ensuring that adoption aligns with human values and organizational objectives (Fenwick, Molnar, & Frangos, 2024). An HRM system encompasses HR planning, including selection and recruitment, talent development, learning and development, compensation and rewards, performance management, employee relations, and payroll-related matters.

This traditional view of the HRM system needs to be revised by optimizing a human-centric AI through knowledge management, leadership initiatives, policies, regulations, and various business tools. Additionally, the HRM system must address the concerns associated with adopting AI, as well as the ethical issues and potential risks to employee well-being that may arise from the use of AI in business.

According to the proposed conceptual framework (known as Humanising AI, Figure 11.2) by Fenwick, Molnar & Frangos (2024), the focus of AI implementation is to have an HRM system that places the human component at the forefront, using a human-centric approach by identifying the key internal stakeholders and how they align with the overall organizational goals. This represents a paradigm shift from the traditional emphasis on the technical transformation of digitalization. Hence, the role

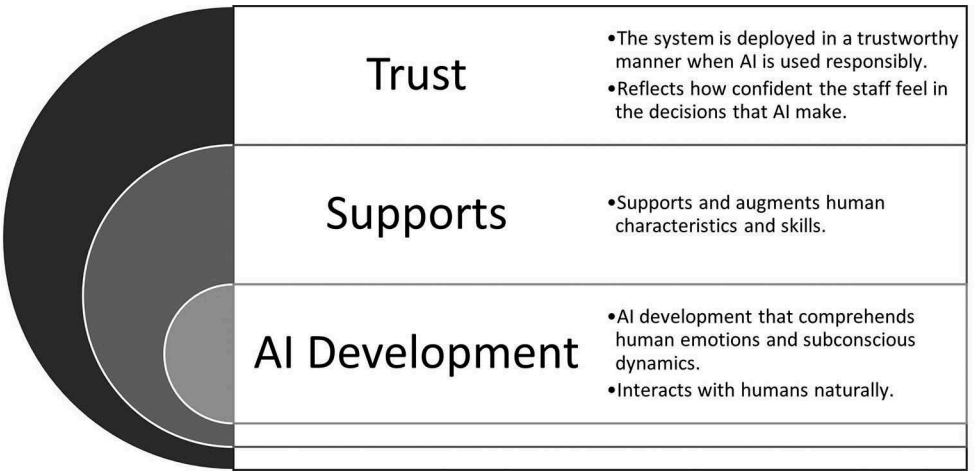


FIGURE 11.2 Humanizing AI framework
Source: Fenwick, Molnar, & Frangos (2024)

of HRM is more strategic, as it helps staff become more agile and adapt to the changes introduced by the implementation of AI.

In the context of humanizing AI, the goal focuses on utilizing AI to complement and augment human skills and decision-making processes, rather than replacing them entirely. This approach will enhance the staff experience with AI by fostering trust in the machines. By doing so, they can make AI more intuitive and empathetic, thereby increasing their skills and potential.

Implementation of AI-Driven Digital Transformation

A successful implementation of an AI-driven digital transformation extends beyond the mere technological ability, as it requires a holistic methodology involving strategic alignment, cross-functional collaboration between management and departments, and effective leadership (Abonamah & Abdelhamid, 2024). There must be an effective alignment between AI integration and machine learning technologies with the organizational strategy, priorities in resources, and business development. In Fenwick, Molnar, & Frangos–s (2024) research, they proposed five key determinants for a successful implementation of AI-driven digital transformation in a business. Each of these elements will be explored and evaluated, along with specific KPIs that organizations can use to measure AI implementation success.

1 Culture

An organization's culture is often described as its soul—the essence that embodies its vision, mission, and capacity to achieve goals. Consequently, an effective HRM system is vital in bridging the interaction between staff and machines, enabling the adoption of AI technologies through a shared understanding of norms, rituals, and implicit assumptions that define how the organization operates. The aim is to foster an

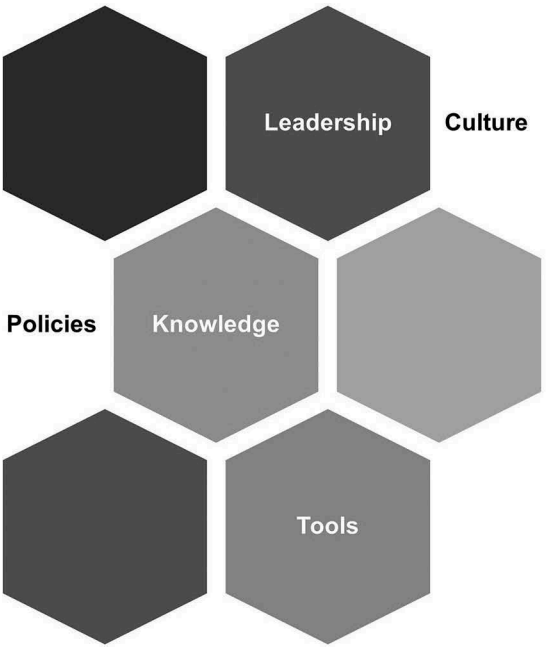


FIGURE 11.3 Five key elements for AI-driven digital transformation
Source: Fenwick, Molnar, & Frangos (2024)

innovative culture that is receptive to change. Hallmarks of such a culture during AI transformation include innovation drive, a learning orientation, transparent decision-making, and a risk appetite (Fenwick, Molnar, & Frangos, 2024). Ultimately, the successful integration of AI-driven technologies depends largely on employees' willingness to adapt these tools to improve performance (Abonamah & Abdelhamid, 2024).

Managers are central in cultivating this climate of trust and openness. Employees who lack trust in AI are significantly less inclined to use it, whereas those who can interpret AI-generated outcomes are nearly three times more likely to develop trust in its use (Ransbotham et al., 2022). To this end, HRM systems must be designed to nurture an AI-friendly ecosystem by shaping culture in ways that encourage human-centric collaboration with technology. Employees should feel psychologically safe when engaging with AI, supported by an environment that values experimentation, calculated risk-taking, and continuous learning. Mechanisms for regular monitoring and feedback should be in place to capture staff experiences and concerns, ensuring their voices are acknowledged and acted upon (Fenwick, Molnar, & Frangos, 2024).

KPIs for Culture: Employee adoption rates of AI tools, the percentage of employees trained in AI, employee satisfaction and trust scores (survey-based), the number of innovation initiatives launched, and the reduction in reported resistance-to-change incidents.

2 Leadership

A good culture is designed and driven by solid leadership, and different leadership styles will have varying impacts. Strong leadership in AI transformation will play a crucial role in how staff are open to change and how they adapt to these new technologies (Tabata et al., 2025). By realigning the company's vision and mission with an AI vision, the potential for using AI will be beneficial, as there will be a greater focus on innovation, increased awareness of risk management in terms of ethical concerns, and how to prioritize AI training and upskilling (Fenwick, Molnar, & Frangos, 2024).

A lack of digital and AI skills among executives can create barriers to setting a clear vision. To overcome this, leaders must serve as role models, upskill themselves, and communicate AI's productivity and strategic potential with clarity and transparency. Levi Strauss's global AI boot camps exemplify how leaders can champion organization-wide education that not only teaches technical use but also cultivates new mindsets. Employees emerge with a stronger sense of competence, autonomy, and connectivity—qualities that help scale cultural change globally. (Ransbotham et. al., 2022)

Leaders must also strike a balance between mandating early AI use to overcome resistance and preserving employee agency. Evidence shows that requiring use triples the likelihood of adoption, but adoption is more sustainable when employees can still override AI recommendations. By combining trust-building with encouragement of agency, leaders foster both compliance and ownership (Ransbotham et. al., 2022)

KPIs for Leadership: Number of leadership training sessions on AI completed, percentage of leaders certified in AI strategy, employee perception of leadership effectiveness in digital change (survey results), AI projects successfully sponsored and delivered on time, and reduction in leadership skill gaps identified in AI governance.

3 Knowledge

Organizational culture is shaped and sustained by strong leadership, and the choice of leadership style directly influences how employees respond to change and adapt to new technologies. Effective leadership in AI transformation is therefore pivotal, as it determines staff openness, adaptability, and confidence in integrating AI. By aligning the company's mission and vision with an AI-driven outlook, leaders can promote innovation, enhance awareness of ethical risk management, and ensure AI training and upskilling are prioritized (Fenwick, Molnar, & Frangos, 2024; Tabata et al., 2025).

However, limited digital and AI expertise among executives can hinder the articulation of a clear vision. Leaders must act as role models by upskilling themselves and communicating AI's productivity and strategic value with transparency. A practical example is Levi Strauss's global AI boot camps, where leadership initiatives extend beyond technical training to cultivate new mindsets. Employees emerge from such programs with stronger competence, autonomy, and connectivity—qualities essential for scaling cultural change across the organization (Ransbotham et al., 2022).

Leadership must also balance early adoption with employee autonomy. Research shows that mandatory initial use of AI tools can triple adoption rates, yet sustained integration is achieved when employees retain the ability to override AI

recommendations. This dual approach—building trust while encouraging agency—fosters compliance without eroding ownership (Ransbotham et al., 2022).

KPIs for Leadership: Number of AI-focused leadership training sessions completed, percentage of leaders certified in AI strategy, employee survey results on leadership effectiveness in digital transformation, successful sponsorship and on-time delivery of AI projects, and measurable reduction in leadership skill gaps related to AI governance.

4 Policies

Policies play a critical role in ensuring that AI adoption remains safe, ethical, and compliant. Organizations must adhere to industry standards, data protection regulations, and ethical frameworks from the outset to safeguard both staff and stakeholders. Well-designed policies not only provide legal and regulatory protection but also foster trust, thereby increasing employee willingness to engage with AI systems (Abonamah & Abdelhamid, 2024).

Beyond compliance, policies should actively promote transparency and contextual awareness. Employees need clarity on when AI is being used, how it functions, and how its outputs align with the broader organizational strategy. Research shows that workers who are consciously aware of using AI derive greater value and job satisfaction compared to those who remain unaware. Extending transparency to include these dimensions strengthens trust and reinforces alignment between AI use, corporate values, and individual contributions (Ransbotham et al., 2022). To maintain accountability, organizations must implement regular audits and monitoring mechanisms to evaluate AI's ethical and operational impacts. At the same time, policies should protect employee rights, privacy, and job security, striking a balance between accountability and innovation.

KPIs for Policies: Frequency of AI ethics audits conducted annually, number of AI-related compliance breaches identified, percentage of AI projects meeting ethical standards, rate of data privacy incidents, and employee survey scores on AI policy awareness.

5 Tools

The core objective of AI-driven digital transformation is to enhance productivity, support data-informed decision-making, and generate sustainable competitive advantage. Generative AI exemplifies this potential by equipping professionals with cognitive tools that extend their expertise, enabling the creation of text, images, and other outputs that amplify human capability (Prasad Agrawal, 2024).

Yet, the successful adoption of AI tools depends on more than technical deployment; it requires cultivating employee agency and alignment with organizational needs. While mandating initial use can help overcome resistance, employees must retain the ability to override AI-generated recommendations. This balanced approach—encouraging adoption while safeguarding autonomy—promotes trust and ensures long-term sustainability. Equally important, HRM must ensure that tool selection reflects organizational values and cultural context. Criteria such as ethical compliance, explainability, and usability should guide procurement and

implementation, allowing employees to view AI not as opaque "black boxes" but as transparent, reliable partners in decision-making (Ransbotham et al., 2022).

KPIs for Tools: AI system uptime and availability, reduction in manual processing hours, measurable cost savings from automation, improvements in accuracy and error reduction for AI-enabled tasks, and return on investment (ROI) derived from AI tool deployment.

AI-Enabled Key Performance Indicators

For AI implementation to succeed, organizations must go beyond traditional metrics and adopt AI-driven KPIs. Traditional KPIs have long benchmarked outcomes such as sales growth, customer satisfaction, and operational efficiency, supported by analytics and reporting systems. Yet, these measures have traditionally focused on what is being tracked rather than how performance is evaluated. In a rapidly changing business environment—characterized by economic volatility, rising consumer expectations, and digital transformation—this approach is no longer adequate. Managers are increasingly recognize the need for change, both in scope and measurement, is vital for sound decision-making. Executives further reinforce that enhancing KPIs with AI has become an immediate priority across industries.

AI-enhanced "Smart" KPIs represent a step change from human-driven, judgment-based approaches that provide only incremental insights. Leveraging advanced analytics, machine learning, and predictive capabilities. These AI-driven measures generate more precise, timely, and forward-looking performance insights. The business benefits extend well beyond efficiency gains: they strengthen organizational strategic alignment, enable proactive decision-making, and drive measurable financial improvements.

Many large organizations are investing in sophisticated algorithmic innovations to make their Smart KPIs more intelligent, adaptive and predictive so that they can improve their performance, generate more revenue and continue to expand their business. Hence, these innovation-driven KPIs powered by AI are now the sources rather than merely measures of strategic differentiation and value creation (Schrage et al., 2024). By giving more insight into the business and its operations, Smarter KPIs are leading to better outcomes. AI-enabled KPIs deliver advantages that outweigh costs, fostering resilience, competitiveness, and long-term growth while uncovering value sources and preventing risks from unexamined assumptions in traditional performance measures.

Making KPIs truly "smart" requires looking beneath the surface of traditional metrics and understanding the interplay of three distinct variables. First, there are strategic outcome variables—the overarching goals every organization recognizes, such as revenue, profit, or growth. Second, operational drivers—factors that organizations can directly influence, like pricing models, customer feedback, or digital traffic. Third, contextual factors—external forces that sit beyond management's control, such as consumer spending trends, global supply chain conditions, or government regulations.

The advantage of applying AI to this framework lies in its ability to uncover causal links between these variables. By connecting drivers and contextual influences to strategic outcomes, AI transforms KPIs into dynamic leading indicators rather than static measures. In practice, this can result in an improved version of an existing KPI, the

TABLE 11.2 Three types of smart KPIs

Types	*Description*	*Benefits*	*Example/Case Study*
Smart Descriptive KPIs	Integrate historical and real-time data to explain what has happened or is happening.	Enhance situational awareness, reveal performance gaps, and uncover root causes.	Sanofi's "snackable AI" tool identifies critical interdependencies across KPIs to improve operational insight.
Smart Predictive KPIs	Apply advanced analytics to anticipate future performance and generate leading indicators.	Provide foresight, detect hidden patterns beyond human capacity, and enable proactive decision-making.	GE uses AI to analyze order pipelines against its installed base, identifying growth opportunities and boosting revenue margins.
Smart Prescriptive KPIs	Use AI recommendations to optimize actions and strategies.	Highlight performance gaps, recommend targeted corrective measures, and align strategy with execution.	Sanofi leverages smart KPIs to link operations and sales, adjusting sales priorities based on supply chain performance.

Source: Schrage et al. (2024)

creation of an entirely new one, or explicit insights into cause-and-effect relationships across metrics. For instance, Tokopedia integrated operational data and external factors to develop a merchant-quality KPI that not only improved buyer-seller matching but also empowered merchants to refine their performance. This demonstrates how smart KPIs, powered by AI, extend beyond measurement to shape strategic success actively (Schrage et al., 2024).

Transforming Legacy KPI Practices

KPIs, once grounded almost entirely in human judgment and experience, are now being redefined by the disruptive force of algorithmic innovation. This shift is not evolutionary but revolutionary, eroding the traditional monopoly humans held over how KPIs are designed, managed, and governed. AI-driven approaches are challenging legacy models by demanding new levels of precision, adaptability, and insight. As a result, leadership teams must integrate AI into the KPI development and use, recognizing that "smart" KPIs are no longer optional.

Measuring a return on AI depends on understanding how different forms of AI use contribute to value creation at both the individual and organizational levels. At the individual level, employees derive value when AI strengthens their competence, autonomy, and relationships (Ransbotham et al., 2022). Workers need to feel effective and capable in their roles; they will not find value in technologies that diminish their confidence. AI enhances competence by providing decision support, surfacing relevant data, and synthesizing insights that individuals could not easily generate alone.

AI also empowers individuals by increasing autonomy. Rather than creating dependency, AI tools often reduce the need for direct oversight and routine managerial intervention. In addition, AI fosters stronger relationships. Self-determination theory highlights individuals' need to connect and care for others. AI tools can

TABLE 11.3 Traditional KPIs vs Smart KPIs

Traditional KPIs	*Smart KPIs*	*Explanation and Example*
Performance Tracking	Redefining Performance	Traditional KPIs focus on tracking targets and "keeping score," but investments in old metrics yield diminishing returns. AI-enabled KPIs redefine performance by uncovering unexpected drivers, much like new instruments revolutionize scientific discovery. Example: Google's unsupervised algorithms revealed surprising markers for media buys; Tokopedia created a merchant-quality KPI improving buyer–seller matching.
Static Benchmarks	Dynamic Predictors	Legacy "set-and-forget" benchmarks are replaced by dynamic, adaptive KPIs that anticipate markets, supply chains, and consumer behavior. Example: Schneider Electric models weather-driven trade-offs among energy costs, carbon footprints, and load-balancing risks; predictive KPIs are already central to algorithmic trading and programmatic advertising.
Judgement First	Algorithmically Defined Strategic Metrics	Historically, KPIs were designed by executives drawing on expertise and trends. Smart KPIs use AI, including generative models, to define innovative metrics previously unimagined. Organizations must ensure transparency, interpretability, and human oversight to align "machine-made" KPIs with values and objectives.
KPI Management	Smart KPI Governance and Oversight	Management optimizes outcomes for individual metrics, often in silos. Governance emphasizes enterprise-wide alignment, balancing trade-offs across KPI portfolios. Smart governance requires algorithms that refine KPIs themselves and "meta-KPIs" that evaluate the performance of KPIs as a system.
Monitoring KPIs	KPI Dialogues and Discussions	Checking dashboards passively is replaced by interactive, AI-powered dialogues. Generative AI enables managers to test counterfactuals, generate scenarios, and engage with KPIs in real time. Example: Sanofi's Plai initiative fosters future-oriented collaboration with predictive analytics.
Strategy with KPIs	Strategy for and with KPIs	Traditionally, KPIs acted as tools within a strategy. Smart KPIs require strategy themselves—designed for continuous optimization. The smartest KPIs both guide and reshape strategy, creating a symbiotic relationship between metrics and management decisions.

Source: Schrage et al. (2024)

support this by enabling employees to build better connections with coworkers, customers, and even the technology itself. Whether through more personal claims processing, faster pharmacy services, or improved collaboration on complex decisions, AI strengthens relational value, making work more connected and meaningful.

At the organizational level, these individual benefits align with a broader shift in how performance is measured. When traditional, static KPIs are transformed into AI-enabled "smart" KPIs, executives and boards gain more than performance snapshots—they gain predictive insight, agility, and foresight. Smart KPIs are not merely

about tracking performance but about redefining it: anticipating change, identifying opportunities, and embedding accountability directly into decision-making (Schrage et al., 2024). In this case, AI becomes both a personal enabler of competence, autonomy, and connection, and an organizational driver of strategic transformation—turning performance measurement from a passive record of outcomes into an active engine of value creation.

Conclusion: The Human Factor in AI Integration

Artificial intelligence has become a transformative force across industries, offering unprecedented opportunities to innovate, optimize operations, and generate value. However, the promise of AI will only be realized through thoughtful and strategic change management that places people—rather than technology—at the center. As explored in this chapter, the successful adoption of AI within supply chains and business value chains hinges not only on technical sophistication but also on cultural alignment, leadership commitment, and the creation of a resilient and adaptive workforce.

Businesses must resist the temptation to treat AI as a plug-and-play solution. Instead, they must reframe AI adoption as a process of co-evolution, where human intelligence and machine intelligence collaborate and grow in tandem. This requires organizations to develop dynamic, human-centric strategies that promote learning, reduce fear, and build trust in AI systems. Change must be approached holistically, where the technological, structural, behavioral, and social aspects of the organization are redesigned in harmony.

Upskilling, clear communication, ethical AI governance, and responsive leadership are not peripheral considerations; they are essential enablers of success. The value of AI lies not just in automating tasks but in enhancing human decision-making, creativity, and responsiveness to customers and markets. From predictive inventory management to personalized customer engagement, AI becomes a multiplier of human capability when embraced with vision and care (Xu et al., 2020).

Nevertheless, challenges remain. Fear of job displacement, cultural inertia, data privacy concerns, and uneven digital literacy continue to impede progress. To overcome these barriers, organizations must invest in AI literacy, adopt transparent policies, and redesign their human resource systems to ensure inclusivity, adaptability, and ethical resilience. AI does not replace people; it empowers them, provided the right scaffolding is in place.

The adoption of AI is not a technological event; it is an organizational journey. It challenges long-standing practices, disrupts traditional mindsets, and demands new forms of leadership. Yet, when change is managed well—with empathy, foresight, and strategic alignment—AI can unlock new avenues of growth, redefine operational excellence, and enable organizations not only to survive but to thrive in the era of intelligent transformation. The future belongs to those who are brave enough to integrate technology with humanity.

References

Abonamah, A. A., & Abdelhamid, N. (2024). Managerial insights for AI/ML implementation: A playbook for successful organizational integration. *Discover Artificial Intelligence*, 4(1), 22–34. doi:10.1007/s44163-023-00100-5.

Ångström, R. C., Björn, M., Dahlander, L., Mähring, M., & Wallin, M. W. (2023). Getting AI implementation right: Insights from a global survey. *California Management Review, 66*(1), 5–22. doi:10.1177/00081256231190430.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., & Trench, M. (2017). Artificial intelligence: The next digital frontier? McKinsey Global Institute. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/artificial-intelligence-the-next-digital-frontier (Accessed: 2025)

Cannas, V. G., Ciano, M. P., Saltalamacchia, M., & Secchi, R. (2024). Artificial intelligence in supply chain and operations management: A multiple case study research. *International Journal of Production Research*, 62(9), 3333–3360. doi:10.1080/00207543.2023.2232050.

Chao, G., Cheung, J., & Haller, K. (2019, January 14). The coming AI revolution in retail and consumer products: Intelligent automation is transforming both industries in unexpected ways. IBM Institute for Business Value / National Retail Federation. https://nrf.com/sites/default/files/2019-01/The%20coming%20AI%20revolution.pdf (Accessed: 2025).

Chui, M., Manyika, J., & Miremadi, M. (2018). How artificial intelligence can deliver real value to companies. McKinsey Quarterly. https://www.mckinsey.com (Accessed: 2025).

Climent, R. C., Haftor, D. M., & Staniewski, M. W. (2024). AI-enabled business models for competitive advantage. *Journal of Innovation & Knowledge*, 9(3), Article 100532. doi:10.1016/j.jik.2024.100532.

Dubey, R., Gunasekaran, A., & Papadopoulos, T. (2024). Benchmarking operations and supply chain management practices using generative AI: Towards a theoretical framework. *Transportation Research Part E: Logistics and Transportation Review*, 189, 103689. doi:10.1016/j.tre.2024.103689.

EY Australia. (2024). Investors confident in long-term return on AI-related investments [Press release]. https://www.ey.com (Accessed: 2025).

Farrow, E. (2021). Mindset matters: How mindset affects the ability of staff to anticipate and adapt to Artificial Intelligence (AI) future scenarios in organizational settings. *AI & Society*, 36(3), 895–909. doi:10.1007/s00146-020-01101-z.

Fenwick, A., Molnar, G., & Frangos, P. (2024). The critical role of HRM in AI-driven digital transformation: A paradigm shift to enable firms to move from AI implementation to human-centric adoption. *Discover Artificial Intelligence*, 4(1), 16–34. doi:10.1007/s44163-024-00125-4.

Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), 1–19. doi:10.1145/2843948.

Ha, L. T. (2024). Impacts of digital business on global value chain participation in European countries. *AI & Society*, 39(3), 1039–1064. doi:10.1007/s00146-022-01524-w.

Hangl, J., Behrens, V. J., & Krause, S. (2022). Barriers, drivers, and social considerations for AI adoption in supply chain management: A tertiary study. *Logistics*, 6(3), 63. doi:10.3390/logistics6030063.

Helo, P., & Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: An exploratory case study. *Production Planning & Control*, 33(16), 1573–1590. doi:10.1080/09537287.2021.1882690.

Huang, Y., Sun, X., & Fan, Q. (2025). AI orientation, global value chain collaboration, and the international performance of entrepreneurial firms: A technology affordance theory perspective. *Chinese Management Studies*. Advance online publication. doi:10.1108/CMS-10-2024-0780.

Jeans, D. (2020, October 20). Companies will spend $50 billion on artificial intelligence this year with little to show for it. Forbes. https://www.forbes.com/sites/davidjeans/2020/10/20/bcg-m

it-report-shows-companies-will-spend-50-billion-on-artificial-intelligence-with-few-results (Accessed: 2025).

Leavitt, H. J. (1965). Applied organizational change in industry: Structural, technological and humanistic approaches. In J. G. March (Ed.), *Handbook of organizations* (Vol. 20). Rand McNally.

Microsoft. (2019). Microsoft AI Business School. https://www.microsoft.com (Accessed: 2025).

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100–1). doi:10.6028/NIST.AI.100-1.

Oosthuizen, K., Botha, E., Robertson, J., & Montecchi, M. (2021). Artificial intelligence in retail: The AI-enabled value chain. Australasian Marketing Journal, 29(3), 264–273. doi:10.1016/j.ausmj.2020.07.007.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental selection of forecasting models. *Journal of Operations Management*, 60, 34–46. doi:10.1002/joom.1011.

Prasad Agrawal, K. (2024). Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems*, 64(1), 1–16. doi:10.1080/08874417.2023.2240744.

Ransbotham, S., Kiron, D., Candelon, F., Khodabandeh, S., & Chu, M. (2022, October 31). Achieving individual—and organizational—value with AI. MIT Sloan Management Review. https://sloanreview.mit.edu (Accessed: 2025).

Reddy, T. (2017, October). How chatbots can help reduce customer service costs by 30%. IBM Blogs. https://www.ibm.com/blogs/watson/2017/10/how-chatbots-reduce-customer-service-costs-by-30-percent/ (Accessed: 2025).

Robertson, J., Botha, E., Oosthuizen, K., & Montecchi, M. (2025). Managing change when integrating artificial intelligence (AI) into the retail value chain: The AI implementation compass. *Journal of Business Research, 189*, 115198. doi:10.1016/j.jbusres.2025.115198.

Schrage, M., Kiron, D., Candelon, F., Khodabandeh, S., & Chu, M. (2024). The future of strategic measurement: Enhancing KPIs with AI . *MIT Sloan Management Review*. https://sloanreview.mit.edu (Accessed: 2025).

Sicular, S., Hare, J., & Brant, K. (2019, July 25). Hype cycle for artificial intelligence, 2019 (G00369840). Gartner. https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/.

Tabata, M., Wildermuth, C., Bottomley, K., & Jenkins, D. (2025). Generative AI integration in leadership practice: Foundations, challenges, and opportunities. *Journal of Leadership Studies*, 18(4), 41–54. doi:10.1002/jls.70005.

Trotter, C. (2018, April 12). 50 best AI retail applications. *Insider Trends*. https://www.insider-trends.com/50-best-ai-retail-applications (Accessed: 2025).

Verhoef, P. C., Broekhuizen, T., & Bart, Y. (2019). Digital transformation: A multi-disciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. doi:10.1016/j.jbusres.2019.09.022.

Xu, Y., Shieh, C. H., van Esch, P., & Ling, I. L. (2020). AI customer service: Task complexity, problem-solving ability, and usage intention. *Australasian Marketing Journal*, 28(4), 189–199. doi:10.1016/j.ausmj.2020.03.005.

12

BRAND PURPOSE IN ETHICAL ALGORITHM DESIGN AND IMPLEMENTATION

Hospitality Industry Cases

Francesco Derchi and Roberto Mario De Stefano

1 Introduction

Recent environmental challenges, political turmoil, and manifold social inequalities have created a scenario of profound economic uncertainty (George et al., 2022). However, many companies have reacted to these challenges by implementing a "Brand Purpose," which is a guide for aligning a company's values, identity, and actions with its image and reputation (Fernandes, et al., 2024). Therefore, the topic of Brand Purpose has gained enormous momentum and relevance in both academic and professional circles (Ignatius, 2019). In this regard, companies are required to identify their proposition not only in terms of distinctive offerings in the market, but also to fulfill a social function by pursuing their purpose, which goes (also) beyond the achievement of profit (Williams et al., 2022, p. 703).

Recent studies (Carah, 2017; Srinivasan et al., 2021) have addressed the relationship between brands and algorithmic decisions, highlighting the role of algorithms in value creation (Kozinets, 2022), in pricing strategies (Nunan & Di Domenico, 2022) as well as in leading business model innovation (Boswijk, 2017). Despite the growing importance of algorithms in corporate decision-making processes, the area of study between Brand Purpose and algorithms remains largely unexplored. While several scholars (e.g., K–chling & Wehner, 2020; Lee & Shin, 2020) have already examined the ethical implications of algorithms, it remains unclear how a company's stated purpose can effectively guide the design, implementation, and control of these systems. At the same time, other recently published research (e.g., Williams et al., 2022; Blocker et al., 2024) highlights how Brand Purpose is now a strategic lever that drives companies beyond simple profit logic. However, it's still unexplored ground to understand how this "purpose" then turns into concrete technical and organizational choices, especially when it comes to configuring algorithms. Specific cases, such as Airbnb (Zhang et al., 2021), show that companies can intervene to modify algorithmic logic when automatic results clash with brand values. The question we ask ourselves, however, is whether these are isolated actions or whether they represent a true systemic model of purpose-driven governance.

DOI: 10.4324/9781003632429-17
This chapter has been made available under a CC-BY-NC-ND 4.0 license

Our analysis has highlighted some critical areas that deserve further investigation.

Firstly, there is a lack of analysis on whether and how brand purpose can serve as a reliable guide for the ethical adoption of algorithms, especially in certain sectors, such as tourism and hospitality, where customer relations are central. In fact, according to Ariffin and Maghzi (2012) in hospitality, the relationship between host and guest is crucial. Taking this perspective, if the organization is genuinely customer-oriented (Lamberti, 2013), or customer-centric (Inversini et al., 2020), it will build effective relationships with them to the point of creating positive behavioral responses (Gallarza and Gil Saura, 2006), such as favorable reviews. In this regard, Inversini et al. (2020) have shown that customer-focused accommodation facilities receive higher scores in user reviews.

Secondly, there is a gap in how companies translate their purpose into algorithmic design in practice, and what tensions arise between operational efficiency and value consistency. In particular there is also a lack of information on how Brand Purpose can influence the ethical design and implementation of algorithms, focusing on aspects such as dynamic pricing that have a direct impact on the customer experience.

Therefore, the two research questions that lead our investigation are:

RQ1: "How does brand purpose affect ethical design and the implementation of algorithmic systems in the hospitality industry?"
RQ2: "How is Brand Purpose actually translated into algorithmic design, and what tensions arise between operational effectiveness and consistency with stated values?"

To address these questions, we performed a multiple-case study analysis, concentrating our endeavors on two emblematic companies in the hospitality sector, Airbnb and Barceló Group, to investigate whether and how organizational purpose is integrated into business practice through the use of algorithms.

This essay contributes to the literature in several ways. First, on a theoretical level, it deepens our understanding of how organizational purpose is not just a theoretical statement but a real, practical, and concrete application within a company. Secondly, this research emphasizes the role of purpose in algorithm design. Thirdly, we expand on previous research on brand purpose (Blocker et al., 2024; Gray et al., 2024), showing how purpose guides technical decision-making in technological contexts, both in digital native and non-digital native companies.

Fourth, we adopt a classification that distinguishes efficiency-oriented top-down models from community-sensitive bottom-up models in the integration of purpose and algorithm, explaining variations in how organizations translate values into technological practice.

The chapter is structured as follows: Section 2 presents the theoretical background behind algorithmic decision-making and Brand Purpose literature. Section 3 describes the methodological approach used, the case selection criteria, and the data collection strategy. Section 4 presents the results of the empirical analysis. Section 5 emphasizes the differences between the two cases under analysis. A conclusion with theoretical and managerial implications is covered in Section 6.

2 Theoretical Background

2.1 Algorithms in Management

Algorithms as Decision-Making Structures: The Problem of Dehumanization

Algorithms in a managerial context can be seen as actual decision-making architectures (Floridi, 2018a, 2018b; Shrestha et al., 2019), which have the power to influence numerous aspects within an organization. For this reason, their use is crucial in processes such as hiring, dynamic pricing, and personalizing offers. However, it is important to note that these systems are not neutral as asserted by Floridi (2013). In this regard, authors such as Kaplan and Haenlein (2019) warn against the idea that algorithms are unbiased. The power of those who design them, although often hidden, is fundamental: programming choices and optimization objectives can introduce systemic biases, discrimination, or adverse effects, regardless of whether they are intentional or not. This risk is particularly high in customer-facing industries, such as hospitality, where algorithmic decisions have a direct impact on people.

It is in this context that Schultz et al. (2025) introduced the concept of "algorithmic dehumanization": the tendency of systems to treat individuals as mere numbers or objects, ignoring their complexity, rights, and dignity. The automation of evaluation (e.g., in ratings or recommendations) risks reducing people to simple patterns, thus ignoring fundamental qualitative aspects.

Connections to our cases: This theoretical angle is notably prominent in our investigation. As we will demonstrate in the empirical analysis, Airbnb has taken measures to design some algorithms that have sometimes produced discriminating outcomes (e.g., price biases), countering the phenomenon of algorithmic dehumanization. In contrast, the Barceló Group, while striving to ensure customer satisfaction in its stated purpose, has not yet systematically integrated anti-dehumanization principles into its algorithmic governance. In its algorithmic management, its mechanisms remain anchored to efficiency.

In their model, the authors propose strategies to avoid such dehumanization: transparency, accountability, and, above all, alignment with the values declared by the organization. It is in this space that Brand Purpose can act as a "guarantor," creating trust (Williams, M., 2024). In fact, if well integrated, it can guide the design of algorithms towards more humane, inclusive, and consistent objectives, while increasing performance (Baumer, 2017). The evidence is provided by digitally native organizations, or exponential organizations (ExO)—companies that leverage exponential technologies and organizational strategies to achieve an impact disproportionate to their resources (Ismail et al., 2014). ExOs identify a specific massive transformative purpose (MTP), a statement that defines their identity, provides confidence and direction in navigating markets, and that influences internal and the external part of the organization, including its competitive strategy (Derchi, 2025; Prattico et al., 2025).

Algorithms and Relationships: The Loss of Control and Agency

Xu et al. (2024) introduce the notion of "aesthetic loss of control," showing how, in creative contexts, algorithms can alter the perceived authorship and agency. When

algorithms make decisions without transparency, customers may perceive manipulation, information asymmetry, or loss of autonomy. This can be particularly relevant and dangerous for an industry such as hospitality. The real challenge for companies in the sector is to design algorithms that are not only practical, but also relevant, consistent with the brand identity and value proposition, this is another important element that contributes to the creation of trust—a key factor for tourism and hotel services (Williams et al., 2022; Xu et al., 2024).

Algorithms in Business Models: Value and Governance

Robertson et al. (2025) propose an "AI Implementation Compass," which can also be adapted to traditional algorithms, to help companies assess the trade-offs between efficiency and social impact. Algorithmic adoption should take into account not only economic KPIs, but also reputational, regulatory, and cultural impacts. Filosa et al. (2025), studying B2B platforms, show that the logic of algorithms is intertwined with the strategic identity of companies: the algorithm is part of the value promise and must therefore be consistent with the stated purpose. Overall, with the term algorithmic management it is considered a strategic management tool in the digital economy (Cheng & Foley, 2019).

Moving into detail, from a more practical perspective, this outlook would suggest that organizations cannot manage algorithms as merely technical tools, but rather they must be integrated into strategy. On the contrary, algorithmic systems epitomize organizational values and directly affect how customers, employees, and communities perceive the brand. This will be demonstrated in our case studies, where Airbnb's algorithmic changes reflect strategic choices of inclusion, while Barceló's efficiency-focused algorithms still reveal a gap between stated values, oriented towards regeneration, and technical implementation.

2.2 Brand Purpose

Brand Purpose beyond Storytelling: A Catalyst for Strategic Transformation

It is within this context that the contributions linking Brand Purpose to the use of algorithms emerge. Brand Purpose is commonly understood as the underlying "why" of a brand's existence, emphasizing the creation of value that transcends mere profit (Biraghi et al., 2020) while also considering its broader social implications (Bartlett & Ghoshal, 1994; Hollensbe et al., 2014). Brand Purpose is a real catalyst that can profoundly influence management strategies (Calder, 2022; Fernandes et al., 2024), guiding the redefinition of the value chain, and promoting a more conscious and ethical adoption of algorithms (Mittelstadt et al., 2016). Blocker et al. (2024) point out that a purpose-oriented company cannot limit itself to stating values but must transform its design choices (including algorithms) to be consistent with those values. Purpose thus becomes a lens through which to filter technical decisions, including objectives, data sets, performance metrics, and decision thresholds. Williams et al. (2022) reinforces this from the consumer's perspective: the perceived alignment between brands, behavior, and algorithms strengthens trust and produces eudaimonic experiences, i.e.,

meaningful and rewarding ones. This is particularly relevant in tourism, where the experience is not only functional but also emotional and relational. Finally, Gray et al. (2024) provide a useful taxonomy: purpose-driving companies are the most consistent, as they integrate purpose into all dimensions of action, including technology. The concept of Brand Purpose has evolved from a simple identity statement to a strategic lever capable of transforming the entire business model (Blocker et al., 2024). According to Blocker et al. (2024), purpose is no longer just a communicative message, but a true "transformative logic" that guides business choices in a systemic way, impacting not only consumers but also governance, innovation, and the supply chain. The proposed framework distinguishes between communicated, pursued, and lived purpose, introducing the notion of "purpose orientation" as a form of organizational intentionality oriented toward the collective good.

Types of Purpose and Perceived Authenticity

As briefly explained in the previous paragraph, Gray et al. (2024) propose a taxonomy of companies based on the level of integration of purpose into corporate strategy, distinguishing four categories:

1. Purpose-washing: organizations that superficially declare purposes merely for marketing purposes without any real and authentic endorsement.
2. Purpose-aligning: organizations that align some operational practices with the declared goals, but do not have complete integration.
3. Social purpose-driven: companies where goals significantly influence decision-making. The commitment is substantial and concrete.
4. Social purpose orientation: organizations that act as agents of proactive change, systematically integrating purpose into all operational decisions.

The paper also highlights how the perceived authenticity of Purpose is a key predictor of trust and loyalty, which are crucial elements in algorithm-intensive contexts such as tourism and hospitality. Authentic alignment between stated purpose and actual practices is essential. "Woke washing" (superficial purpose claims) can harm brand credibility, especially among skeptical or highly involved consumers (Mirzaei et al., 2022).

Brand Purpose as a Guide for the Ethical Adoption of Algorithms

The adoption of new exponential technologies, including algorithms and AI, requires informed and ethical choices that need to be rooted in a purpose or "brand purpose." This is essential to ensure that implementation decisions are aligned with broader values and regulatory imperatives (Gama & Magistretti, 2023).

Below is a guide to fundamental values according to various authors.

Human values and trust: As briefly mentioned above, algorithms are not neutral. They are, in essence, mirror images, reflecting the values, assumptions, and biases present in the data and logic with which they are designed (Davenport et al., 2019) and in their developers themselves (Martin, 2019). This can be an impediment, a

concern, as discrimination or ineffective decisions may occur in the long term (Bruner, 2020).

However, having a clear and well-defined brand purpose can provide support to the company and the group of developers themselves, acting as an ethical guide to orient the algorithm toward values such as respect for freedom, equality, and human dignity (Schultz et al., 2025). In this way, it is easier to gain the trust of customers by meeting their moral and ethical expectations of companies (Robertson et al., 2025). Customer relations in highly relational sectors: In the hospitality industry, interactions with various technologies such as algorithms and AI can be viewed in a negative light or as somewhat unsettling. In particular, Granulo et al. (2021) have analyzed that people can perceive manipulated or dehumanized feelings (Puntoni et al., 2020). It is now essential for companies to pay attention to the opinions of their customers, who tend to be skeptical of technology, considering it unsuitable for tasks involving emotions and feelings (Reese, 2018).

However, the existence of a clear Brand Purpose that is truly aligned with corporate values can prove to be a guarantee for a variety of reasons:

- by guiding the development of AI that improves customer relationships by promoting the integration of "emotive" AI technologies;
- by promoting transparency;
- by supporting algorithmic governance, following the *Dehumanization Avoidance* model.

(Schultz et al., 2025)

Translating brand purpose into algorithmic design: Many companies struggle to integrate purpose into algorithmic design, acting through sporadic, isolated, or sometimes purely instrumental implementations (Bruner, 2020). The most common challenges in this context stem from the conflict between operational efficiency and values, and the difficulty of defining algorithmic objectives in line with non-financial interests (Shrestha et al., 2019). However, a purpose-driven approach involves explicitly defining social impact objectives and actively listening to customers at the heart of the design process (Filosa et al., 2025).

Brand purpose and algorithmic pricing: The issue of dynamic pricing is central to this debate. In the hospitality sector, dynamic pricing based on algorithmic choices has a direct and significant impact on customers' perceptions of fairness and autonomy (Huang and Rust, 2021; K–chling & Wehner, 2020). In this regard, the mainn challenges from an ethical perspective include:

- the risk of algorithmic discrimination, caused by the use of data that may contain implicit biases (Bruner, 2020; K–chling & Wehner, 2020);
- privacy violations and perceptions of manipulation, linked to the *paradoxical concept of privacy-personalization*, whereby customers want highly personalized services but are concerned about data sharing (Nunan & Di Domenico, 2022);
- compromised trust, especially when pricing criteria are opaque or perceived as unfair.*(Williams et al., 2022)*

In this scenario, a clear and consistent Brand Purpose can serve as a guide for ethical price design, promoting:

- transparency, through understandable explanations of price determination (Robertson et al., 2025);
- fairness, excluding the use of sensitive characteristics in predictive models, in line with the principles of dignity and justice (K–chling & Wehner, 2020);
- customer autonomy, offering customers the possibility to negotiate (Nunan & Di Domenico, 2022);
- the relational experience, valuing trust, empathy, and consistency with the values stated by the brand, going beyond profit maximization (Williams et al., 2022; Blocker et al., 2024).

3 Methodology

For this research, we employed a multiple case study strategy, a qualitative research methodology that helps explore a phenomenon within a specific context (Rashid et al., 2019), in order to seek out and emphasize different facets of the phenomenon (Baxter & Jack, 2008). Following Yin's guidelines (2013), we opted for purposive sampling, based on a number of criteria set in advance in order to conduct a study on non-randomly selected cases.

In Table 12.1, the list of criteria are outlined.

Both cases selected meet the five criteria outlined in the table above. Both have a clear purpose statement, adopt algorithms, and satisfy the industry and organizational requirements. Furthermore, both belong to the social purpose alignment category outlined by Gray et al. (2024). In detail of the last criterion, Barceló has articulated a clear regenerative purpose ("REGEN") that influences operational practices such as local sourcing and community involvement. Similarly, Airbnb exhibits the characteristics of socially purpose-driven organizations. The company's Massive Transformative Purpose, "Belong Anywhere," operates as an active governance principle that has triggered systematic changes to algorithmic systems when conflicts arise.

Primary data collection was carried out following Arsel guidelines (2017), adopting a semi-structured interview format for executives involved in strategic decision-making

TABLE 12.1 Sampling criteria

Number of Criteria	*Criteria in details*
Criteria 1	Cases must belong to the hospitality sector
Criteria 2	They must have a stated and applied brand purpose
Criteria 3	They must use algorithms
Criteria 4	Diversity of organizational a (digital native vs. traditional) and geographical context
Criteria 5	They must belong to one of the category explained by Gray et al. (2024)

Source: Authors' own elaboration, based on Yin (2013), Rashid et al. (2019), and Baxter & Jack (2008).

or corporate social responsibility. For the purposes of data collection, we conducted a total of four interviews, selecting two informants from each organization in order to obtain a comprehensive picture of our survey.

The interviews were conducted via Google Meet and Microsoft Teams and transcribed verbatim as suggested by Krippendorff's guidelines (2011). The interviews were tailored to investigate several recurring themes: (1) how brand purpose is understood and operationalized within each organization; (2) the role of algorithms in key business processes; (3) governance mechanisms for algorithmic decision-making; and (4) instances where brand purpose and algorithmic logic have come into tension or alignment.

We relied on triangulation of sources to ensure data richness and reliability (Yin, 2013). In this regard, the interviews were supplemented with an extensive review of gray literature, specifically:

- public statements made by executives during industry conferences, press releases, and blog articles;
- corporate sustainability reports and purpose statements;
- related press releases;
- blog articles.

The analysis of the interviews followed an iterative process: reading, coding, and interpretation (Salda–a, 2015). The authors conducted the coding process independently and then validated the information using researcher triangulation (Patton, 2002). Following is the explanation of the two case studies.

4 Results

Barceló Group Case

With a history spanning almost a century, the Barceló Group (more details in Table 12.2) began its journey in business as a transport company. Today, it is one of the leading Spanish companies in the hospitality sector, with around 300 hotels located around the world. The core of its business is no longer just that of a traditional hotel operator, but rather a provider of complete tourist experiences, with a growing focus on sustainability, particularly on generating a positive impact on local communities.

TABLE 12.2 General information about Barceló Group

Foundation	*1931*
Headquarter	Palma de Mallorca, Spain
Industry	Hospitality
Turnover 2024	€7,500 million
Net Profit/Loss 2024	€301, 8 million profit
Ebitda	> €660 million

Source: Authors' own elaboration, based on Barceló Group Annual Report (2024) and primary interview data.

The Barceló Group's brand purpose, "REGEN," is about promoting a positive, fair, and inclusive economy with the goal of becoming a leading player in tourism. The three pillars of this purpose are: "Glocal. Sustainable investment in local communities with global impact"; "Fair, Ethical. Our social role"; "Caring for the Planet – Caring for People."

The goal of this approach is not just to "sell" rooms but to create experiences focused on generating customer happiness. In this vein the informant affirmed: "Our purpose is not just to provide a room, but to create experiences that make people happier and leave a positive footprint on society and the environment." Or even the interviewer affirms: "We don't do campaigns to try to sell more for the day of the planet's environment. What we are doing, we do the job."

Another pillar is the valorization of the local community and local people, in this regard the informant affirmed: "We work with local suppliers, local chefs, local artists, because what matters is that the customer feels part of the community where the hotel is located."

Going into detail, for the company, the adoption of algorithms is a widespread and constantly expanding practice. Its use is pervasive, especially in the digital, marketing, revenue management, and customer intelligence departments. One of the areas most influenced by algorithms is pricing. AI systems support the analysis of demand, seasonality, and anomalous behavior. In this context the informant affirmed: "Today, around 70–80% of our pricing decisions are automated by algorithms. They help us anticipate demand and anomalies, while managers can focus on more strategic tasks." The same is true for website personalization and commercial offerings, where the use of data and algorithms is increasingly widespread. In this case, user behavior is tracked to suggest tailored offers, adapt navigation, and improve the likelihood of conversion. The group has also developed a system called "Marketing Hub" that integrates data on customers, hotel performance, and sales trends to automatically recommend different advertising campaigns depending on the property, season, or target audience. The company has also introduced innovations in customer service, digitizing its call centers and introducing intelligent chatbots and IVR to handle calls, automatically assign the best operators, and recognize customers by voice or digital history. Speech analysis technology allows satisfaction levels to be tracked, the most frequently used keywords to be identified, and the quality of interactions to be measured. Call flow forecasting and staff scheduling are now also based on predictive models. Looking ahead, the company is developing personalized upselling systems for ancillary services (such as excursions, transportation, or local experiences), although these algorithms are not yet active on a large scale. Despite the extensive use of artificial intelligence, there is currently no structural connection between algorithms and brand purpose. Regenerative principles guide operational and branding strategies, but they do not inform algorithm design or governance. Regenerative practices are not integrated into decision-making models, nor do they represent a variable in campaign optimization or customer segmentation. In summary, the Barceló Group's brand purpose is integrated with algorithms through an ex-ante approach, which prioritizes consistency of values over ex-post optimization.

In answering RQ1, brand purpose impacts the ethical design of algorithms mainly through a voluntary omission of regenerative aspects from their communication. This

is to avoid the perception that commitment to the planet is just a strategy to improve profits, avoiding greenwashing. Algorithms are not coded to be ethical upstream, but they are set up to avoid over-exploiting their values for commercial gain.

About RQ2, the translation of brand purpose into algorithmic design occurs in an indirect and emergent way. Brand purpose is not directly coded into algorithms. Instead, the company creates content and experiences based on its values, such as being "very local." The algorithm, in turn, observes what customers like best and what leads to the most conversions. If content that reflects the company's values performs well, the algorithm automatically promotes it more, demonstrating that the values are effective in the market. This creates a potential conflict between "selling more" and "staying true to your principles." The Barceló Group resolves this tension in a clear-cut way: corporate values win. The company has decided not to launch commercial campaigns based on sustainability, even if they could be profitable. This policy acts as an insurmountable barrier, preventing the algorithm from even testing a strategy that would go against the brand's ethics.

Airbnb

Airbnb is a publicly traded corporation operating as an online marketplace connecting private homes with travelers in search of local experiences. Scholars define it as market leader in peer-to-peer accommodation (Boswijk, 2017), while for the CEO Brian Chesky, Airbnb is "a new kind of economy powered by people" to underline the role of human engagement behind a database of over 8 million listings in more than 220 countries and regions (Airbnb, 2025). Airbnb operates properties and interactions inspired by its Massive Transformative Purpose: "Belong Anywhere" (Derchi, 2022). This evocative statement invites everyone to participate being a guest and traveler or a host, allowing global interactions and local exploration into one virtual and physical community of global connections (Oskam & Boswijk, 2015). This "sharing philosophy" (Guttentag, 2015) together with the powerful and real community is considered the great attractor and the secret of Airbnb's marketing success (Yannopoulou et al., 2013). However, the importance of the purpose for the firm is not just at a marketing level rather than about its competitive advantage. In our interviews we learned that the firm has been very intentional in the way to design and in the community values. An example for this initiative is the creation of a "community commitment" proposed to guests and hosts to be flagged at the end of the sign-up process; this makes explicit the

TABLE 12.3 General information about Airbnb

Foundation	*2007*
Headquarter	San Francisco, United States
Industry	Hospitality
Turnover 2024	$11.1 billion
Net Profit/Loss 2024	$2.648 billion
EBITDA	$4 billion (adjusted)

Source: Authors' own elaboration, based on Airbnb Annual Report (2024) and primary interview data.

intention for participating to "create a world where anyone can belong anywhere." The statement is straightforward: "I agree to treat everyone in the Airbnb community—regardless of their race, religion, national origin, ethnicity, disability, sex, gender identity, sexual orientation, or age—with respect, and without judgment or bias."

The purpose here is to clarify the intentions of the firm to connect the MTP with the customers in the deepest possible binding form.

Accordingly, all community members, when leveraging the platform, cannot discriminate based on criteria language, culture, gender, sexual orientation. The bold move came with the cost. Our interviewee told us that more than a million people declined to sign the proposal since it became operational around 2015, "and probably doubled by now." That is the number of potential customers that did not align with Airbnb's values, and when we discussed the impact for this move from a competitive standpoint, this was explained by the interviewee "we believe long term will have a positive effect…we believe to create positive growth."

Being a data-driven company, Airbnb has been leveraging algorithms in multiple areas of business. These can be easily distinguished between the ones dedicated to managing hosts and the supply side of the platform and the ones related to marketing and value creation for guests: the demand side. Among them, dynamic pricing is one of the main algorithms' applications for the firm (Hill, 2015). In this case, the analysis underlined racial disparity in daily revenue earned by Airbnb hosts, due to biases in the original algorithm model (Edelman et al., 2017).

On the supply side, algorithmic management is considered influencing hosts' practices, creating challenges, influencing results or frictions according to how they react to it (Cheng & Foley, 2019). Another area is related to managing interactions between hosts and potential guests in the negotiation phase. This is a crucial part of the value creation process for the firm because the interaction between people is considered the connector for allowing the sense of belonging. Here the algorithm plays a distinctive role connecting parts and allowing specific information to be visible, like the name of guests and the booking intentions and by blocking other information like phone numbers or autonomous e-mails to avoid communication between parts outside of the Airbnb messaging platform.

The case study investigated the firm's strategy, driven by the purpose "belong anywhere," which led to a structural change in its algorithm, affecting the revenues of the various hosts. Despite the relevance of such a topic (K–chling & Wehner, 2020), growing with the progressive implementation of algorithms in business contexts (Lee & Shin, 2020) it is still unclear whether this influence is an isolated phenomenon or represents a systematic pattern.

Finally, our interviews revealed another interesting case of internal processes where interfaces and algorithms were not aligned with the firm's purpose statement.

The Diversity Recruiting App is a tool used for recruitment purposes by the firm. It was introduced using the software Greenhouse for talent management. At the same time, inspired by the purpose, the Airbnb community raised the point about proving an assessment that could happen "without regards to race, color, religion, national origin, age, sex, marital status, ancestry, physical or mental disability, veteran status or sexual orientation."[1] In order to make recruitment choices, the entire process for information input has been shaped into two different steps: the one on Greenhouse with the collection of jobs-related information. A second flow has been managed

remotely. Candidates have been reached privately, by email, asking to submit a voluntary survey which decision would not affect the hiring process or thereafter, and all information would be kept in a confidential file.

Ultimately, this is in line with what emerged from the materials collected. At the World-Bélo event, in fact, Chesky was quoted in a presentation referring to the company being "not in the business of spaces rather than in the business of belonging." This statement is very important in clarifying the role of the MTP as crucial inspiration for algorithm development as well. According to this perspective in fact: "Houses do not create belonging; people do."

The Airbnb case responds to RQ1, showing that its MTP "Belong Anywhere" is not just a motto or marketing declaration but has ethical purposes. In this regard, the company has introduced a "community commitment" involving both hosts and guests, encouraging them to treat each other without any kind of discrimination. This has repercussions in the field of algorithms. In fact, once racial biases were detected in the platform, the brand purpose guided its modification in an ethical manner.

Airbnb's purpose is oriented toward design and algorithmic governance choices that emphasize and prioritize value consistency over economic efficiency. This responds to RQ2. There are many examples, such as in the management of host–guest interactions, where the algorithm obscures private information to maintain the relationship on the platform, preserving security and trust, or in the pricing strategy, where the company has faced tensions due to the need to optimize economically with the perception of having to balance these needs with a purpose oriented towards inclusion, where Airbnb wants to be aligned with its founding values.

Leveraging this perspective, we can read all the previously described cases and initiatives within a clear light: technology is seen as a powerful tool that allows Airbnb to maximize its impact in all the ecosystem areas, guided and inspired by its purpose (Boswijk, 2017).

5 Discussion: Difference between Cases

The qualitative analysis of two well-known cases in the hospitality sector reveals different ways in which *brand purpose* and algorithms relate to each other. In the Barceló case, the stated *purpose* is regenerative (REGEN) and it is reflected operationally in practices that benefit local communities. At the algorithmic level, the values expressed in the purpose are not yet fully utilized. They remain linked to predominantly market-driven logic such as efficiency, dynamic pricing, personalization, and marketing automation. This results in a structure in which purpose and technology coexist and interact indirectly, through what can be defined as a top-down approach, in which purpose is defined at a strategic level, while systems remain governed by traditional economic performance metrics. Airbnb introduces an alternative configuration, more oriented toward a community-centered and bottom-up approach. Airbnb's *brand purpose* (focused on inclusion, belonging, and trust in the community) serves as the basis for governance that can guide technological choices when critical issues arise. This happened, for instance, when the firm received criticism over racial bias of their booking engine algorithm (Time.com, 2016). In response to feedback and reports from the community, management has shown a willingness to modify the algorithmic design and

decision-making flows in sensitive areas, with the aim of reducing bias and improving perceived fairness. In other words, the purpose has a protective and safeguarding role, which allows for the redesign of algorithms. This difference is also reflected in the differences in organizational genesis. Airbnb is a digital-native platform and it has a habit and tendency to implement mechanisms for listening to and integrating feedback, which facilitate the alignment of *purpose and algorithms* "from bottom up." In contrast, a non-digital-native operator such as Barceló strongly values its relationship with local areas and communities in terms of experience and offering but has not yet implemented this orientation into constraints and project metrics within algorithmic architectures, which remain focused on economic efficiency. In both cases, customer centricity is crucial, but it operates according to different approaches. Airbnb's model depends on community feedback, which can trigger system revisions. Example for such revision is the implementation of an algorithm that obscures private information for maintaining the host-guest relationship on platform, while keeping security and trust. On the other end, Barceló's model is linked to the quality of the customer experience. At the algorithmic level, however, it risks being highly quantitative, with potential risks of dehumanization. In this case, the support of an integrated purpose becomes essential. From a theoretical point of view, the relationship between brand purpose and algorithms is constantly evolving in both cases. To maximize the potential of purpose, it would be desirable for both to be 100% embedded in the algorithms.

6 Conclusion

The analysis emphasizes how Brand Purpose could be pivotal in encouraging the ethical adoption of algorithms, particularly in a specific sector such as hospitality. The cases analyzed demonstrate different patterns. On the one hand, the Barceló Group, despite having defined a purpose oriented toward having a positive impact on people, finds it difficult to translate this into algorithms, which remain anchored to efficiency- and optimization-based logic. On the other hand, Airbnb integrates its founding purpose into technological governance and modifies algorithmic processes when they

TABLE 12.4 Comparative resolution of efficiency-values tensions

Tension Point	*Barceló Approach*	*Airbnb Approach*
Dynamic Pricing	Algorithms are set to optimize revenue; purpose is not considered in the pricing logic.	Fairness constraints added; accepts reduced predictive accuracy to prevent discrimination.
Marketing Optimization	Sustainability messages in conversion algorithms are not used to avoid greenwashing.	Actively incorporates belonging metrics; tests not just conversion but inclusive outcomes.
Customer Segmentation	Efficiency-oriented: segments based on spending potential and likelihood of booking.	Values-influenced: considers community fit, implements monitoring for discriminatory patterns.

Source: Authors' own elaboration, based on Airbnb and Barceló Annual Report (2024) and primary interviews data.

conflict with the values of inclusion and belonging. These findings reveal some important considerations. First, brand purpose can effectively orient the design and implementation of algorithms only if it is incorporated into the design criteria. Second, translating purpose into algorithmic architectures unavoidably generates tensions between the need for operational efficiency and the need to ensure value consistency. From a theoretical point of view, the main contribution of this study is to highlight the intersection between corporate purpose and algorithmic governance, a field of research that is still relatively unexplored but destined to grow in importance. From a managerial perspective, the evidence suggests that tourism companies should integrate purpose into algorithmic design processes, establishing metrics that go beyond economic indicators alone. Scholars consider the firm's purpose as a key element for managing organizations in the age of AI (Birkinshaw, 2019). The more technologies evolve and expand in pervasiveness, the more the role of purpose is considered key in attracting customers, staff on demand and new forms of leveraged assets (Derchi, 2022). However, despite some cases, typically among digital native organizations and ExOs (Boswijk, 2022), the relationship between corporate and brand purpose and artificial intelligence is just at its dawn. At the same time, the increasing distribution and vulgarization of algorithms and their application (Bogost, 2015) creates the risk of misalignment between application and impact (Martin, 2019). Here, purpose driven brands can be an example for more virtuous models (Grey et al., 2024). Differences in the way to design and deliver algorithms depend not only on the digital savviness but also on the presence of maturity in terms of sustainability and sense of purpose. Looking ahead, we can anticipate that firms with a deep sense of purpose will be able to raise stronger awareness regarding the importance of an algorithm design and management practice in coherence with the brand purpose. This will include firms to address, design, and correct some algorithms applications, orientating their business decisions and maximizing impact for more coherent decision-making.

Note

1 https://time.com/4484084/airbnb-is-overhauling-its-anti-discrimination-policy/

References

Ariffin, A. A. M., & Maghzi, A. (2012). A preliminary study on customer expectations of hotel hospitality: Influences of personal and hotel factors. *International Journal of Hospitality Management*, 31(1), 191–198. doi:10.1016/j.ijhm.2011.04.012.

Arsel, Z. (2017). Crafting qualitative data analysis. In *The SAGE handbook of qualitative research* (pp. 1092–1107). SAGE Publications.

Bartlett, C. A., & Ghoshal, S. (1994). Changing the role of top management: Beyond structure to processes. *Harvard Business Review*, 72(1), 86–96.

Baumer, E. (2017). Toward human-centered algorithm design. *Big Data & Society*, 4(2). doi:10.1177/2053951717718854.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13(4), 544–559.

Biraghi, S., Gambetti, R. C., Quigley, S., & Quigley, S. (2020). Brand purpose as a cultural entity between business and society. In D. S. Marques (Ed.), *Social entrepreneurship and*

corporate social responsibility. Management for professionals (pp. 401–422). Springer. doi:10.1007/978-3-030-39676-3_26.

Birkinshaw, J. (2019). What is the value of firms in an AI world? In *The future of management in an AI world: Redefining purpose and strategy in the fourth industrial revolution* (pp. 23–35). Springer International Publishing.

Blocker, C. P., Cannon, J. P., & Zhang, J. Z. (2024). Purpose orientation: An emerging theory transforming business for a better world. *Journal of the Academy of Marketing Science*, 53(2), 367–393. doi:10.1007/s11747-023-00989-5.

Bogost, I. (2015, January 15). The cathedral of computation. *The Atlantic*. https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/.

Boswijk, A. (2017). Transforming business models in the hospitality industry with algorithms. *Service Science*, 9(3), 233–239.

Boswijk, A. (2022). The power of the economy of experiences: New ways of value creation. In *Handbook on the experience economy* (pp. 171–176). Edward Elgar Publishing. doi:10.4337/9781781004227.

Bruner, C. M. (2020). Distributed ledgers, artificial intelligence and the purpose of the corporation. *The Cambridge Law Journal*, 79(3), 431–458. doi:10.1017/S0008197320000756.

Calder, B. J. (2022). Customer interaction strategy, brand purpose and brand communities. *Journal of Service Management*, 33(4–5),747–757. doi:10.1108/JOSM-11-2021-0410.

Carah, N. (2017). Algorithmic brands: A decade of brand experiments with mobile and social media. *New Media & Society*, 19(3), 384–400. doi:10.1177/1461444815605463.

Cheng, M., & Foley, C. (2019). Algorithmic management: The case of Airbnb. *International Journal of Hospitality Management*, 83, 33–36. doi:10.1016/j.ijhm.2019.04.009.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. doi:10.1007/s11747-019-00696-0.

Derchi, F. (2022). Tales from the disruption world. *Impresa Progetto: Electronic Journal of Management*, 1. doi:10.15167/1824-3576/IPEJM2022.1.1442.

Derchi, F. (2025). From buzzword to framework: A systematic review of the massive transformative purpose concept. *Administrative Sciences*, 15, Article x. doi:10.3390/admsci15120472.

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 9(2), 1–22. doi:10.1257/app.20160213.

Fernandes, T., Guzmán, F., & Mota, M. (2024). Purpose is the new branding: Understanding conscientious purpose-driven marketing and its impact on brand outcomes. *Journal of Product & Brand Management*, 33(6), 761–782. doi:10.1108/JPBM-08-2023-4667.

Filosa, C., Jovanovic, M., Agostini, L., & Nosella, A. (2025). Pivoting B2B platform business models: From platform experimentation to multi-platform integration to ecosystem envelopment. *International Journal of Production Economics*, 280, Article 109466. doi:10.1016/j.ijpe.2024.109466.

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Floridi, L. (2018a). Artificial intelligence, ethics, and Big Data. *Philosophy & Technology*, 31(1), 1–8.

Floridi, L. (2018b). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Gallarza, M. G., & Gil Saura, I. (2006). Value dimensions, perceived value, satisfaction and loyalty: An investigation of university students' travel behaviour. *Tourism Management*, 27(3), 437–452. doi:10.1016/j.tourman.2004.12.002.

Gama, F., & Magistretti, S. (2023). Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. *Journal of Product Innovation Management*. Advance online publication. doi:10.1111/jpim.12698.

George, G., Haas, M. R., McGahan, A. M., Schillebeeckx, S. J. D., & Tracey, P. (2022). Purpose in the for-profit firm: A review and framework for management research. *Journal of Management*. Advance online publication. doi:10.1177/01492063211006450.

Granulo, A., Fuchs, C., & Puntoni, S. (2021). Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts. *Journal of Consumer Psychology*, 31(1), 72–80. doi:10.1002/jcpy.1181.

Gray, H., Dolan, R., Wilkie, D. C., Conduit, J., & Burgess, A. (2024). Social purpose branding approaches: A typology of how brands engage with a social purpose. *European Journal of Marketing*. Advance online publication. doi:10.1108/EJM-08-2022-0588.

Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, 18(12), 1192–1217. doi:10.1080/13683500.2013.827159.

Hill, D. (2015, August 20). The secret of Airbnb's pricing algorithm. *IEEE Spectrum*. http://spectrum.ieee.org/computing/software/the-secret-of-airbnbs-pricing-algorithm.

Hollensbe, E., Wookey, C., Hickey, L., George, G., & Nichols, C. V. (2014). Organizations with purpose. *Academy of Management Journal*, 57(5), 1227–1234.

Huang, Y., Jiao, J., & Zhu, D. (2025). Empowering the traveler: The impact of algorithm transparency and customer control on customer experience. *Asia Pacific Journal of Tourism Research*, 1–19. Advance online publication. doi:10.1080/10941665.2025.2523339.

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. doi:10.1007/s11747-020-00749-9.

Ignatius, A. (2019, March–April). Profit and purpose. *Harvard Business Review*. https://hbr.org/2019/03/profit-and-purpose.

Inversini, A., De Carlo, M., & Masiero, L. (2020). The effects of customer-centricity in hospitality. *International Journal of Hospitality Management*, 86, Article 102436. doi:10.1016/j.ijhm.2019.102436.

Ismail, S., Malone, M., & Van Geest, Y. (2014). *Exponential organizations*. Diversion Books.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.

Köchling, A., & Wehner, M. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848. doi:10.1007/s41685-020-00134-w.

Kozinets, R. (2022). Algorithmic branding through platform assemblages: Core conceptions and research directions for a new era of marketing and service management. *Journal of Service Management*, 33(3), 437–452.

Krippendorff, K. (2011). *Content analysis: An introduction to its methodology*. SAGE Publications.

Lamberti, L. (2013). Customer centricity: The construct and the operational antecedents. *Journal of Strategic Marketing*, 21(7), 588–612. doi:10.1080/0965254X.2013.817476.

Lee, D., & Shin, H. (2020). Algorithmic bias in hospitality and tourism. *Tourism Economics*, 26(7), 1404–1423.

Martin, K. (2019). Artificial intelligence, business, and society: Are we ready for algorithmic responsibility? *Business Horizons*, 62(6), 761–769.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. doi:10.1177/2053951716679679.

Mirzaei, A., Wilkie, D. C., & Siuki, H. (2022). Woke brand activism authenticity or the lack of it. *Journal of Business Research*, 139, 1–12. doi:10.1016/j.jbusres.2021.09.044.

Nunan, D., & Di Domenico, M. L. (2022). Value creation in an algorithmic world: Towards an ethics of dynamic pricing. *Journal of Business Research*, 150, 451–460. doi:10.1016/j.jbusres.2022.06.032.

Oskam, J., & Boswijk, A. (2015). Airbnb: The future of networked hospitality businesses. *Journal of Tourism Futures*, 2(1), 22–42. doi:10.1108/JTF-11-2015-0048.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. SAGE Publications.

Prattico, K., Derchi, F., & De Stefano, R. M. (2025). *Massive transformative purpose and nuances of corporate purpose as leverage for value re-generation: RePlanet case study*. Proceedings of the EURAM 2025 (European Academy of Management), Florence, Italy, 22–25 June, ISBN 978–972-9602195–9602197-9.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2020). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, 85(1), 131–151. doi:10.1177/0022242920953847.

Rashid, Y., Rashid, A., Akib, M., Sabir, S., & Waseem, A. (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods*, 18, 1–13. doi:10.1177/1609406919862424.

Reese, B. (2018). *The fourth age: Smart robots, conscious computers, and the future of humanity*. Atria Books.

Robertson, J., Botha, E., Oosthuizen, K., & Montecchi, M. (2025). Managing change when integrating artificial intelligence (AI) into the retail value chain: The AI implementation compass. *Journal of Business Research*, 189, Article 115198. doi:10.1016/j.jbusres.2025.115198.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. SAGE Publications.

Schultz, M. D., Clegg, M., Hofstetter, R., & Seele, P. (2025). Algorithms and dehumanization: A definition and avoidance model. *AI and Society*, 40(4), 2191–2211. doi:10.1007/s00146-024-02123-7.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. doi:10.1177/0008125619862257.

Srinivasan, R., & Abi, G. S. (2021). When algorithms fail: Consumers' responses to brand harm crises caused by algorithm errors. *Journal of Marketing*, 85(5), 74–91. doi:10.1177/0022242921997082.

Time.com. (2016, September 8). Airbnb is overhauling its anti-discrimination policy. https://time.com/4484084/airbnb-is-overhauling-its-anti-discrimination-policy/.

Williams, J., et al. (2022). Eudaimonic experiences in hospitality: The perceived alignment between brand and algorithms. *Tourism Management*, 90, Article 104512.

Williams, M. (2024). Multi-objective reinforcement learning from AI feedback. *ArXiv*. doi:10.48550/arXiv.2406.07295.

Xu, C., Sun, Y., & Zhou, H. (2024). Artificial aesthetics and ethical ambiguity: Exploring business ethics in the context of AI-driven creativity. *Journal of Business Ethics*. Advance online publication. doi:10.1007/s10551-024-05837-2.

Yannopoulou, N., Moufahim, M., & Bian, X. (2013). User-generated brands and social media: Couchsurfing and AirBnb. *Contemporary Management Research*, 9(1), 85–90.

Yin, R. K. (2013). *Case study research: Design and methods*. SAGE Publications.

Zhang, S., Mehta, N., Singh, P. V., & Srinivasan, K. (2021). Can an artificial intelligence algorithm mitigate racial economic inequality? An analysis in the context of Airbnb. *Marketing Science*, 40(5), 813–820. doi:10.1287/mksc.2021.1276.

PART V

Specialized Applications and Governance: AI in Critical, Regulated, and High-Stakes Environments



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

13

DATA MANAGEMENT AND PRIVACY IN AI FOR AEROSPACE AND DEFENSE

Architecting for Assurance in Critical and Cyber-Physical Systems

Sylvester Kaczmarek

1 Introduction

The aerospace and defense (A&D) sector, a global industry valued in the trillions of dollars, is demonstrably at the forefront of the transformation driven by artificial intelligence (AI) (PwC, 2025). AI is no longer a futuristic concept but a present-day enabler, driving advancements from enhanced situational awareness through intelligent sensor fusion and optimized logistics powered by predictive analytics, to the deployment of autonomous operational systems and sophisticated predictive maintenance schedules that increase platform availability (Deloitte, 2024). The global AI market in A&D is projected for significant growth, with some estimates projecting it to more than double in the next decade, underscoring its escalating importance (Precedence Research, 2024). At the heart of this revolution lies data—vast, varied, and vital. Data serves as the critical fuel for AI algorithms, training them to recognize patterns, make predictions, and execute complex tasks. However, this profound reliance on data introduces a dual challenge: while data propels innovation and operational advantage, it simultaneously presents significant and multifaceted risks to security, privacy, and ethical integrity, particularly within the high-stakes A&D environment, which has zero tolerance for failure. This chapter uses the aerospace and defense sector as its primary lens because it represents the most demanding proving ground for the principles of data and AI assurance. The lessons learned in securing data for these critical terrestrial systems are directly applicable to the ultimate benchmark: architecting trustworthy AI to operate safely in the unforgiving environment of space.

The aerospace and defense landscape is characterized by its inherently critical nature, where system failures or data breaches can trigger severe, far-reaching consequences. Such incidents can impact national security, compromise mission success, lead to substantial economic loss, erode technological superiority, and, in some scenarios, endanger human lives. A data breach within the supply chain of a major defense program, for instance, can lead to the exfiltration of sensitive design schematics and operational parameters, eroding a nation's technological edge and placing

DOI: 10.4324/9781003632429-19
This chapter has been made available under a CC-BY-NC-ND 4.0 license

personnel at risk (Reuters, 2013; The Diplomat, 2017). Unlike many commercial applications, AI systems in A&D frequently operate with highly sensitive, classified, or mission-critical information. These systems must function with unwavering reliability in complex, dynamic, and potentially contested or denied environments, demanding exceptional levels of robustness, resilience, and trustworthiness. The characteristically long lifecycles of A&D platforms, the intricacy of global supply chains involving numerous partners, and the imperative for seamless interoperability in coalition or multi-national operations further compound the complexities of data governance and AI integration (Deloitte, 2024; McKinsey, 2023, 2024). Consequently, the imperative to safeguard data and ensure its ethical handling is not merely a matter of regulatory compliance but a fundamental pillar of operational viability, mission assurance, and sustained strategic advantage.

The integration of AI into A&D systems surfaces a unique array of data management and privacy challenges that demand specialized, proactive attention. A&D organizations face the task of securely collecting, storing, processing, and sharing enormous and diverse datasets: ranging from high-resolution intelligence, surveillance, and reconnaissance (ISR) outputs and telemetry from myriad sensors, to intricate operational logs, sensitive personnel information, and performance data from advanced weaponry (GE Aerospace, 2017; HCLTech, 2024; Waharte & Trigoni, 2010). Ensuring the integrity (accuracy and consistency), provenance (verified origin and history, which is crucial for auditable decision-making), and timeliness of this data is critical, as AI-driven decisions in A&D can have profound and irreversible outcomes. Moreover, the sophisticated nature of many AI models, often perceived as "black boxes," raises significant concerns regarding transparency, explainability, accountability, and the potential for embedded biases that could lead to unintended or discriminatory outcomes. These are all vital aspects for maintaining human–AI trust and ethical operational control. This has led to the development of foundational guidelines, such as the U.S. Department of Defense (DoD) Ethical AI Principles, to steer development toward responsible and equitable systems (Gunning & Aha, 2019; Philosophy & Technology, 2021; Taddeo & Floridi, 2018; U.S. Department of Defense, 2020). Another critical layer of concern is protecting the valuable AI models themselves from theft, reverse engineering, tampering, or sophisticated adversarial attacks (malicious inputs designed to deceive AI models). These models are critical intellectual property and encapsulate sensitive information gleaned from training data (Belfer Center, 2019; IEEE Spectrum, 2020; Kaczmarek, 2025; MIT Technology Review, 2019; National Security Commission on Artificial Intelligence, 2021; Papernot et al., 2018). These challenges are often amplified in scenarios involving highly autonomous systems or platforms operating at the tactical edge, where communication constraints and latency sensitivities necessitate novel approaches to distributed data governance, robust security, and privacy-preserving techniques such as federated learning (Military Embedded Systems, 2021; Simply NUC, 2025; Washington Technology, 2025). The complex regulatory environment, encompassing data protection mandates, sector-specific cybersecurity standards like CMMC, NIST SP 800–171, and international considerations such as export controls (e.g., ITAR, EAR) for sensitive AI technologies, adds further layers of complexity to data handling in A&D (Bureau of Industry and Security, 2022; Center for Security and Emerging Technology, 2023; Daniel & Meg,

2022; Dr. Beyza & Ulysse, 2024; National Institute of Standards and Technology, 2024; Sharetru, 2023).

Traditional data management paradigms, often designed for more static, centralized, and less overtly adversarial IT environments, frequently prove insufficient to address the dynamic, distributed, and highly sensitive nature of AI-driven operations in aerospace and defense. There is an urgent and growing need for a principled design framework that moves beyond ad hoc solutions by integrating security and adaptability at the architectural level. While existing literature addresses specific technical vulnerabilities or high-level ethical principles, there remains a gap in providing an integrated, managerially oriented framework that bridges technical security, data governance, and organizational strategy specifically for the A&D context. This chapter addresses this gap by asking the following research question: How can aerospace and defense organizations develop and implement a holistic framework to manage data security and privacy challenges in AI-driven critical systems? In answering this question, the chapter makes a dual contribution. For practitioners, it offers a structured, actionable framework and strategic roadmap. For academics, it provides a synthesized model that integrates disparate streams of research on data security, AI ethics, and technology management within the critical systems domain. Drawing on insights from current practices, documented challenges, and emerging technological solutions, this chapter will offer practical guidance for business leaders, policymakers, AI practitioners, and legal and compliance professionals seeking to harness the power of AI responsibly and securely. To formalize these competing demands, we first introduce a conceptual model that frames the core strategic challenge.

To achieve these objectives, this chapter is structured as follows: Section 1.1 introduces the Assurance Trilemma, a conceptual model that frames the core challenges of the A&D sector. Section 2 examines the specific data lifecycle challenges encountered in critical and cyber-physical systems within A&D. Section 3 analyzes the evolving regulatory landscape and compliance requirements pertinent to AI and data privacy in this domain. Section 4 introduces a comprehensive framework for data management and privacy in AI, designed to navigate the Assurance Trilemma by covering key principles, governance structures, and technological enablers. Section 5 presents illustrative case studies from the aerospace and defense sector, showcasing practical approaches to secure and ethical data handling in AI systems. Finally, Section 6 discusses strategies for bridging organizational AI readiness with resilient, secure and compliant implementation practices, followed by a concluding summary of key findings, recommendations, and future trends in this rapidly evolving field.

1.1 The Assurance Trilemma in Aerospace and Defense

The unique challenges of the A&D sector can be conceptualized as an assurance trilemma, a persistent tension between three critical and competing objectives: mission capability, operational constraints, and adversarial resilience (see Figure 13.1). Mission capability represents the drive for performance and innovation, the advanced functionalities that AI promises. Operational constraints encompass the strict, real-world limitations of viability and compliance, including size, weight, and power (SWaP), cost, and the complex regulatory landscape. Finally, adversarial resilience is the non-

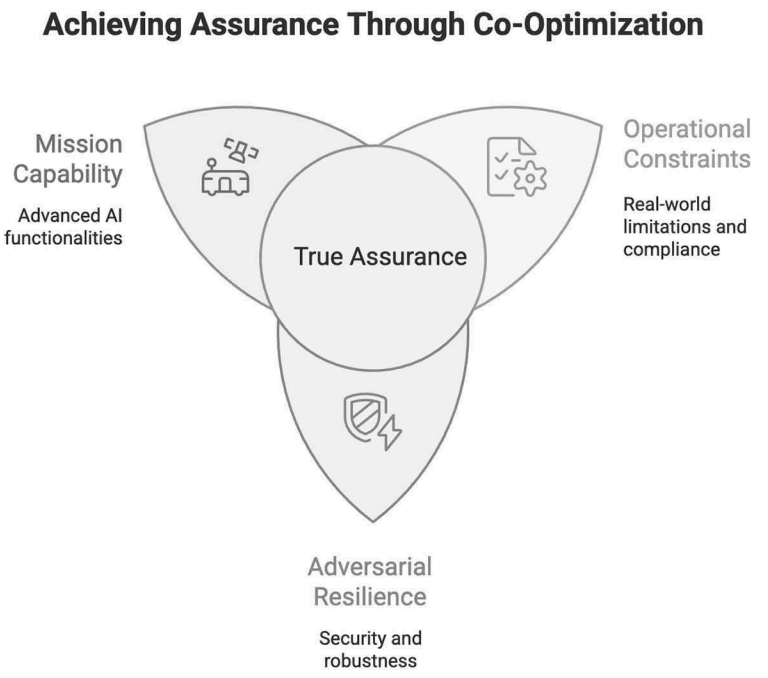


FIGURE 13.1 The assurance trilemma for AI in aerospace and defense, illustrating the tension between mission capability, operational constraints, and adversarial resilience.
Source: Author's own elaboration.

negotiable requirement for security and robustness against both intelligent threats and environmental hazards. A naive approach that maximizes one vertex, such as pursuing maximum performance without regard for security, creates profound vulnerabilities. This chapter argues that achieving true assurance requires a principled design framework that explicitly seeks to co-optimize all three vertices of this trilemma.

1.2 Methodology

This chapter develops a conceptual framework for managing data and AI security in the aerospace and defense sector. The methodology is grounded in a narrative review and synthesis of three distinct streams of literature: academic research on AI security and data governance; industry best practices and white papers from leading A&D and technology firms; and official reports, standards, and strategic documents from government and defense organizations. This multi-source approach ensures that the resulting framework is both theoretically informed and practically relevant to the unique operational and regulatory challenges of the A&D environment.

The case studies presented in Section 5 are synthesized as illustrative archetypes, an approach that allows for the examination of realistic challenges while respecting the sensitive and often classified nature of specific A&D programs. They are informed by

the author's primary empirical experience and publicly available information on industry challenges and technological solutions. The rationale for their selection was to represent three distinct and critical A&D challenges that test different aspects of the proposed framework:

1. handling highly classified data in a national security context (Case 5.1);
2. enabling secure multinational data collaboration under strict sovereignty and export control constraints (Case 5.2);
3. balancing mission-critical safety with human-centric privacy for personnel (Case 5.3).

By clarifying this methodological approach, the chapter aims to provide a robust conceptual model that is transparent in its construction and directly applicable to practitioners, while also laying the groundwork for future empirical validation.

2 Data Lifecycle Challenges in Critical A&D Systems

The data management and privacy challenges inherent in aerospace and defense are not monolithic; they manifest differently and with varying intensity across each stage of the data lifecycle. From the moment data is generated to its final disposal, the unique operational and security demands of A&D amplify the risks and complexities involved. This four-stage model of collection, storage, sharing, and disposal is adopted for its clarity and direct alignment with the operational data flows typical in A&D environments, from initial sensor acquisition to final archival and decommissioning. A systematic examination of this lifecycle provides a clear framework (as illustrated in Figure 13.2) for understanding the specific vulnerabilities that must be addressed to ensure secure and responsible AI implementation.

2.1 Data Collection and Generation

In the A&D sector, data is generated from an exceptionally diverse and distributed array of sources. These include high-fidelity sensors on satellites, aircraft, and ground vehicles; intelligence, surveillance, and reconnaissance (ISR) platforms capturing imagery and signals intelligence; operational logs from complex missions; physiological and biometric data from pilots and soldiers; and telemetry from weapons systems and manufacturing processes. The primary challenge at this initial stage is ensuring data integrity and traceable provenance from the point of origin, which is paramount for AI systems. AI models, whether for predictive maintenance or real-time threat detection, are fundamentally dependent on the quality and trustworthiness of their training and operational data. In a contested operational environment, data streams are susceptible to adversarial manipulation, such as GPS spoofing or the injection of false information to mislead AI-driven analytical systems. Therefore, establishing a trusted and verifiable chain of custody for data is a mission-critical necessity. Furthermore, the collection of sensitive information, particularly biometric or personnel data, immediately introduces privacy concerns, requiring adherence to data protection principles and ethical guidelines to ensure lawful basis for collection and to prevent



FIGURE 13.2 The four stages of the A&D data lifecycle, highlighting the primary security and integrity challenge at each step.
Source: Author's own elaboration.

unauthorized use (Department of Defense, 2019; Kufakunesu et al., 2025; Morgan et al., 2020; Papernot et al., 2015; Rettore et al., 2024; Zabala-L–pez et al., 2024; Zhou et al., 2022).

2.2 Data Storage and Processing

Once collected, A&D data must be stored and processed under stringent security protocols due to its often highly sensitive or classified nature. The challenge is twofold: securing vast repositories of data ("data at rest") and protecting data during computation ("data in use"). Storing petabytes of mission data requires scalable and resilient infrastructure, but centralized repositories can become high-value targets for state-level cyber adversaries and insider threats. For AI, this risk extends beyond the raw data to the trained models themselves, which represent significant intellectual property and can be reverse-engineered to reveal sensitive information about their training data if compromised. Protecting these models from theft, extraction, or tampering is a

critical security consideration. At the tactical edge, the shift to local data processing on devices like autonomous drones or integrated combat systems introduces new attack surfaces. Securing the data and AI models on these distributed nodes, while ensuring they can process information effectively with low latency under potentially disconnected conditions, represents a significant architectural and security challenge. Storing personal data, such as health metrics for performance monitoring, also requires robust encryption, data sovereignty controls, and strict access controls to comply with privacy regulations and prevent misuse (Cloud Security Alliance, 2024; Jin et al., 2022; Ristenpart et al., 2009; Schellman, 2024; Tonge et al., 2024).

2.3 Data Sharing and Dissemination

Effective data sharing is fundamental to modern A&D operations, particularly in coalition environments requiring technology transfer approvals, and across complex, multi-tiered supply chains. However, this operational imperative exists in direct tension with security and privacy requirements. The challenge is to enable the "need to share" for mission success while enforcing the "need to know" to protect sensitive information. When sharing AI-generated intelligence, it is crucial to ensure that the shared insights do not inadvertently reveal the underlying sensitive training data or proprietary model architectures through inference attacks. Techniques such as differential privacy can be employed to add statistical noise, allowing for the dissemination of valuable analytical products while protecting the source data. Similarly, sharing personnel data across international partners for coalition operations necessitates robust privacy safeguards, such as anonymization or pseudonymization aligned with partner-nation regulations, to comply with varying cross-border data transfer regulations. Achieving seamless and secure interoperability between disparate systems, each with its own security policies, data classification levels, and data formats, remains one of the most persistent and difficult challenges in the A&D data lifecycle (Deni et al., 2025; Galvin, 2025; Gormus, 2024; Pettai & Laud, 2015; U.S. Government Accountability Office, 2021; Ziani, 2022).

2.4 Data Retention and Disposal

The A&D sector is subject to rigorous legal, regulatory, and operational requirements for long-term data retention. Data from flight recorders, weapons testing, and mission logs may need to be archived for decades for accident investigation, performance analysis, algorithmic accountability, or legal purposes. For AI systems, retaining historical data is also essential for model retraining, performance evaluation, and auditing algorithmic decisions. However, this long-term retention must be balanced with data minimization principles and privacy regulations, especially when dealing with personal data that should not be kept indefinitely without a clear purpose. The technologies used to encrypt this archived data may become obsolete or vulnerable over time, requiring continuous cryptographic agility and risk management. Equally critical, and often more challenging, is the secure and permanent disposal of data at the end of its lifecycle across diverse media types (e.g., embedded systems, legacy storage). Simply deleting files is insufficient, as it often leaves recoverable data remnants. The

secure sanitization of media, whether hard drives in a data center or solid-state drives on a decommissioned autonomous vehicle, requires certified methods like physical destruction or cryptographic erasure to ensure that sensitive or classified information is rendered completely unrecoverable, preventing it from falling into the wrong hands (Department of Defense, 2022; Kissel et al., 2014). These lifecycle challenges, from generation to disposal, are governed by the complex regulatory and compliance landscape that will be examined in the next section.

3 The Regulatory and Compliance Landscape for AI in Aerospace and Defense

The data lifecycle challenges in aerospace and defense are not addressed in a vacuum; they are governed by a complex, multi-layered, and often overlapping landscape of regulations and compliance mandates (Office of the Assistant Secretary of Defense for Industrial Base Policy, 2025). For A&D organizations seeking to leverage AI, navigating this terrain is a formidable task. It requires a holistic understanding of rules spanning general data privacy, defense-specific cybersecurity standards, stringent export controls, and emerging AI-specific ethical guidelines. Failure to comply can result in severe penalties, loss of contracts, reputational damage, and significant damage to national security. This section examines the principal domains of this regulatory environment, highlighting their direct implications for the development and deployment of AI systems.

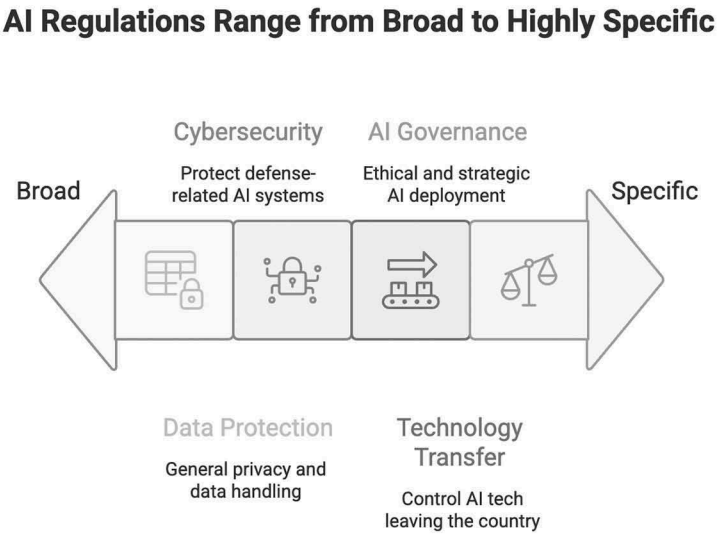


FIGURE 13.3 The multi-layered regulatory and compliance landscape for AI in A&D, showing the four principal domains of governance.
Source: Author's own elaboration.

3.1 Foundational Data Protection Principles

While A&D operations may have unique exemptions, they are not entirely insulated from the global trend toward stronger data privacy rights. Regulations like the European Union's General Data Protection Regulation (GDPR) have established a global benchmark for data handling, built on core principles such as lawfulness, fairness, transparency, data minimization, and purpose limitation. GDPR's provisions on automated decision-making (Article 22) are particularly relevant, creating rights for individuals related to decisions made solely by AI systems and often necessitating human oversight. Other jurisdictions, such as California with its Consumer Privacy Act (CCPA), impose similar obligations. While national security exemptions exist within these frameworks, allowing for necessary defense activities, A&D organizations must still ensure compliance for all other personal data they process, such as employee records or information from commercial or dual-use partners. This requires building privacy-by-design into their systems to manage data in a compliant and ethical manner and necessitates clear documentation of exemption justifications (European Commission, 2024; European Parliament, 2023; European Union, 2016; State of California, 2018).

3.2 Defense-Specific Cybersecurity Mandates

Moving from general privacy to specific security, A&D organizations operate under exacting cybersecurity mandates designed to protect sensitive government information. In the United States, for example, the Cybersecurity Maturity Model Certification (CMMC) program is a cornerstone of this effort, requiring contractors to implement and be audited against standards derived from NIST SP 800–171 to protect Controlled Unclassified Information (CUI) and comply with DFARS clauses. Requirements for handling classified data (e.g., under JSIG) impose even stricter controls. In addition to these government mandates, industry-led standards like the Aerospace Industries Association's (AIA) NAS9933 provide further guidance on best practices for data security across the supply chain. For AI, these mandates are critical: they define the secure environment in which AI systems must be developed, trained, deployed, and maintained. Allied nations have instituted similar cybersecurity requirements, such as the UK's Cyber Essentials Plus scheme or Australia's ISM, creating a global expectation for robust and verifiable security postures across the A&D ecosystem, though harmonization remains challenging (Aerospace Industries Association, 2018; Australian Signals Directorate, 2025; Department of Defense, 2020a, 2020b, 2025; National Institute of Standards and Technology, 2024; UK Ministry of Defence, 2021).

3.3 Export Controls and Technology Transfer

Perhaps one of the most acute regulatory challenges for AI in A&D comes from export control laws. To protect national security interests, governments strictly regulate the export of sensitive defense-related technologies, software, technical data, and data. In the U.S., the International Traffic in Arms Regulations (ITAR) and the

Export Administration Regulations (EAR) are the primary legal frameworks. An AI model trained on sensitive defense data or specifically designed for a military application can be classified as a "defense article" under ITAR or controlled under EAR's Category 4 (Computers) or Category 11 (Sensitive Technology List). This classification has profound implications, severely restricting the ability to share the technology with foreign nationals (even within domestic facilities) and complicating international research collaboration and cloud computing infrastructure choices, often requiring specific Technology Assistance Agreements (TAAs) or exemptions. Recent executive orders have expanded these controls to include restrictions on U.S. investments in foreign AI technologies and on bulk sensitive data transfers to specific countries of concern, reflecting heightened national security considerations. These controls create significant hurdles for multinational programs, such as those within NATO, requiring A&D companies to navigate a complex web of varying export regulations across allied nations (Biden, 2023, 2024; EU AI Act, 2025; European Commission, 2024; U. S. Department of Commerce, Bureau of Industry and Security, 2025; U.S. Department of State, 2025).

3.4 Emerging AI-Specific Governance and Ethics

Recognizing that traditional regulations were not designed to address the unique properties of AI, defense organizations are now developing AI-specific governance principles. The U.S. Department of Defense has established five principles for the ethical use of AI: responsible, equitable, traceable, reliable, and governable. These are complemented by the DoD's Data, Analytics, and AI Adoption Strategy, which provides a practical roadmap for integrating these principles into military operations though operationalizing them remains complex. Similarly, NATO has revised its AI Strategy and is developing an AI certification standard focused on military applications to ensure systems used by the Alliance meet ethical and legal requirements. Globally, the debate on military AI regulation is intensifying, including discussions on autonomous weapons systems. The EU's AI Act, while excluding purely military applications, sets a human-centric, risk-based precedent for dual-use technologies that may influence future international norms. This contrasts with other approaches that prioritize flexible, innovation-friendly standards. Furthermore, standards bodies in adjacent sectors, like the European Union Aviation Safety Agency (EASA) with its AI Roadmap emphasizing airworthiness, are setting precedents for AI certification that could inform the defense sector. Navigating this intricate and evolving legal and ethical terrain requires more than a compliance checklist; it demands a culturally embedded, proactive and integrated framework that addresses liability and accountability gaps, which will be the focus of the subsequent section (Department of Defense, 2020b, 2023; European Union, 2024; European Union Aviation Safety Agency, 2023; NATO, 2024).

4 A Framework for Secure Data Management and Privacy in AI

Addressing the lifecycle vulnerabilities and regulatory complexities detailed in the previous sections requires more than a reactive, checklist-based approach to

compliance. A proactive and integrated framework is essential to embed security and privacy into the very fabric of an organization's AI initiatives. This section proposes such a framework, grounded in a principled design methodology where security and adaptability are treated as emergent properties of the system's architecture, not as bolt-on features. The framework is multi-layered, combining robust governance structures that culminate in meaningful human control, with fundamental security principles, advanced technological enablers including emerging computational paradigms, and dedicated protections for the AI models themselves (see Figure 13.4). It is designed to provide a holistic and defensible strategy for managing data and AI in high-stakes A&D environments.

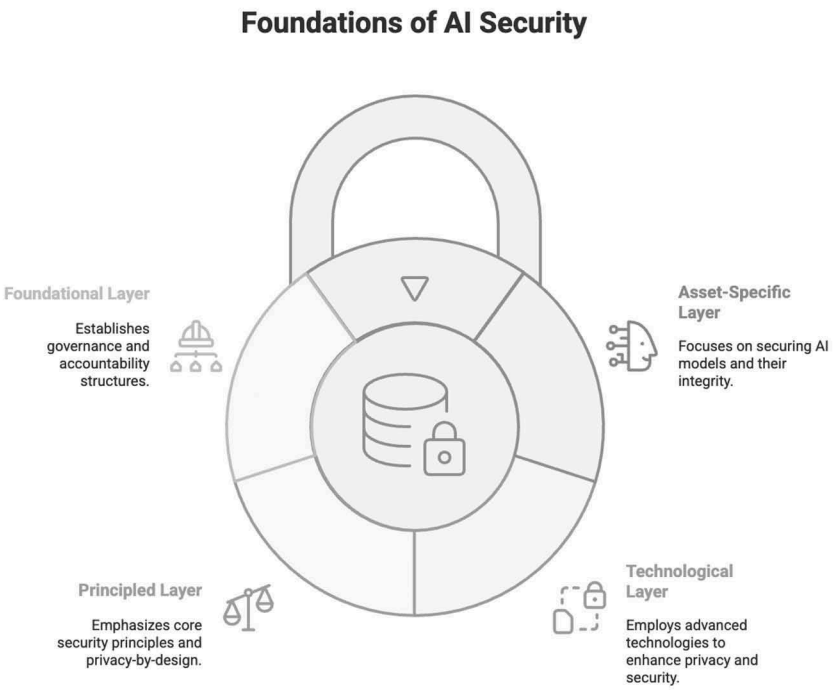


FIGURE 13.4 A multi-layered framework for secure data management and privacy in AI, built upon a foundation of governance and culminating in asset-specific model protection.

Source: Author's own elaboration.

4.1 Foundational Layer: Data Governance and Accountability

Technology alone cannot ensure security or privacy; the foundation of any effective framework is a robust governance structure with clear lines of accountability. This layer is grounded in established principles of information governance and stewardship theory, which move beyond the concept of IT as the sole custodian of data to establish clear ownership at the business and mission levels (Thomas, 2006). Key components of this foundational layer include:

- **Defined Roles and Responsibilities:** Establishing specific roles such as data stewards, responsible for the quality and use of specific data domains, and an AI ethics board or review council with executive sponsorship. This board should have a defined charter to assess the ethical implications, legal compliance, and operational risks of new AI projects, ensuring alignment with standards like the DoD Ethical AI Principles.
- **Comprehensive Policies and Standards:** Developing and enforcing clear, documented policies for data classification, handling, access, and secure deletion, aligning with standards like NIST SP 800-88. This includes explicit requirements for tamper-evident data provenance tracking to ensure a reliable and auditable data lineage from source to model.
- **Continuous Training and Awareness:** Implementing mandatory and ongoing training programs to ensure that all personnel, from engineers to operators, understand their responsibilities under the governance framework, the nature of the threats (including insider risks), and the importance of a security-first culture.

This governance layer ensures that human oversight and accountability are central to the AI ecosystem, creating a culture where security and ethical considerations are a shared responsibility (Blackman, 2022; Department of Defense, 2020, 2024; National Institute of Standards and Technology, 2023; Schuett et al., 2024). Ultimately, this culminates in a commitment to meaningful human control (MHC). MHC is an architectural and operational principle ensuring that humans retain ultimate moral and strategic authority over autonomous systems, a concept codified in foundational defense policy (Department of Defense, 2023). It is achieved by designing systems for interpretability and validating them in high-fidelity digital twins, ensuring that the goal of autonomy is to enhance human capability, not replace human accountability.

4.2 Principled Layer: Core Security and Privacy-by-Design

Built upon the governance foundation, this layer embeds core principles of security and privacy directly into system design and operational workflows. Adopting a "security-by-design" and "privacy-by-design" methodology, a concept rooted in proactive engineering, is critical for preventing vulnerabilities rather than attempting to patch them after the fact. Essential principles include:

- **Data Minimization and Purpose Limitation:** Strictly adhering to the practice of collecting only the data that is absolutely necessary for a specific, defined purpose.

Data collected for one mission should not be repurposed for another without a clear justification and authorization, reducing the overall attack surface and simplifying compliance.

- **Least Privilege and Zero Trust:** Implementing specific access control models like Role-Based Access Control (RBAC) or Attribute-Based Access Control (ABAC), where users and systems are granted only the minimum level of access required. This is part of a broader "zero trust" architecture requiring network micro-segmentation, which assumes no implicit trust and continuously verifies every access request.
- **Continuous Auditing and Monitoring:** Deploying robust systems for logging all data access and system interactions for real-time threat detection. This must include monitoring for data and model drift and integrity anomalies, where the performance of an AI model degrades over time as real-world data deviates from the training data, to ensure continued reliability and accuracy.

These principles form the operational rulebook for data handling, while acknowledging the need to balance security measures with performance requirements, as the computational overhead of some security controls can impact latency in real-time mission systems (CISA et al., 2023; Hu et al., 2014; Kore et al., 2024; Rose et al., 2020; Sandhu et al., 1996).

4.3 Technological Layer: Privacy-Enhancing Technologies (PETs) and Secure Computing

To address the inherent tension between data utility and data protection, organizations can leverage a growing class of advanced technologies. These tools enable data analysis and collaboration while minimizing the exposure of sensitive raw data. Key technologies relevant to A&D include:

- **Privacy-Enhancing Technologies (PETs):** This includes federated learning for collaborative model training without data sharing, differential privacy for anonymizing analytical outputs, secure multi-party computation (SMPC) for joint calculations on encrypted data, and homomorphic encryption for processing encrypted data in untrusted environments via confidential computing.
- **Secure Computing Environments:** Utilizing hardware-based security such as trusted execution environments (TEEs) certified to relevant standards (e.g., FIPS 140-3) or hardware security modules (HSMs) to create isolated, protected enclaves where sensitive data and AI models can be processed, even on a compromised host system.
- **Future-Proofing with Advanced Cryptography:** Planning for the advent of quantum computing by beginning to integrate cryptographic agility and quantum-resistant cryptography (QRC) for critical functions like digital signatures and data encryption, ensuring long-term data security against future threats as recommended by NIST PQC standardization efforts.

The strategic deployment of these technologies provides a technical means to achieve mission objectives that would otherwise be blocked by security or privacy constraints

(National Institute of Standards and Technology, 2001, 2022, 2024a, 2024b, 2024c; OECD, 2025; Pelekis et al., 2025; Sabt et al., 2015; The Royal Society, 2023).

4.3.1 Strategic Enabler: Emerging Computational Paradigms

The practical implementation of this framework, particularly in the most resource-constrained edge environments, is deeply influenced by the underlying computational paradigm. While the principles outlined are hardware-agnostic, achieving the required balance of high performance, low latency, and extreme energy efficiency may necessitate moving beyond traditional von Neumann architectures for certain critical applications. Neuromorphic computing, a paradigm inspired by the brain's event-driven and massively parallel structure, represents a key strategic enabler for this framework. As validated in extensive doctoral research, and demonstrated on platforms like Intel's Loihi processor (Davies et al., 2018) and the BrainChip Akida processor (Lutes et al., 2024), neuromorphic systems can offer orders-of-magnitude improvements in energy efficiency and processing latency for tasks involving sparse, temporal data, such as real-time anomaly detection (Kaczmarek, 2025). By performing computations only in response to significant events ("spikes"), these systems can achieve inference with energy budgets measured in microjoules and latencies in milliseconds. While the technology is still maturing and faces challenges in software ecosystem development and integration, its potential is transformative for deploying sophisticated, adaptive AI on SWaP-constrained platforms like autonomous drones, satellites, and rovers (Schuman et al., 2022). The adoption of such emerging paradigms should be considered a core part of an organization's long-term technology strategy for achieving resilient and sustainable autonomy at the tactical edge.

4.4 Asset-Specific Layer: AI Model Security and Integrity

Finally, the framework must recognize that AI models are themselves critical assets that require dedicated protection. This layer applies principles of AI-specific threat modeling to identify and mitigate risks unique to machine learning systems (Marshall et al., 2019). Securing the data is necessary but not sufficient; the model itself is a new and valuable attack surface. This layer focuses on:

- **Robust Adversarial Defense:** Implementing model hardening techniques like "adversarial training" to make models more resilient and utilizing established libraries and tools for testing and defense against attacks (Kaczmarek, 2025). This also includes regular AI red teaming and penetration testing to proactively identify and mitigate vulnerabilities specific to the operational context.
- **Data and Model Integrity:** Protecting the training data pipeline from data poisoning via strict vetting and ensuring model integrity through the use of digital signatures for model files and secure bootstrapping. This includes securing the model weights through strict access controls, encryption, and centralized, secure storage to prevent theft or tampering. These weights represent the core parameters of the trained model and are a primary target for theft.

- **Model Provenance and Transparency:** Maintaining a detailed record of the model's lineage, often referred to as an "AI Bill of Materials" (AI BOM), which documents the datasets, libraries, and parameters used in its creation and tracks vulnerability disclosures. This aids in traceability, vulnerability management, and ensuring the model has not been tampered with.
- **Intellectual Property Protection:** Employing techniques such as digital watermarking or model obfuscation within tamper-resistant hardware to deter the theft of valuable, proprietary AI models.

By treating the AI model as a key asset to be defended, this layer completes the holistic framework, ensuring that both the data and the intelligence derived from it are secure across the supply chain. The practical application of these framework layers will be further explored through the case studies in the following section (Barrett et al., 2024; Bennet et al., 2025; National Institute of Standards and Technology, 2024; OWASP Foundation, 2025; Perez et al., 2022; RAND Corporation, 2024; Rando et al., 2025; Wiz, 2025).

TABLE 13.1 Summary of the AI security framework

Framework Layer	*Key Components*	*Primary A&D Challenge Addressed*	*Managerial Implication*
Foundational	Defined Roles & Responsibilities; Comprehensive Policies; Continuous Training	Ambiguous accountability; Insider threats; Inconsistent security practices	Establishes clear ownership and a defensible governance posture, reducing human error and compliance risk.
Principled	Data Minimization; Zero Trust Architecture; Continuous Monitoring	Data spillage; Network breaches; Undetected model degradation	Reduces the attack surface, hardens the infrastructure against attack, and ensures ongoing system reliability.
Technological	PETs (Federated Learning, etc.); Secure Computing (TEEs); Quantum-Resistant Crypto	Data sovereignty conflicts; Processing classified data; Long-term data compromise	Enables secure collaboration and analysis that would otherwise be impossible, future-proofing critical data assets.
Asset-Specific	Adversarial Defense; Model Integrity; AI Bill of Materials; IP Protection	Model evasion/ deception; Data poisoning; Model theft; Supply chain vulnerabilities	Protects the AI model as a critical asset, ensuring its trustworthiness and preserving competitive advantage.

Source: Authors' own elaboration, based on Marshall et al. (2019), Kaczmarek (2025), Barrett et al. (2024), National Institute of Standards and Technology (2024), OWASP Foundation (2025), and RAND Corporation (2024)

5 Illustrative Case Studies from Aerospace and Defense

The principles and technologies outlined in the preceding framework can seem abstract without concrete examples. To bridge the gap between theory and practice, this section presents three illustrative case studies. As described in the methodology, these cases are not empirical studies but illustrative archetypes, synthesized from industry practices and designed to represent distinct and critical A&D challenges. These cases demonstrate how different layers of the framework can be applied to solve complex data management and privacy challenges, enabling the secure and effective use of AI in mission-critical contexts.

Case Study 5.1: Secure Data Enclaves for AI Model Training on Classified Intelligence

Scenario: A national intelligence agency needs to develop a sophisticated AI model to automatically identify emerging threats from vast quantities of classified satellite imagery and signals intelligence (SIGINT). The raw data is highly sensitive, and its exposure would compromise sources and methods, posing a grave risk to national security.

Challenge: How can data scientists train a high-performance AI model without directly exposing the classified raw data to their development environment or risking data leakage during model iteration or debugging?

Framework Application:

- **Foundational Layer (Governance):** A strict governance policy is established, defining who can access the enclave, under what conditions, and for what specific purpose. All actions are logged and auditable by independent security officers, ensuring compliance with relevant security standards like ICD 503 (Intelligence Community Directive).
- **Principled Layer (Security-by-Design):** The principle of least privilege is strictly enforced. Data scientists are granted access only to the tools within the enclave, not the underlying data itself. The data is subject to purpose limitation; it cannot be exfiltrated or used for any other project and is automatically deleted after training completion.
- **Technological Layer (Secure Computing):** The core solution is a secure data enclave, a highly isolated and controlled computing environment, potentially leveraging technologies like AWS Trusted Secure Enclaves or NVIDIA's confidential computing capabilities. The classified data is loaded into this enclave via one-way data diodes, which is air-gapped or connected only via highly monitored data diodes. Data scientists interact with the data only through secure, remote terminals with no external connectivity.
- **Asset-Specific Layer (Model Security):** Once trained, the final model weights are extracted from the enclave only after a rigorous security review for embedded sensitive information and an automated process that verifies the code used for extraction has been approved and digitally signed, preventing unauthorized code from exfiltrating data. The model itself is then treated as a classified asset, digitally signed to ensure its integrity before deployment to operational systems, a process

underpinned by a hierarchical defense philosophy to protect the model from both physical and adversarial perturbations (Kaczmarek, 2025).

Outcome: The intelligence agency successfully develops a state-of-the-art AI model for threat detection. The secure enclave allows for the use of highly sensitive data to achieve maximum model performance while ensuring that the data itself remains protected from exposure, adhering to both security protocols and the principle of need-to-know and enabling faster accreditation. This approach enables the development of highly responsive systems suitable for mission-critical AI applications (Amazon Web Services, 2023; Director of National Intelligence, 2024; OpenMined, 2024; Rathi et al., 2021; Shevlane, 2022; Waterman, 2025).

Case Study 5.2: Federated Learning for Collaborative Predictive Maintenance in a Coalition Fleet

Scenario: A coalition of allied nations operates a common fleet of advanced fighter aircraft. Each nation collects detailed operational and maintenance data from its own aircraft. They wish to build a shared AI model for predictive maintenance to improve fleet readiness and reduce costs, but they are prohibited by national regulations and ITAR constraints from sharing their raw, sovereign operational data with each other.

Challenge: How can the allies collaboratively train a single, superior AI model without centralizing their sensitive, sovereign data while ensuring no single nation can infer another's proprietary data or operational patterns?

Framework Application:

- **Foundational Layer (Governance):** The coalition establishes a joint governance agreement under a NATO Technical Arrangement that defines the data standards, the model's objectives, and the rules of engagement for the federated learning process, similar to frameworks explored in research initiatives like Europe's H2020 ReMAP project for aircraft maintenance.
- **Principled Layer (Privacy-by-Design):** The core principle is data minimization and sovereignty. No raw operational data ever leaves the sovereign control of each nation. Only the mathematical updates (gradients) from the model training process are shared after sanitization checks.
- **Technological Layer (PETs):** The coalition implements a federated learning architecture with secure aggregation. Each nation trains a copy of the global AI model on its local data within its own secure boundary. The resulting model updates, not the data itself, are then encrypted using homomorphic encryption and sent to a central server managed by the coalition. This server aggregates the updates without decrypting individual contributions to create an improved global model, which is then sent back to the nations for the next round of training.
- **Asset-Specific Layer (Model Security):** The aggregated global model is protected as a shared coalition asset with access controls based on nationality clearance levels. The communication channels used to transmit the model updates are secured with end-to-end encryption, potentially using quantum-resistant algorithms to future-proof against emerging threats and monitored for anomalies.

Outcome: The coalition successfully develops a highly accurate predictive maintenance model that outperforms any model they could have trained individually. Federated learning allows them to achieve a collective benefit while fully respecting national data sovereignty, security regulations, and export controls, enhancing the readiness of the entire fleet and reducing maintenance costs by an estimated 15–20%. The viability of deploying such sophisticated models on individual, SWaP-constrained aircraft is underpinned by the extreme energy efficiency of emerging computational paradigms (Bonawitz et al., 2016; H2020 ReMAP Consortium, 2022; Jin et al., 2023; Kaczmarek, 2025; Landau et al., 2025; Rahman & Morshed, 2025; Tsai & Morshed, 2025).

Case Study 5.3: Privacy-Preserving AI for Pilot Health and Fatigue Monitoring

Scenario: An air force wants to implement an AI-driven system to monitor pilot health and cognitive fatigue in real-time using in-cockpit sensors that collect biometric data (e.g., heart rate, eye movement, EEG). The goal is to enhance safety by detecting early signs of incapacitation. However, pilots have significant privacy concerns about the continuous collection and analysis of their personal health data and potential misuse for career decisions.

Challenge: How can the air force leverage AI for safety enhancement while protecting the personal privacy of its pilots and complying with relevant regulations like HIPAA (Health Insurance Portability and Accountability Act, 1996) and DoD Instruction 6025.18 (Department of Defense, 2019) (Privacy of Individually Identifiable Health Information)?

Framework Application:

- **Foundational Layer (Governance):** An AI ethics board, including pilot representatives and medical privacy experts, and legal counsel, is established to create a strict policy governing the use of the biometric data. The policy clearly states that the data will be used *only* for real-time safety alerts and anonymized aggregate analysis, not for performance evaluation, and must comply with relevant health data privacy standards, such as principles found in HIPAA or FAA Advisory Circular 120-108 (Continuous Monitoring).
- **Principled Layer (Privacy-by-Design):** The system is built on the principles of data minimization and purpose limitation. The AI model running on the edge (in the cockpit) processes the raw biometric data locally. It only transmits an alert metadata (e.g., "Fatigue Alert Level 2") if a critical safety threshold is crossed. The raw data stream is not continuously transmitted or stored centrally beyond a short buffer for incident reconstruction.
- **Technological Layer (PETs):** For long-term analysis to identify fleet-wide fatigue trends, differential privacy is applied with tight epsilon parameters. When anonymized, aggregated data is periodically uploaded to a central server, statistical noise is added in such a way that aggregate trends can be identified, but it is mathematically impossible to reverse-engineer the specific biometric data of any individual pilot from the dataset.
- **Asset-Specific Layer (Model Security):** The on-board AI model is secured in a hardware-secured module to prevent tampering, ensuring that the alert thresholds cannot be maliciously altered or disabled.

Outcome: The system is successfully deployed, enhancing flight safety by providing critical real-time alerts with no false negatives in initial trials. By using a combination of edge processing, clear governance, and differential privacy, the air force builds trust with its pilots, demonstrating that the system's primary purpose is safety, not surveillance. This balanced approach allows for the adoption of a life-saving technology while respecting individual privacy and adhering to regulatory principles and military medical ethics. These practical applications demonstrate that a structured framework is not a barrier to innovation but an essential enabler for the responsible deployment of AI in the complex A&D landscape (Chen et al., 2025; Dyda et al., 2021; Federal Aviation Administration, 2023; Kaczmarek, 2025; Rathi et al., 2021; Tsai & Morshed, 2025; Vincze et al., 2025).

6 From Readiness to Resilience: A Strategic Roadmap for AI Implementation

The successful integration of AI into the aerospace and defense sector is a continuous strategic process. Achieving enduring operational resilience, which is the ability to withstand, adapt to, and recover from disruptions while maintaining secure operations, requires a deliberate and structured roadmap. This roadmap translates the principles of the framework from the previous section into a sequence of actionable, strategic phases, as shown in Figure 13.5. Although presented sequentially, these phases are iterative and often overlap, requiring continuous attention.

6.1 Phase 1: Foundational Readiness Assessment

Before embarking on significant AI initiatives, an organization must first conduct a comprehensive and honest assessment of its current capabilities and maturity. This foundational audit goes beyond technical infrastructure to provide a holistic view of the organization's readiness to manage AI securely and responsibly. Key assessment domains include:

- **Data Maturity and Governance:** Evaluating the organization's control over its data assets and governance processes using a structured approach, such as the data management maturity (DMM) model. This includes assessing compliance with relevant data privacy regulations (e.g., GDPR, CCPA) for all personal data holdings and identifying gaps in provenance tracking for critical datasets.
- **Technical Infrastructure and Security Posture:** Assessing the adequacy of infrastructure to support AI workloads and existing cybersecurity controls against relevant standards, such as NIST SP 800-171 for protecting controlled unclassified information, as a baseline for CMMC compliance and readiness for secure AI deployment.
- **Skills and Talent:** Identifying gaps in essential personnel, including data scientists, AI/ML engineers, AI security specialists, and cybersecurity experts with domain-specific AI knowledge. This should also include developing strategies for talent retention and acquisition, such as partnerships with universities and defense research labs.

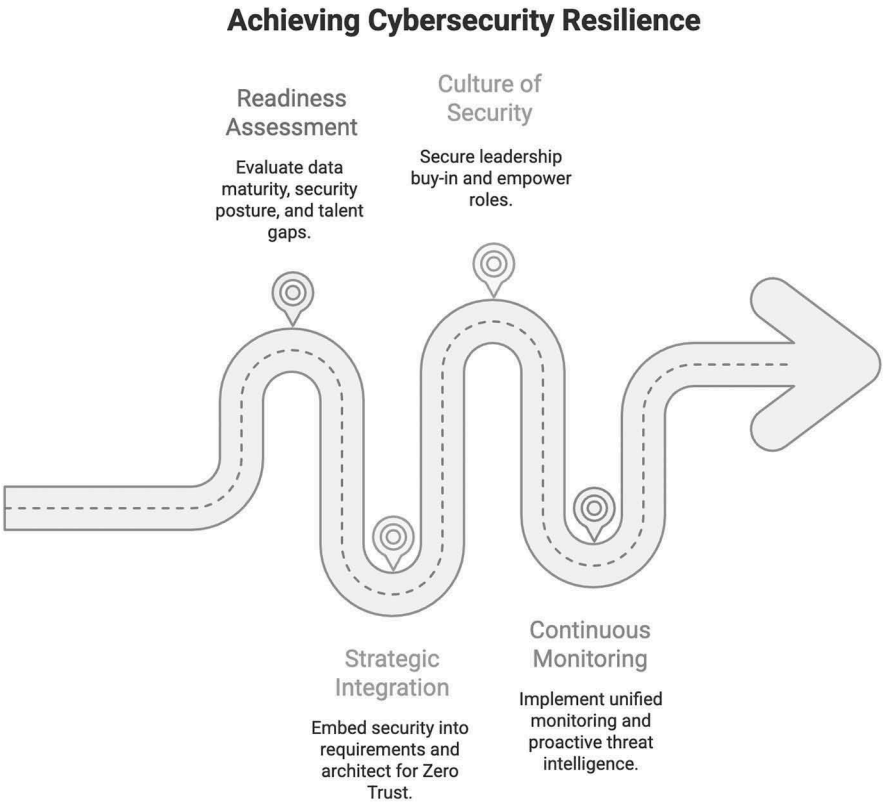


FIGURE 13.5 A strategic roadmap for AI implementation, illustrating the four iterative phases from foundational readiness to continuous resilience.
Source: Author's own elaboration.

The output of this phase is a clear gap analysis that identifies specific weaknesses and provides a risk-based, prioritized list of foundational investments required before scaling AI implementation informed by mission criticality (Boston Consulting Group, 2025; CMMI Institute, 2014; Weill et al., 2024).

6.2 Phase 2: Strategic Integration and Security-by-Design

With a clear understanding of its starting position, the organization must embed the principles of the framework directly into its established development and procurement lifecycles. Bolting on security as an afterthought is a recipe for failure; it must be a core requirement from inception. This phase involves:

- **Integrating Security and Privacy into Requirements:** Defining specific security, privacy, ethical, and export control requirements alongside functional performance requirements at the very beginning of any AI project (e.g., specifying that

an AI system handling classified data must meet a certain DoD security level and ITAR compliance).

- **Architecting for Security:** Making deliberate design choices based on the framework, such as implementing a zero-trust architecture, which verifies every access request and assumes no implicit trust, to protect against both external and internal threats. This includes selecting appropriate privacy-enhancing technologies (PETs) during the system design phase and defining cryptographic standards.
- **Secure Supply Chain Management:** Enforcing security requirements on all third-party vendors and partners through contractual obligations leveraging standards like CMMC, continuous monitoring, and software bill of materials (SBOM) verification, addressing the complexities of multi-tiered A&D supply chains.
- **Robust Testing and Validation:** Expanding testing protocols beyond functional validation to include dedicated security testing, vulnerability scanning, and AI-specific red teaming focused on adversarial tactics relevant to the operational domain, which involves actively testing models for vulnerabilities like adversarial evasion or embedded biases before deployment.

This security-by-design approach ensures that the resulting AI systems are not only capable but also inherently defensible and accreditable (Cimphony AI, 2024; National Institute of Standards and Technology, 2020, 2022, 2023).

6.3 Phase 3: Fostering a Culture of Security and Accountability

A successful AI strategy is as much about people and culture as it is about technology. A resilient organization is one where security and accountability are shared responsibilities, championed from the top down. This requires a concerted effort in leadership, culture, and change management:

- **Leadership and Buy-in:** Senior leadership must visibly and consistently champion the strategic importance of secure and ethical AI. This involves moving beyond passive approval to active sponsorship, which can be secured by presenting clear risk assessments and case studies that frame security investment not as a cost center, but as a critical mission enabler.
- **Organizational Culture:** A resilient culture is one where security is ingrained in daily routines and decision-making. This is fostered by establishing cross-functional teams that break down organizational silos, bringing together mission operators, data scientists, security experts, and legal counsel. In collaborative projects, these teams must include representatives from partner organizations to ensure shared governance and alignment on security protocols.
- **Change Management:** Implementing this framework constitutes a significant organizational change. Success depends on applying established change management principles, such as clear communication about the "why" behind new security processes, empowering roles like data stewards with genuine authority to enforce policies, and aligning incentives by integrating security compliance and data stewardship metrics into performance evaluations to reinforce desired behaviors (Boston Consulting Group, 2025; Corradini, 2020; Kotter, 2012).

This cultural shift transforms the framework from a document into a living, operational reality essential for coalition interoperability.

6.4 Phase 4: Continuous Monitoring and Adaptation for Resilience

In the dynamic A&D environment, resilience is not a static state but a process of continuous adaptation. Once an AI system is deployed, the work of securing it has only just begun. This final, ongoing phase focuses on:

- **Unified Monitoring:** Implementing a unified monitoring strategy that uses automated tools and AI-driven analytics to track not only for cybersecurity threats but also for AI-specific issues like data and model drift or integrity compromises to ensure the system remains accurate, reliable, and secure over time.
- **Proactive Threat Intelligence:** Actively monitoring the evolving landscape of AI attack techniques and defense countermeasures by leveraging sources like government alerts (e.g., CISA, NCSC), industry reports, and AI-specific threat intelligence platforms tailored to critical infrastructure threats.
- **AI-Specific Incident Response:** Developing and rehearsing incident response plans that are specifically designed for AI-related failures including supply chain compromises. These plans should outline clear steps for isolating affected systems, revoking model access, initiating model retraining with sanitized data, notifying stakeholders, and meeting regulatory reporting obligations.
- **Iterative Framework Improvement:** Establishing a formal process (e.g., annual reviews, post-incident audits) to regularly review and update the organization's governance framework, security controls, PET implementations, and AI strategies based on lessons learned from monitoring, near-misses, incidents, technology refresh cycles, and changes in the regulatory and threat landscapes.

This strategic, lifecycle approach to AI implementation is essential for achieving not just initial capability, but lasting operational advantage and trusted autonomy, a theme that will be synthesized in the concluding section of this chapter (Beer Mohamed et al., 2021; Condon, 2022; National Institute of Standards and Technology, 2011, 2023; Radiant Security, 2024).

6.5 The C-Suite Assurance Toolkit: Operational KPIs for AI

To translate this strategic roadmap into immediate executive action, leaders should track the following key performance indicators (KPIs) and governance checks. These metrics move beyond simple performance to measure the assurance of the AI ecosystem.

1. **Governance Check:** Has an AI ethics and review board with veto power over non-compliant projects been formally chartered?
2. **Security Metric (Adversarial Success Rate):** Is the system tested against active adversarial perturbations, and is the ASR < 5% in simulated environments?

3. **Resilience Metric (Recovery Time Objective):** In the event of a detected anomaly or attack, can the system revert to a safe state or failover within the mission-critical latency window (e.g., < 100ms)?
4. **Data Integrity (Provenance Coverage):** What percentage of the training and operational data has a cryptographically verifiable lineage? (Target: 100% for critical systems.)
5. **Adaptability Metric (Novelty Response Time):** How quickly can the system adapt to a previously unseen anomaly class without requiring a full ground-based retraining cycle?
6. **Culture Check (Security Literacy Rate):** What percentage of the engineering and operations workforce has completed role-specific training on AI security risks and counter-adversarial techniques?
7. **Conclusion**

The integration of AI into the aerospace and defense sector represents a paradigm shift, offering unprecedented opportunities to enhance strategic advantage, operational efficiency, and mission effectiveness. However, as this chapter has detailed, these opportunities are inextricably linked to the profound and complex challenges of data management, security, and privacy. In the A&D sector, where data often includes highly classified information and operations involve complex multinational partnerships under strict export controls, these challenges are particularly acute. The journey from data collection to disposal is fraught with vulnerabilities that are amplified by the high-stakes, zero-fail nature of A&D operations, demanding a holistic, integrated, and proactive approach focused on resilience (National Academies of Sciences, Engineering, and Medicine, 2022; U.S. Department of Defense, 2018).

This chapter has argued that a reactive, compliance-focused posture is insufficient for navigating this complex landscape. Instead, it has proposed a structured pathway toward achieving genuine operational resilience through a multi-layered framework and a strategic implementation roadmap.

7 Implications, Limitations, and Future Research

7.1 Theoretical and Managerial Implications

The primary contribution of this chapter is twofold. For academics, it offers a synthesized conceptual framework that bridges disparate fields of research, including AI security, data governance, privacy-enhancing technologies, and strategic management within the specific context of critical A&D systems. By structuring the challenges and solutions into interdependent layers, it provides a model that can be used as a basis for future empirical investigation into the efficacy of different security and governance strategies.

For managers and practitioners, the implications are direct and actionable. The framework and strategic roadmap provide the necessary tools to *shift organizational strategy from a narrow focus on performance to a deep commitment to assurance*. They serve as a practical guide for assessing readiness, architecting for security, and leading the necessary cultural changes. The central managerial takeaway is that investing in

secure and ethical data management for AI should be viewed as a strategic enabler. This approach builds trust, ensures mission assurance, and ultimately underpins sustainable competitive advantage in a data-driven battlespace.

7.2 Limitations of the Chapter

This chapter has several limitations that should be acknowledged. First, while the proposed framework is deeply informed by the author's primary empirical experience within the A&D sector, this chapter presents these insights in a synthesized, conceptual format. It does not constitute a formal empirical study with detailed, replicable data analysis as one might find in a peer-reviewed journal article. Second, the case studies are presented as illustrative archetypes designed to demonstrate the framework's application in realistic scenarios. While grounded in real-world challenges and insights, they are synthesized for clarity and do not follow a formal, in-depth case study methodology. Third, the technological landscape, particularly concerning privacy-enhancing technologies and adversarial AI techniques, is evolving at a rapid pace. The specific technologies discussed represent a snapshot in time and will undoubtedly be supplemented by new innovations. Finally, the chapter's focus is primarily on the regulatory and operational contexts of Western and allied nations; the framework would require adaptation for different geopolitical and legal environments.

7.3 Future Research Directions

The limitations of this chapter point toward several promising avenues for future research. There is a clear need for empirical studies to validate and refine the proposed framework. Future work could test the *hypothesis that organizations with higher measured adoption of this framework exhibit statistically lower rates of AI-related security incidents and faster accreditation times.* Such research could take the form of in-depth case studies of organizations that have successfully implemented secure AI systems, identifying key success factors and barriers.

Quantitative research could explore the performance trade-offs and return on investment associated with implementing different layers of the framework, answering the question: *What is the quantifiable impact of investing in the technological layer (e.g., PETs) on the reduction of data breach risk in a simulated coalition environment?* Further research is also needed on the human and cultural factors of AI adoption, testing the *hypothesis that leadership training programs focused on the principles in Section 6.3 lead to a measurable increase in security-conscious behaviors among engineering teams.* Finally, technical research into the scalability and robustness of PETs in resource-constrained, tactical-edge environments remains a critical area for continued investigation.

Ultimately, for the aerospace and defense sector, mastering the complexities of data management and privacy is more than a risk mitigation exercise. It is the essential foundation upon which the transformative potential of AI can be responsibly and securely built, ensuring a lasting strategic advantage through operational resilience in an increasingly complex and contested world.

References

Section 1

American Action Forum (AAF). (2024). AI export controls: Balancing national security and AI innovation. https://www.americanactionforum.org/insight/ai-export-controls-balancing-national-security-and-ai-innovation/.

Belfer Center. (2019). Attacking artificial intelligence: AI's security vulnerability. https://www.belfercenter.org/publication/AttackingAI.

Beyza, U., & Ulysse, R. (2024). Governance of artificial intelligence in the military domain. The United Nations Office for Disarmament Affairs (UNODA). https://digitallibrary.un.org/record/4062924/files/1417181-EN.pdf.

Bureau of Industry and Security. (2022). Commerce implements new export controls on advanced computing and semiconductor manufacturing items to the People's Republic of China (PRC). https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file.

Center for Security and Emerging Technology (CSET). (2023). For export controls on AI, don't forget the "catch-all" basics. https://cset.georgetown.edu/article/dont-forget-the-catch-all-basics-ai-export-controls/.

Daniel, A., & Meg, K. (2022). The impact of artificial intelligence on military defence and security. Centre for International Governance Innovation (CIGI). https://www.cigionline.org/static/documents/no.263.pdf.

Deloitte. (2024). 2025 aerospace and defense industry outlook. https://www.deloitte.com/us/en/insights/industry/aerospace-defense/aerospace-and-defense-industry-outlook.html.

GE Aerospace. (2017). Big data has landed in the aviation industry. https://www.geaerospace.com/news/articles/technology/big-data-has-landed-aviation-industry.

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44–58. doi:10.1609/aimag.v40i2.2850.

HCLTech. (2024). Unlocking new possibilities with AI in aerospace and defense. https://www.hcltech.com/white-papers/ai-in-aerospace-and-defense-unlocking-new-possibilities.

IEEE Spectrum. (2020). How adversarial attacks could destabilize military AI systems. https://spectrum.ieee.org/adversarial-attacks-and-ai-systems.

Kaczmarek, S. (2025). A bio-inspired hierarchical temporal defense for securing spiking neural networks against physical and adversarial perturbations. *Proceedings of the NeurIPS 2025 Workshop on Machine Learning and the Physical Sciences (ML4PS)*.

McKinsey. (2023). Scaling complex manufacturing in the aerospace industry. https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/taking-off-scaling-complex-manufacturing-in-the-aerospace-industry.

McKinsey. (2024). Overcoming challenges in aerospace procurement. https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/overcoming-challenges-in-aerospace-procurement.

Military Embedded Systems. (2021). Enabling AI at the tactical edge. https://militaryembedded.com/ai/cognitive-ew/enabling-ai-at-the-tactical-edge.

MIT Technology Review. (2019, June). The Pentagon has a laser that can identify people from a distance—by their heartbeat. https://www.technologyreview.com/2019/06/27/238884/the-pentagon-has-a-laser-that-can-identify-people-from-a-distanceby-their-heartbeat/.

National Institute of Standards and Technology. (2024). Protecting controlled unclassified information in nonfederal systems and organizations (NIST Special Publication 800–171 Rev. 3). https://csrc.nist.gov/pubs/sp/800/171/r3/final.

National Security Commission on Artificial Intelligence. (2021). Final report. U.S. Government Printing Office. https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf.

Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. P. (2018). Sok: Security and privacy in machine learning. *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, 399–414. doi:10.1109/EuroSP.2018.00035.

Philosophy & Technology. (2021). Ethical principles for artificial intelligence in national defence. https://link.springer.com/article/10.1007/s13347-021-00482-3.

Precedence Research. (2024). AI in aerospace and defense market size, share & growth report, 2034. https://www.precedenceresearch.com/ai-in-aerospace-and-defense-market.

PwC. (2025). Global aerospace and defense: Annual industry performance and outlook. PricewaterhouseCoopers. https://www.pwc.com/us/en/industries/industrial-products/library/assets/pwc-global-aerospace-and-defense-annual-industry-performance-and-outlook.pdf.

Reuters. (2013). Theft of F-35 design data is helping U.S. adversaries -Pentagon. https://www.reuters.com/article/business/theft-of-f-35-design-data-is-helping-us-adversaries-pentagon-idUSL2N0EV0T3/.

Sharetru. (2023). ITAR compliance: Understanding the basics. https://www.sharetru.com/blog/itar-compliance-guide/.

Simply NUC. (2025). Edge AI for government: Real-time decision making at the tactical level. https://simplynuc.com/blog/edge/edge-ai-for-government-real-time-decision-making-at-the-tactical-level/.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, 361(6404), 751–752. doi:10.1126/science.aat5991.

The Diplomat. (2017). What does the latest F-35 data breach teach us about defense industrial espionage? https://thediplomat.com/2017/10/what-does-the-latest-f-35-data-breach-teach-us-about-defense-industrial-espionage/.

U.S. Department of Defense. (2020). DOD adopts ethical principles for artificial intelligence. https://www.defense.gov/News/Releases/release/article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

Waharte, S., & Trigoni, N. (2010). Supporting search and rescue operations with UAVs. *2010 International Conference on Emerging Security Technologies*, 142–147. doi:10.1109/EST.2010.31.

Washington Technology. (2025). The future of defense: AI at the edge for real-time tactical advantage. https://www.washingtontechnology.com/sponsors/sponsor-content/2025/03/future-defense-ai-edge-real-time-tactical-advantage/403996/.

Section 2

Cloud Security Alliance. (2024). Security guidance for critical areas of focus in cloud computing (Version 5). https://cloudsecurityalliance.org/artifacts/security-guidance-v5.

Deni, J. R., MacLeod, M. R., Stewart, S. E., Banko, K. M., & Jones, A. (2025). Measuring interoperability within NATO: Adapted off-the-shelf or bespoke solution? *Parameters*, 55(1). doi:10.55540/0031-1723.3331.

Department of Defense. (2019). DoD instruction 5400.11: Privacy and civil liberties programs (Change 1). Office of the Assistant to the Secretary of Defense. https://www.esd.whs.mil/Directives/issuances/dodi/540011p.pdf.

Department of Defense. (2022). DoD 5200.01: DoD information security program. https://exwc.navfac.navy.mil/Portals/88/Documents/EXWC/DoD_Locks/PDFs/DODM-5200-01-vol3-AFI-16-1404.pdf.

Galvin, K. (2025). Advancing modelling and simulation in NATO federated mission networking. *Proceedings of the NATO STO Modelling and Simulation Group (MSG-207) Meeting*. NATO Science and Technology Organization. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-MSG-207/MP-MSG-207-07.pdf.

Gormus, E. (2024). NATO's artificial intelligence strategy and interoperability challenges: The case of Turkey. *Journal of Balkan and Near Eastern Studies*. doi:10.1080/19448953.2024.2414174.

Jin, X., Katsis, C., Sang, F., Sun, J., Kundu, A., & Kompella, R. (2022). Edge security: Challenges and issues. *arXiv*. doi:10.48550/arXiv.2206.07164.

Kissel, R., Regenscheid, A., Scholl, M., & Stine, K. (2014). Guidelines for media sanitization (NIST Special Publication 800–888 Revision 1). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/88/r1/final.

Kufakunesu, R., Myburgh, H., & De Freitas, A. (2025). The internet of battle things: A survey on communication challenges and recent solutions. *Discover Internet of Things*, 5, 226–246. https://link.springer.com/article/10.1007/s43926-025-00093-w.

Morgan, F. E., Boudreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K., & Grossman, D. (2020). Military applications of artificial intelligence: Ethical concerns in an uncertain world (Research Report RR-3139). RAND Corporation. https://www.researchgate.net/publication/340991921_Military_Applications_of_Artificial_Intelligence_Ethical_Concerns_in_an_Uncertain_World.

Papernot, N., McDaniel, P. D., Jha, S., Fredrikson, M., Celik, Z. B., & Swami, A. (2015). The limitations of deep learning in adversarial settings. *arXiv*. https://arxiv.org/abs/1511.07528.

Pettai, M., & Laud, P. (2015). Combining differential privacy and secure multiparty computation (Cryptology ePrint Archive, Paper 2015/598). https://eprint.iacr.org/2015/598.pdf.

Rettore, P. H. L., Zißner, P., Alkhowaiter, M., Zou, C., & Sevenich, P. (2024). Military data space: Challenges, opportunities, and use cases. *IEEE Communications Magazine*, 62(1), 70–76. doi:10.1109/MCOM.001.2300396.

Ristenpart, T., Tromer, E., Shacham, H., & Savage, S. (2009). Hey, you, get off of my cloud: Exploring information leakage in third-party compute clouds. *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS)*, 199–212. https://rist.tech.cornell.edu/papers/cloudsec.pdf.

Schellman. (2024, August). What are the 2024 DoD cloud computing security requirements (SRG) for NIST 800–53 revision 5? Schellman Blog. https://www.schellman.com/blog/federal-compliance/the-dod-csp-srg-and-dod-mo-srg.

Tonge, P., Jiwane, N., & Rakhade, V. (2024). Security challenges in edge computing. *International Journal of Futuristic Innovation in Arts, Humanities and Management*, 3(1), 431–442. doi:10.59367/0a04qv85.

U.S. Government Accountability Office. (2021). Weapon systems cybersecurity: DOD just beginning to grapple with scale of vulnerabilities (GAO-21–179). https://www.gao.gov/assets/720/712828.pdf.

Zabala-López, A., Linares-Vásquez, M., Haiduc, S., & Donoso, Y. (2024). A survey of data-centric technologies supporting decision-making before deploying military assets. *Defence Technology*, 42, 226–246. https://www.sciencedirect.com/science/article/pii/S221491472400182X.

Zhou, S., Liu, C., Ye, D., Zhu, T., Zhou, W., & Yu, P. S. (2022). Adversarial attacks and defenses in deep learning: From a perspective of cybersecurity. *ACM Computing Surveys*, 55(8), 1–39. doi:10.1145/3547330.

Ziani, J. (2022). How differential privacy impacts data elicitation. *ACM SIGecom Exchanges*, 20 (2), 75–81. doi:10.1145/3699804.3699811.

Section 3

Aerospace Industries Association. (2018). Critical security controls for effective capability in cyber defense (NAS9933). https://store.accuristech.com/standards/aia-nas9933?product_id=2603292.

Australian Signals Directorate. (2025). Essential eight. https://www.cyber.gov.au/resources-business-and-government/essential-cybersecurity/ism.

Biden, J. R. (2023). Executive Order 14105—Addressing United States investments in certain national security technologies and products in countries of concern. https://www.presidency.ucsb.edu/documents/executive-order-14105-addressing-united-states-investments-certain-national-security.

Biden, J. R. (2024). Executive Order 14117—Preventing access to Americans' bulk sensitive personal data and United States government-related data by countries of concern. https://www.presidency.ucsb.edu/documents/executive-order-14117-preventing-access-americans-bulk-sensitive-personal-data-and-united.

Department of Defense. (2020a). DOD adopts 5 principles of artificial intelligence ethics. https://www.defense.gov/News/News-Stories/article/article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/.

Department of Defense. (2020a). DOD adopts ethical principles for artificial intelligence. https://www.defense.gov/News/Releases/release/article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

Department of Defense. (2023). DOD releases AI adoption strategy. https://www.defense.gov/News/News-Stories/Article/Article/3578219/dod-releases-ai-adoption-strategy/.

Department of Defense. (2025). Cybersecurity Maturity Model Certification (CMMC). https://dodcio.defense.gov/cmmc/About/.

EU AI Act. (2025). Export control implications for AI technologies. https://artificialintelligenceact.eu/.

European Commission. (2024). AI Act: Regulatory framework for AI. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Parliament. (2023). EU AI Act: First regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

European Union. (2024). Artificial Intelligence Act. https://artificialintelligenceact.eu/the-act/.

European Union Aviation Safety Agency. (2023). Artificial intelligence roadmap 2.0. https://www.easa.europa.eu/en/domains/research-innovation/ai.

National Institute of Standards and Technology. (2024). Protecting controlled unclassified information in nonfederal systems and organizations (NIST Special Publication 800–171 Rev. 3). U.S. Department of Commerce. https://csrc.nist.gov/pubs/sp/800/171/r3/final.

NATO. (2024). Summary of NATO's revised artificial intelligence (AI) strategy. https://www.nato.int/cps/en/natohq/official_texts_227237.htm.

North Atlantic Treaty Organization. (2024). *Revised AI strategy*. https://www.nato.int/cps/en/natohq/official_texts_227237.htm.

Office of the Assistant Secretary of Defense for Industrial Base Policy. (2025). Request for information (RFI) on defense industrial base (DIB) adoption of artificial intelligence (AI) summary and analysis report. U.S. Department of Defense. https://www.businessdefense.gov/ibr/pat/docs/AI-RFI-Summary-Report.pdf.

State of California. (2018). California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=.

UK Ministry of Defence. (2021). Cyber security for defence suppliers (Def Stan 05–138 Issue 4). https://www.gov.uk/government/publications/cyber-security-for-defence-suppliers-def-stan-05-138.

U.S. Department of Commerce, Bureau of Industry and Security. (2025). Export Administration Regulations (EAR). https://www.bis.gov/regulations/ear.

U.S. Department of State. (2025). The International Traffic in Arms Regulations. https://www.pmddtc.state.gov/ddtc_public/ddtc_public?id=ddtc_kb_article_page&sys_id=24d528fddbfc930044f9ff621f961987.

Section 4

Barrett, A. M., Jackson, K., Murphy, E. R., Madkour, N., & Newman, J. (2024). Securing AI model weights: Preventing theft and misuse of frontier models. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2849-1.html.

Bennet, K., et al. (2025). Implementing AI bill of materials (AI BOM) with SPDX 3.0: A comprehensive guide to creating AI and dataset bill of materials. *arXiv*. https://arxiv.org/abs/2504.16743.

Blackman, R. (2022, July–August). Why you need an AI ethics committee. *Harvard Business Review*. https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee.

CISA, NSA, FBI, ACSC, NCSC-UK, CCCS, BSI, NCSC-NL, CERT NZ, NCSC-NZ. (2023). Shifting the balance of cybersecurity risk: Principles and approaches for security-by-design and -default. Cybersecurity and Infrastructure Security Agency. https://www.cisa.gov/sites/default/files/2023-04/principles_approaches_for_security-by-design-default_508_0.pdf.

Davies, M., Srinivasa, N., Lin, T.-H., et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1), 82–99. https://ieeexplore.ieee.org/document/8259423.

Department of Defense. (2020). DoD data strategy. U.S. Department of Defense. https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF.

Department of Defense. (2023). Directive 3000.09: Autonomy in weapon systems. Official DoD Publication. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

Department of Defense. (2024). Cybersecurity Maturity Model Certification (CMMC) program. *Federal Register*, 89(198), 66268–66347. https://www.federalregister.gov/documents/2024/10/15/2024-22905/cybersecurity-maturity-model-certification-cmmc-program.

Hu, V. C., Ferraiolo, D., Kuhn, R., Schnitzer, A., Sandlin, K., Miller, R., & Scarfone, K. (2014). Guide to attribute based access control (ABAC) definition and considerations (SP 800–162). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/162/upd2/final.

Kore, A., Abbasi Bavil, E., Subasri, V., Abdalla, M., Fine, B., Dolatabadi, E., & Abdalla, M. (2024). Empirical data drift detection experiments on real-world medical imaging data. *Nature Communications*, 15(46142). https://www.nature.com/articles/s41467-024-46142-w.

Lutes, N., Nadendla, V. S. S., & Krishnamurthy, K. (2024). Few-shot transfer learning for individualized braking intent detection on neuromorphic hardware. *arXiv*. https://arxiv.org/abs/2408.03336.

Marshall, A., Parikh, J., Kiciman, E., & Siva Kumar, R. S. (2019). Threat modeling for AI/ML systems and dependencies. Microsoft Security. https://learn.microsoft.com/en-us/security/engineering/threat-modeling-aiml.

National Institute of Standards and Technology. (2001). Security requirements for cryptographic modules (FIPS PUB 140–142). https://csrc.nist.gov/pubs/fips/140-2/upd2/final.

National Institute of Standards and Technology. (2022). NIST announces first four quantum-resistant cryptographic algorithms. https://www.nist.gov/news-events/news/2022/07/nist-announces-first-four-quantum-resistant-cryptographic-algorithms.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://www.nist.gov/itl/ai-risk-management-framework.

National Institute of Standards and Technology. (2024a). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST AI 100–102 E2023). https://csrc.nist.gov/pubs/ai/100/2/e2023/final.

National Institute of Standards and Technology. (2024b). Module-lattice-based digital signature standard (FIPS 204). https://csrc.nist.gov/pubs/fips/204/final.

National Institute of Standards and Technology. (2024c). Module-lattice-based key-encapsulation mechanism standard (FIPS 203). https://csrc.nist.gov/pubs/fips/203/final.

National Institute of Standards and Technology. (2024d). Stateless hash-based digital signature standard (FIPS 205). https://csrc.nist.gov/pubs/fips/205/final.

OECD. (2025). Sharing trustworthy AI models with privacy-enhancing technologies. https://www.oecd.org/en/publications/sharing-trustworthy-ai-models-with-privacy-enhancing-technologies_a266160b-en.html.

OWASP Foundation. (2025). GenAI red teaming guide. https://genai.owasp.org/2025/01/22/announcing-the-owasp-gen-ai-red-teaming-guide/.

Pelekis, S., et al. (2025). Adversarial machine learning: A review of methods, tools, and critical industry sectors. *Artificial Intelligence Review*, 58(226). doi:10.1007/s10462-025-11147-4.

Perez, E., et al. (2022). Red teaming language models with language models. *arXiv*. https://arxiv.org/abs/2202.03286.

RAND Corporation. (2024). A playbook for securing AI model weights. https://www.rand.org/pubs/research_briefs/RBA2849-1.html.

Rando, J., Zhang, J., Carlini, N., & Tramèr, F. (2025). Adversarial ML problems are getting harder to solve and to evaluate. *arXiv*. https://arxiv.org/abs/2502.02260.

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero trust architecture (SP 800–207). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/207/final.

Sabt, M., Achemlal, M., & Bouabdallah, A. (2015). Trusted execution environment: What it is, and what it is not. *2015 IEEE Trustcom/BigDataSE/ISPA*, 1, 57–64. doi:10.1109/Trustcom.2015.357.

Sandhu, R., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *IEEE Computer*, 29(2), 38–47. https://csrc.nist.gov/csrc/media/projects/role-based-access-control/documents/sandhu96.pdf.

Schuett, J., Reuel, A. K., & Carlier, A. (2024). How to design an AI ethics board. *AI and Ethics*, 5, 863–881. doi:10.1007/s43681-023-00409-y.

Schuman, C. D., Kulkarni, S. R., Parsa, M., Mitchell, J. P., Date, P., & Kay, B. (2022). Opportunities for neuromorphic computing algorithms and applications. *Nature Computational Science*, 2, 10–19. doi:10.1038/s43588-021-00184-y.

The Royal Society. (2023). From privacy to partnership: The role of privacy enhancing technologies in data governance and collaborative analysis. https://royalsociety.org/news-resources/projects/privacy-enhancing-technologies/.

Thomas, G. (2006). The DGI data governance framework. *The Data Governance Institute*. https://www.datagovernance.com/the-dgi-data-governance-framework/.

Wiz. (2025). AI-BOM: Building an AI-bill of materials. https://www.wiz.io/academy/ai-bom-ai-bill-of-materials.

Section 5

Amazon Web Services. (2023). *AWS Nitro Enclaves*. https://docs.aws.amazon.com/enclaves/latest/user/nitro-enclave.html.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2016). Practical secure aggregation for federated learning on user-held data. *arXiv*. https://arxiv.org/abs/1611.04482.

Chen, Y., Liu, X., Zhang, W., & Zhao, H. (2025). Fatigue detection method for UAV remote pilot based on YOLOv8. *2024 IEEE 5th International Conference on Electronic Communication and High-speed Circuits (ECCE)*, 545–551. doi:10.1109/ECCE59342.2024.10988450.

Department of Defense. (2019). DoD Instruction 6025.18: Health Insurance Portability and Accountability Act (HIPAA) privacy rule compliance in DoD health care programs. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/602518p.pdf.

Director of National Intelligence. (2024). Intelligence Community Directive 503: Intelligence Community Information Environment Risk Management. https://www.dni.gov/files/documents/ICD/ICD-503.pdf.

Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, P., Ricardo, K., Weng, H., Moore, J. C., Hewett, M., Williams, G., & Lau, C. L. (2021). Differential privacy for public health data: An

innovative tool to optimize information sharing while protecting data confidentiality. *Patterns*, 2(12), 100366. doi:10.1016/j.patter.2021.100366.
Federal Aviation Administration. (2023). Guide for Aviation Medical Examiners: Privacy of medical information. https://www.faa.gov/ame_guide/app_process/general/privacy.
H2020 ReMAP Consortium. (2022). H2020 ReMAP project: Real-time condition-based maintenance for adaptive aircraft maintenance planning. https://h2020-remap.eu/.
Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104–191. (1996). https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf.
Jin, W., Yao, Y., Han, S., Joe-Wong, C., Ravi, S., Avestimehr, S., & He, C. (2023). FedML-HE: An efficient homomorphic-encryption-based privacy-preserving federated learning system. *arXiv*. https://arxiv.org/abs/2303.10837.
Landau, D., et al. (2025). Federated learning framework for collaborative remaining useful life prognostics: An aircraft engine case study. *arXiv*. https://arxiv.org/abs/2506.00499.
OpenMined. (2024). Secure enclaves for AI evaluation. https://openmined.org/blog/secure-enclaves-for-ai-evaluation/.
Rahman, M., & Morshed, B. I. (2025). Resource-constrained on-chip AI classifier for beat-by-beat real-time arrhythmia detection with an ECG wearable system. *Electronics*, 14(13), 2654. https://www.mdpi.com/2079-9292/14/13/2654.
Rathi, V. K., Rajput, N. K., Mishra, S., Grover, B. A., Tiwari, P., Jaiswal, A. K., & Hossain, M. S. (2021). An edge AI-enabled IoT healthcare monitoring system for smart cities. *Computers & Electrical Engineering*, 96, 107524. https://www.sciencedirect.com/science/article/abs/pii/S0045790621004699.
Shevlane, T. (2022). Structured access: An emerging paradigm for safe AI deployment. *arXiv*. https://arxiv.org/abs/2201.05159.
Tsai, I. H., & Morshed, B. I. (2025). Signal-specific and signal-independent features for real-time beat-by-beat ECG classification with AI for cardiac abnormality detection. *Electronics*, 14(13), 2509. https://www.mdpi.com/2079-9292/14/13/2509.
Vincze, M., Molnár, B., & Kozlovszky, M. (2025). Quality assurance of the whole slide image evaluation in digital pathology: State of the art and development results. *Electronics*, 14(10), 1943. https://www.mdpi.com/2079-9292/14/10/1943.
Waterman, S. (2025). Army blocks Air Force's AI program over data security concerns. *Air & Space Forces Magazine*. https://www.airandspaceforces.com/fearing-data-leaks-army-blocks-air-force-ai-program-from-its-networks/.

Section 6

Beer Mohamed, M. I., Hassan, M. F., Safdar, S., & Saleem, M. Q. (2021). Adaptive security architectural model for protecting identity federation in service oriented computing. *Journal of King Saud University – Computer and Information Sciences*, 33(5), 580–592. doi:10.1016/j.jksuci.2019.03.004.
Boston Consulting Group. (2025). Three truths about AI in aerospace and defense. https://www.bcg.com/publications/2025/three-truths-about-ai-in-aerospace-and-defense.
Cimphony AI. (2024). AI incident response plans: Checklist and best practices. https://www.cimphony.ai/insights/ai-incident-response-plans-checklist-and-best-practices.
CMMI Institute. (2014). Data management maturity model introduction. https://cdn.ymaws.com/www.globalaea.org/resource/collection/68814379-BF7E-41C8-B152-18A617F9C0AA/Data_Management_Maturity_Model_Introduction_-_Dec_12_2014.pdf.
Condon, M. (2022). The continuous monitoring lifecycle: 7 steps for building a strong foundation. AuditBoard. https://auditboard.com/blog/continuous-monitoring-lifecycle.
Corradini, I. (2020). *Building a cybersecurity culture in organizations: How to bridge the gap between people and digital technology*. Springer. doi:10.1007/978-3-030-43999-6.

Kotter, J. P. (2012). Leading change. Harvard Business Review Press. https://www.amazon.co.uk/Leading-Change-New-Preface-Author/dp/1422186431.

National Institute of Standards and Technology. (2011). Information security continuous monitoring (ISCM) for federal information systems and organizations (NIST Special Publication 800–137). https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-137.pdf.

National Institute of Standards and Technology. (2020). Zero trust architecture (NIST Special Publication 800–207). https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf.

National Institute of Standards and Technology. (2022). Implementation of DevSecOps for a microservices-based application with service mesh (NIST Special Publication 800–204C). https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-204C.pdf.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100–101). https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

Radiant Security. (2024). AI-driven incident response: Definition and components. https://radiantsecurity.ai/learn/ai-incident-response/.

Weill, P., Woerner, S. L., & Sebastian, I. M. (2024). Building enterprise AI maturity. MIT Center for Information Systems Research. https://cisr.mit.edu/publication/2024_1201_EnterpriseAIMaturityModel_WeillWoernerSebastian.

Section 7

National Academies of Sciences, Engineering, and Medicine. (2022). Human-AI teaming: State-of-the-art and research needs. The National Academies Press. doi:10.17226/26355.

U.S. Department of Defense. (2018). Summary of the 2018 Department of Defense artificial intelligence strategy. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

PART VI

Case Studies—Real-World Lessons: From Theory to Practice across Sectors



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

14

AUGMENTING TRUST

AI and Automation's Impact in Small-Scale Legal Services

Tudor-Andrei Stanciu and Andrei Pavel

Introduction

Most debates on artificial intelligence (AI) in law focus on global firms, not the small practices that serve most clients. This chapter takes the opposite angle, by treating AI and automation not as tools that may be included in any workflow but that only make sense once repeatable work is mapped and stress-tested.

We write from the standpoint of a two-lawyer Romanian boutique that built a semi-automated flow for routine trade registry operations. The work focuses on a simple use-case, is price-sensitive, and repetitive, making it precisely the sort of legal task where existing software tools, structured intake, and a few human checkpoints can turn a classic legal mandate into a productized, fixed-fee service.

Our claim is simple: for small firms, the path to an AI-enabled practice does not have to start with AI. It starts with understanding your own patterns of work, being honest about client expectations, and then letting automation, and, where it truly adds reliability, AI, carry the parts of the journey that do not require legal judgment.

Literature Review

Debates on AI and automation in legal services are dense with surveys, vendor reports, and policy papers, most of them being written from the vantage point of large firms and corporate legal departments. This section maps the main areas covered by that literature and highlights the gaps our Romanian case study is meant to fill.

A first strand concerns the structural shift from a "seller's market" to a "buyer's market" for legal services. Jordan Furlong (Law21, 2017) argues that empowered clients, alternative providers and technology have reconfigured the market so that law firms must rebuild around client needs, pricing transparency and measurable value rather than hours sold. Market reports from Thomson Reuters (Thomson Reuters, 2025) and the ACC (acc.com, 2025) echo this demand-side pressure: corporate clients are pushing for operational efficiency, alternative fee arrangements and demonstrable

DOI: 10.4324/9781003632429-21
This chapter has been made available under a CC-BY-NC-ND 4.0 license

innovation from external counsel, especially in routine work. This literature sets the backdrop for why small firms feel compelled to use AI in any area in their practice, even where their scale and capital base make large-firm strategies unrealistic.

A second, rapidly expanding body of work quantifies AI and automation uptake. Clio's Legal Trends Report (Clio.com, 2024b) suggests that around four-fifths of legal professionals already use some form of AI, that a large share of hourly work is technically subject to automation, and that most clients are at least comfortable with firms using AI tools. Wolters Kluwer's Future Ready Lawyer 2024 report (Wolters Kluwer, 2024) likewise finds that most corporate legal departments and law firms expect AI to materially affect the billable hour and business models over the next few years. The IBA report "The Future Is Now: Artificial Intelligence and the Legal Profession" (IBA, 2024) steps back from vendor marketing and frames AI as both opportunity and systemic risk: it shows widespread but uneven adoption, with larger firms using AI more extensively, predominantly for internal support tasks, and identifies persistent concerns around data governance, bias, and professional ethics. Closer to our focus, Smokeball–s 2025 State of Law report (Smokeball, 2025) finds that AI adoption among small firms and solo practitioners has doubled in two years (from roughly a quarter to over half of firms), driven by pressure to handle routine work, admin and client communication more efficiently.

Another important line of literature looks at pricing and client expectations rather than technology per se. Reports from Clio referenced above document a clear trend towards greater use of fixed fees and other alternative fee arrangements, both in the UK and in continental Europe, as firms react to client demands for cost predictability and transparency. These studies, however, rarely dig into how firms operationalize fixed-fee work: what internal processes are standardized, where human discretion is preserved, and how risk is priced, especially in the SME and startup segments that small boutique firms tend to serve.

Framing Automation and Artificial Intelligence in Legal Services

Automation is the general category; artificial intelligence (AI) is a subset. When initially coined, in the 1940s by a Ford Motor Company manager, "automation" was used to describe the increased use of automatic devices and controls in mechanized production lines. Currently, it refers to the application of machines to tasks once performed by human beings or, increasingly, to tasks that would otherwise be impossible without the help of machines (Britannica, 2025). AI did not appear too far behind, being first coined during a summer research project in the 1950s at Dartmouth College, chosen because it was thought to be neutral to any other theories prevalent at the time in the field of "thinking machines" (Wikipedia, 2025). AI is now widely taught as the study of intelligent agents that perceive and act, with the classic definition initially coined at Dartmouth – "the science and engineering of making intelligent machines" – still referenced.

With all of its subcategories of machine learning, natural language processing, and, more recently, generative models, AI is still just one part of that broader toolbox that automation involves. In the technology management literature, the concept of hyper automation explicitly lists AI and machine learning alongside business process management (BPM), robotic process automation (RPA), low/no-code integration, and

packaged software, all as components of one automation program (Gartner, n.d.). Put differently, AI lives inside automation programs and initiatives, among many other ways at hand to use technology in law firms. Acknowledging this lowers the bar of resources required to implement an automation initiative within a law firm.

Methodology

This chapter adopts a single, practitioner-led case study of a small Romanian law firm that designed and implemented a semi-automated workflow for trade registry procedures. The trade registry use-case was chosen because it combines a highly standardized legal process, visible client pain points, and a strong trust component, making it a useful testing scenario for understanding how far automation can go without eroding perceived quality.

The analysis draws on four main sources of evidence: (i) internal operational data from the automated flow, such as number and type of requests, cycle times from intake to filing, number and nature of client follow-ups, drop-off points in the questionnaire, proportion of files requiring manual intervention; (ii) process items, including intake forms, document templates, automation scenarios, internal checklists, and correspondence with platform providers and partners; (iii) the authors' reflections, captured through structured debriefs between the authors at key milestones (initial design, pilot phase, scaling, and subsequent adjustments), and real-time notes taken during client interactions; and (iv) secondary sources, such as market reports and surveys on AI and technology adoption in small firms and legal services, which provide external benchmarks for client expectations and adoption patterns in comparable jurisdictions.

The study follows a theory-informed analytical strategy. First, we reconstructed the full lifecycle of a trade registry matter before and after automation, mapping each step, actor, and decision point. Second, we compared operational data and practitioner reflections to identify patterns around (i) where automation consistently reduced friction or time, (ii) where human intervention remained critical for client trust, and (iii) how pricing and communication needed to adapt once the workflow was semi-automated. Third, we engaged in pattern-matching against established technology adoption frameworks (in particular, the technology–organization–environment perspective) to explain why certain design choices were made (for example, favouring existing software tools and a partner-led go-to-market strategy) and why others (such as client-facing generative AI) were intentionally postponed.

This perspective comes with clear limitations. The findings are shaped by the specific context of a two-lawyer firm in Romania, focusing on a narrow class of procedures before a particular authority; they cannot be directly extrapolated to larger firms, other jurisdictions, or other types of legal work. The authors are also the designers and operators of the workflow, which may create biases despite our efforts to document failed experiments and trade-offs, not just successes. Operational data are constrained by the scale of the practice and do not allow for robust quantification of client-side business impact beyond cycle times, satisfaction signals, and repeat work. Finally, the regulatory and technological environment is evolving; the observations in this chapter represent a snapshot as of mid-to-late 2025 and may shift as client digital literacy, AI regulation, and trade registry infrastructure change.

An Overview into Our Approach

Prior to this endeavour, we had implemented technology both related and unrelated to automation in other routine tasks, such as e-invoicing and card payments; project management through specialized software; AI-powered email clients on top of our inboxes; bots in our instant messaging internal tools to help us be reminded of different recurring tasks; or email snippets to re-use for repetitive tasks. Working on these internal use-cases led us to a mindset of wanting to find the optimal set of steps to achieve our outcome, when we started working on client-facing improvements.

Focusing on this specific use-case would allow us to be more efficient in already existing processes arising from our existing consultancy clients, thus improving a type of legal service we already provide. It would also expand the addressability of the solution we implement to other clients, that find out about us by using the trade registry automation solution and, as a result, can be clients of our core legal consultancy services.

Our use-case and the way we wanted to approach it using automation, are not new on the Romanian market. This use-case was already being addressed by existing players in the market: law firms or legal services providers that provide legal support; accounting firms that have a legal counsel to support clients with trade registry operations; accounting platforms providing some trade registry services such as entity set-ups with the trade registry. In such a diversified market, without a clear winner but with many providers that had proven the need for these types of services built on top of a legal requirement, we thought about the improvements needed to convince clients to work with us instead of other providers. We decided our solution should aim to improve the current level of solutions available at least on two fronts: (i) personalized customer support, always ultimately handled by human operators. This is one area we consider automation is possible, through e.g. large language model (LLM) training on datasets of customer interactions, to create personalized frequently asked questions (FAQ) for each client, but needs to be addressed with due care, especially for a service that targets a diversity of end-users; and (ii) end-to-end automated process, from client input to confirmation of trade registry approving the change requested, but with human oversight at every intermediary step. Trust is the key factor that legal services providers can showcase, and it's one of the things clients look for when deciding to engage or maintain a relationship with a law firm.

Case Study Context: Local Tailwinds and Constraints in the Romanian Legal Market

We deem it relevant to also look at the broader picture related to digitization of individuals that would be using our solution, and of how digitized the interaction with the local trade registry is. The national office of trade registry in Romania has been supporting online submissions since 2011, via email and a dedicated e-filing portal. The latest update of the portal, made available in August 2024, turned it into a cloud-based application with new features being implemented in real-time. Through the portal, documents may be submitted by a representative of the entity for which changes need to be registered (either a conventional one or one of the appointees the

TABLE 14.1 Technology adoption framework as a decision matrix

Decision point	*Options*	*Selection criteria*	*Choice in this project*
Scope of automation	Internal vs external processes	Internal: efficiency only; External: new revenue/productized service	External client-facing flow
Technology approach	Custom build vs off-the-shelf	Custom: high scale/ differentiation; off-the-shelf: low risk, fast to market	Off-the-shelf tools + light integration
Automation level	Manual vs semi-automated vs fully automated	Depends on pattern stability, authority predictability, and risk tolerance	Semi-automated with human review
AI role	No AI vs internal AI vs client-facing AI	Add AI only after stable workflows; be cautious in regulated, client-facing contexts	No client-facing AI; focus on rules-based automation
Go-to-market	Direct to client vs via partners	Direct: strong marketing capacity; Partners: access to existing client pools	Partner-led distribution via existing ecosystems

Source: Authors' own elaboration.

law allows, such as the director or a proxy mandated specifically for the operation that needs to be performed). All documents submitted to the trade registry must be signed either in wet ink or with a qualified electronic signature according to the eIDAS Regulation and countersigned by the individual making the submission on behalf of the entity.

To address these steps, we conceived our solution to (i) allow the end user and any additional signatories decide how to sign the documents and, if opting for a qualified electronic signature, have the possibility of obtaining one from a partner with a special discount code, and emphasized a qualified electronic signature would be usable for other signing needs of the individual obtaining it; (ii) submit the set of documents signed by the client to the trade registry via automations built in our existing legal services infrastructure, based on human language commands that prompt a chatbot to send the documents to the competent trade registry and referencing the relevant client.

Factors we took into account when deciding on a relevant price, which are applicable in our view for any technical implementation done by a law firm, include: (i) total cost of software solutions we needed to use for our project to be live, both internal and external, such as automation costs or payment processing costs; (ii) price points of other solutions available on the market, especially the ones delivered through similar digital means as ours; we acted as end-users for some of those, to understand the level of automation involved, speed of execution and customer support level to benchmark against them; (iii) relevant discounts we could integrate as an additional selling point for partners we reached out to for integrations; those discounts made sense financially given that we had eliminated marketing costs by selling to an already existing customer base.

Given this rather extensive timeline, even with an online portal at their disposal, we considered potential users of our solution should not incur additional extensive timelines because of our legal flow, as long as the information they submit to us as input is

TABLE 14.2 Cost–benefit analysis of the Romanian trade registry project

Category	*Law firm perspective*	*Client perspective*
Public fees (trade registry)	Mandatory pass-through cost. For most standard operations on a limited liability company, the trade registry fee is approximately **EUR 26** as of August 2025. This cost is fixed and outside the automation scope but must be factored into total price.	Sees the public fee as part of a bundled price. Predictable disclosure of this mandatory cost, together with the firm's fee, improves perceived transparency and trust.
Software & infrastructure	Upfront and recurring cost for: document templates (including those provided by the trade registry), e-filing portal, e-signature, automation layer, payment processing, and storage. These are key inputs in setting a viable flat fee.	Gains access to a complete infrastructure (online intake, document generation, qualified e-signature, submission) without buying tools separately. The complexity is hidden behind the service fee.
Internal lawyer cost	Manual work (collecting data, adapting templates, preparing submission packages, monitoring files) is partially shifted into a structured workflow: standardized intake, document generation, and an automation that sends status updates at each key step. This frees time and thus reduces cost.	Less back-and-forth by email and fewer delays caused by missing information. The client mostly interacts with a guided flow plus human support when needed, rather than with ad-hoc requests for documents and clarifications.
Marketing & distribution	By selling primarily through partners that already aggregate the target clientele, the firm can **reduce direct marketing spend**. The project relies on partner networks and discounts as a selling point instead of heavy brand-building for this narrow use-case.	Finds the solution in channels they already trust (accountants, platforms, other intermediaries). May benefit from preferential pricing or bundled offers negotiated between the firm and partners.
Pricing & revenue model	Automation enables **flat or fixed-fee pricing** for a class of trade registry operations. Price setting explicitly considers software costs, competitor price points and service levels, and discounts offered via partners. This supports more predictable, scalable revenue than purely bespoke work.	Clear expectation of total cost (public fee + law firm fee). Flat pricing makes it easier to compare the service with other providers and reduces anxiety about standard hourly billing.
Process timeline & transparency	The firm cannot compress the statutory timelines (trade registry has up to a few business days to decide, plus additional days to send documents) but can compress and standardize its own handling time. Automation triggers **near real-time status emails** at each internal step and when public status changes.	Even if the objective duration remains constrained by the authority, the client experiences a **more transparent process**: they know when files are submitted, when they are being processed, when decisions are issued, and when documents are dispatched.
Risk & quality control	Standardized templates and flows reduce the risk of missing steps or inconsistent drafting. Human checkpoints remain in the loop for review and for edge cases, which mitigates the risk of over-automation in a regulatory context with 42 different local trade registry offices.	Benefits from a more consistent quality level across similar operations. Built-in controls and human review increase confidence that filings are complete and compliant, compared to purely DIY or low-touch providers.

Source: Authors' own elaboration based on internal data and feedback from clients.

accurate and complete. As of August 2025, our standard delivery term is one business day for each intermediary step that requires our input (initial document generation; submission to the trade registry of client documents; regular checks of the file's status once submitted). In future versions of the tool, we aim to deliver documents in close to real time after the client submitted their request to us.

Bearing in mind those insights we've observed in practice, we realized that a fully automated experience would not be possible. The unpredictability of working with a local authority, the fact that the client has no direct visibility in the status of their file once they submitted all that has been required for them, and the trust that needs to be built into a legal service being provided exclusively online (compared to other traditional legal services), all meant designing ways to inform our clients about any step taken and delays incurred. Given that the sequence of steps is standard, and the reasons for delays are almost standardized, we embedded each step in an automation that, once triggered, sent a specific status update to the client via email. We note that the Romanian trade registry portal provides a searchable function to identify files registered, and see their status (registered, being processed by the trade registry, under resolution, solved, documents issued by the trade registry, issues raised by the trade registry), that can be used by any interested party to check the status of a file, and that we have also replicated in our process.

We wrap up this section by referencing an important constraint in implementing a mass-market, digital solution for a legal need that touches people with different levels of digital skills. Eurostat confirms Romania has the lowest share of adults with at least basic digital skills in the EU (approximatively 28% in 2023), which means our automated journey must include hand-holding artefacts: clear instructions, human customer support that communicates in layman terms, and structured forms that prevent errors (Eurostat, 2023, 2024). This is evident in the fact that based on our internal data and feedback from clients using our automation flow, 6 in 10 of clients accessing our current version of the flow (without submitting a request for a change), and approximately 1 in 10 clients that work with us would like to discuss with a human operator by phone or another instant messaging solution to guide them through the process. In addition, as of August 2025, around 6 in 10 potential users interacting with our flow drop off after learning that the process involves filling in details in a questionnaire. This begs the question of whether any automation project targeted at mass-market individual clients can be fully implemented in a low-digitized environment. At the same time, poor digital skills leads to acute access to justice issues. This pattern of having routine needs unmet or handled informally as a result of process friction being high and digital skills low is exactly the space where well-designed, flat-fee, automated legal solutions can move the needle.

For any legal solution provider that designs a business-to-business (B2B) or business-to-consumer (B2C) flow that involves a public authority, all of the above context matters. On one hand, clients must to be made aware of any limitations of the solution they're using to solve their need, or otherwise the client feedback received could be negative for reasons that cannot be improved without proper communication. On the other, an automation solution needs to understand the system with which it integrates on the public authority side and its limitations, in order to understand how to interact with it, as well as any expectations of the end user over the extent of human customer support required.

Technology Adoption Theoretical Framework

We consider our trade registry automation project as following the technology–organization–environment (TOE) framework (Wikipedia, 2025). On the technology side, the service sits in the "low-complexity, high-compatibility" quadrant, with its structured intake, low-code automation, qualified e-signature, and the existing local trade registry portal. The gain is not that any single tool being used is innovative, but that they are put together into one coherent workflow.

The organization is our two-lawyer boutique law firm with limited capital, no specifically allocated team, and clients who expect partner involvement in the solving of their problems. These constraints favour experiments that are reversible, capable of operation by the existing team, and paid for through the work being automated. Therefore, the choice of a semi-automated service that standardizes and automates the repeatable processes, while also keeping human review on areas that require judgement, potential edge-cases, or reputational risks, is the obvious one.

The environment poses various challenges. The local trade registry provides an electronic submission portal but some case-by-case interpretations of the law make the process cumbersome. On the client side, there are issues arising from low digital literacy and pressure on speed and predictability from the provider, even if the client is not aware of the entire process. The solution space is also crowded, with various providers such accountants, lawyers, and platforms providing similar niches. This requires constant innovation to stay relevant, while also requiring scrutiny in case of over-automation and user experience (UX) that might go ahead of client capabilities of understanding.

A Design Framework for Automation Initiatives in Small Law Firms

Based on our experience, we design below a minimum viable automation plan we would recommend to a small law firm in an emerging market. The guiding principles are: identify the pattern, decide on the approach, set the automation/AI toggle, identify the go-to market.

Identifying the pattern requires you to look for standardized areas of delivery, where either the output created, the process through which the output is created, or both, are repetitive, and where the differences between iterations of the same process only come from the client's input. This identification process should be achieved before deciding on methods and tools for implementation. A workflow should be drafted in plain language: client inputs, decisions to be taken based on client inputs, outputs being provided, service-level agreements (SLAs) made with the client, and other terms of service, and where any legal expert checkpoints need to be. Once the workflow is created, it should be tested in practice without changing the traditional customer process currently being used to solve that need (e.g. asking for information and expecting replies via email or phone). Only when the workflow seems correct and complete, a decision may be made on the technical approach to implement the pattern. In our case, our workflow included the following steps: (i) obtaining relevant information from the end client; making sure the information is complete (i.e. covers all types of scenarios possible related to the client's status prior to the operation), accurate (i.e. no

information request is based on outdated legal provisions, and no information submitted by the client is outdated and/or irrelevant for the operation at hand), and timely (i.e. asked for only if it is related to the operation at hand, rather than requesting all information and then deciding which piece of it is relevant); (ii) design the set of steps that need be taken internally once client data is submitted; this should be refined with each new client that requests a service while delivering it via the current method being used to provide the services, and up until there are no additional changes happening after several client flows undertaken this way; this set of steps should also be designed in the most efficient way possible, if speed of execution is also a factor; (iii) creating the right client expectations over the steps required of the process; this means both accurate (i.e. not mentioning things that the expert knows are not possible only to make the client purchase the service) and clear (i.e. adding explanatory pieces of information at each step of the way, to pre-empt client questions that would hinder the automation process because they need to be addressed manually).

While identifying the pattern and designing the workflow, the goal is eliminating back-and-forth interactions that are the current norm in delivering legal services. But a critical principle at this stage is doing things that do not scale (Graham, 2013), rather than applying automation and AI and, for that matter, any kind of technology that requires implementation, before the pattern is clearly identified and the workflow ideally designed, which can be confirmed by clients obtaining the results they seek for using the workflow.

Deciding on the approach means taking the first step towards technical implementation. We suggest imagining a scale that, at one end, involves not implementing any additional technology than what is currently in place and, at the other end, having an end-to-end fully automated solution. In our view, fully automated solutions for legal needs may be implemented only to the extent they do not erode the trust factor, inherently expected from professional services providers such as lawyers. For this to happen, the pattern and workflow previously designed need to maintain the appearance of a process following legal provisions and leading to an outcome that is valid from a legal standpoint. This is important if the potential clients have obtained legal services of the type being aimed to be deliver through the process (and, as a result, such clients may benchmark what is being provided to them against what they have obtained in the past), but even more so with clients that pose adoption problems for your solution (e.g. low digital skills, sub-optimal or different prior experience with legal services providers – due to result of service, method of service being provided, or customer support). In our experience, a legal services client is more likely to remember a good experience with a legal services they were provided even if that service does not have the expected or implied outcome.

At the same time, a fully manual process to provide a legal service might not be in line with the expectations of end clients. Clients build expectations over different products or services by comparing experiences they deem similar, provided by other businesses, even if those are not from providers you are competing with. At the same time, there's an expectation over how a legal need should be solved is directly connected to how simple, straightforward, and frictionless for an experienced provider, seems to be for the end client.

Deciding on which point of the scale you position your technical solution is a balancing act between customer needs, internal efficiency, and maintaining at least the same level of customer support as without the solution being implemented.

In our specific case and as of August 2025, we have tipped the scale towards more automation of the internal processes, more manual customer support to address corner cases in an as friendly way as possible, while seeking to cover costs with the tools required for technical implementation based on a sub-set of the total number of flows we could process before creating a full go-to market strategy. The process starts with limited automation in the form of a standard way to capture structured data from the client, using branching logic, with webhooks to push input data into our workspace. Once the legal service request is finalized and paid for by the client, internal rules generate the needed set of documents, which are sent to the client, with recurring follow-ups triggered by additional rules to have the client return the signed documents. Once these are received and checked manually for accuracy, they are submitted to the trade registry based on additional triggers. Every client file is indexed as a specific entry in our project management dashboard, and the file's status is always available for our team to check and update, with some of these updates happening automatically.

Setting the AI/automation toggle should happen in two stages. The first involves deciding on the level of proprietary development needed to implement the identified pattern and designed workflow. For an entity that does not develop technology in its regular activity such as law firms, we believe employing specialized resources – e.g. a CTO or an outsourced development team – is not required at the outset and can be added if the implementation scales to relevant revenue numbers that can cover this additional cost. The second step means choosing the right software tools based on both existing tools being used for the legal services which are being provided by the law firm, and new tools that are indispensable for the new experience being designed. We believe that favouring existing tools that have automation capabilities that just need to be turned on should be the approach (e.g. many law firms use the Microsoft 365 suite of tools to conduct their business, so it makes sense to check its automation functionalities first). This also helps in educating existent law firm staff that would be in charge with working with the new solution, as it will be easier for them to use tools they are already accustomed with.

Only after the core use-case arising from the designed flow, and relevant revenue is attained because of the go-to market strategy, does it make sense to think about incremental improvements to an experience that already works. In our case, for instance, these may include, based on the roadmap available as of August 2025: a library of questions asked by past customers, turned into an FAQ that can be accessed for output via a LLM-powered infrastructure, that the client can use in off-hours of customer support; or personalized rather than general follow-ups with the client at certain touchpoints, based on prior conversations with that client. We note that if a proposed automation step does not reliably reduce cycle time or, worse, creates accuracy or trust issues over the delivered output, it should be removed until the right implementation can occur, even if it means reverting to a more manual approach to a task.

Identifying your go-to-market strategy relies on how well-versed the provider's existing team is, in different ways to put an implementation into the hands of users. If

TABLE 14.3 Implementation phases of the trade registry automation project

Phase	*Objective*	*Main activities*	*Key output*
Groundwork	Test basic standardization	Use templates and simple process tweaks internally; observe repeat patterns and client expectations on speed/price	Internal clarity on which workflows are truly repeatable
Use-case selection	Choose the service to productize	Map trade registry operations, required data, templates and authority steps	Defined scope: trade registry changes for Romanian LLCs
Workflow design	Structure client-facing journey	Split process into intake, drafting, signing, filing, follow-up; decide where humans stay "in the loop"	Conceptual client journey with human checkpoints
Build & integrate	Implement semi-automated flow	Configure intake form, document generation, e-signature path, e-filing and status notifications	Working semi-automated flows using off-the-shelf tools
Pricing & go-to-market	Make economics viable	Set flat fees based on costs and benchmarks; design partner-led distribution	Fixed-fee model and partner-based channel strategy
Iteration	Improve and extend	Monitor usage, identify friction points, assess where AI could later add value	Roadmap for incremental optimization and possible AI add-ons

Source: Authors' own elaboration.

a team has successfully launched similar solutions that target the same type of end clients, that experience should be employed. For small law firms that are thinking for the first time about using technology to automate existing flows, those capabilities are not already present in their team. We believe in lowering your go-to market costs by relying on partners that already have access to the market you're targeting and building own channels once the project starts scaling.

In exchange for access to a partner's network, build a commercial model that rewards the partner, either with a commission, a discount on prices for clients coming from that partner, or a fixed licence fee that the partner pays if they implement directly, per client or a specific period of time, with full partner freedom to decide on the final price.

Trust, Professional Judgement, and the Role of Automation

Our journey of building the solution we have described above tells a simple story. Client expectations converge around speed, visibility, and value received; as pricing tilts toward flat fees, following the standardization and increase in speed over routine steps, technology becomes central to more legal services being provided to individuals and business. The reliable way to capitalize on that is to design automation for processes you understand well, because you've executed them in the past many times.

References

Association of Corporate Counsel & Everlaw (2023). The State of Collaboration in Corporate Legal Departments. Retrieved December 9, 2025, from ACC_2023_State_of_Collaboration_Report.pdf.

Association of Corporate Counsel (2025). ACC Chief Legal Officers Survey Key Findings. Retrieved December 9, 2025, from 2025_ACC_CLO_Survey_Key_Findings.pdf.

Best Law Firms/Best Lawyers (2025). Spain Legal Market Report, Inaugural Edition. Retrieved December 9, 2025, from Best Law Firms: Spain Legal Market Report Inaugural Edition.

Bloomberg Law (2025). An Innovation Race Is on to Pick the Top Law Firms of the Future. Retrieved December 9, 2025, from An Innovation Race Is on to Pick the Top Law Firms of the Future.

Britannica (2025). Retrieved December 9, 2025, from Automation | Technology, Types, Rise, History, & Examples | Britannica.

Caserta, S. (2022). New Technologies and Law Firms—An Uneasy Relationship: A European Perspective. *Law, Technology and Humans*, 4(2), 183–196. Retrieved December 9, 2025, from doi:10.5204/lthj.2246.

Clio (2024a). Everything you need to know about AI in the legal industry (Legal Trends hub). Retrieved December 9, 2025, from https://www.clio.com/guides/ai-legal-trends/.

Clio (2024b). Highlights from the Legal Trends Report: the Legal Industry in 2024. Retrieved December 9, 2025, from It's Here: The Newest Legal Trends Report.

Clio (2024c). Is flat fee billing becoming the norm in law? Retrieved December 9, 2025, from https://www.clio.com/guides/flat-fees-legal-trends/.

European Commission, Eurostat (2023). 56% of EU people have basic digital skills. Retrieved December 9, 2025, from https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20231215-3.

European Commission, Eurostat (2024). Digital skills in 2023: Impact of education and age. Retrieved December 9, 2025, from https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240222-1.

European e-Justice Portal (2025). Business registers in EU countries – Romania. Retrieved December 9, 2025, from https://e-justice.europa.eu/topics/registers-business-insolvency-land/business-registers-eu-countries/ro_en.

Furlong, J. (2017). *Law is a Buyer's Market: Building a Client-First Law Firm*. Law21 Press.

Furlong, J. (2023). Imagine there's no partners. And no associates, too. Substack. Retrieved December 9, 2025, from https://jordanfurlong.substack.com/p/imagine-theres-no-partners-and-no.

Gartner (n.d.). Hyperautomation (glossary). Retrieved December 9, 2025, from https://www.gartner.com/en/information-technology/glossary/hyperautomation.

Graham, P. (2013). Do Things That Don't Scale. Retrieved December 9, 2025, from Do Things that Don't Scale.

International Bar Association and the Center for AI and Digital Policy (2024). The Future Is Now: Artificial Intelligence And The Legal Profession. Retrieved December 9, 2025, from document.

LawCareers.net (2025). New research indicates shift in fixed-fee billing. Retrieved December 9, 2025, from https://www.lawcareers.net/Explore/News/New-research-indicates-shift-in-fixed-fee-billing1.

LHH (2025). Retrieved December 9, 2025, from Top Challenges Small Law Firms Face (And How to Overcome Them).

Smokeball (2025). State of Law Report: AI adoption nearly doubles among small firms. Retrieved December 9, 2025, from https://www.smokeball.com/news/smokeball-state-of-law-report.

SRA (Solicitors Regulation Authority) (2025). Sole practitioners' and small firms' use of technology and innovation. Retrieved December 9, 2025, from https://www.sra.org.uk/sra/research-publications/sole-practitioners-small-firms-technology-innovation/.

Thomson Reuters Institute/Law Blog (2025). How AI is transforming the legal profession. Retrieved December 9, 2025, from https://legal.thomsonreuters.com/blog/how-ai-is-transforming-the-legal-profession/.

Thomson Reuters Institute (2025), 2025 Report on the State of the US Legal Market. Retrieved December 9, 2025, from State-of-the-US-Legal-Market-Report-2025.pdf.

Wikipedia (2025a). Dartmouth workshop. Retrieved December 9, 2025, from https://en.wikipedia.org/wiki/Dartmouth_workshop.

Wikipedia (2025b). Technology-organization-environment framework. Retrieved December 9, 2025 from https://en.wikipedia.org/wiki/Technology–organization–environment_framework.

Wolters Kluwer (2022). The Wolters Kluwer Future Ready Lawyer. Retrieved December 9, 2025, from 2022 Future Ready Lawyer Survey Report by Wolters Kluwer.pdf.

Wolters Kluwer (2024). Future Ready Lawyer 2024 (US + 9 EU markets). Retrieved December 9, 2025, from https://www.wolterskluwer.com/en/know/future-ready-lawyer-2024.

15

AI IN THE LEGAL SECTOR

Building Trustworthy Legal AI

Marco Lorenzi and Pierluigi Casale

1 Introduction: AI Adoption and the Legal Industry

Artificial intelligence (AI) has evolved from a collection of experimental applications into a structural component of economic and organizational strategy. Over the past decade, AI has shifted from pilot projects to mainstream deployment, reshaping how businesses operate, make decisions, and deliver value. This transformation represents not only a technological leap but also a cultural one—as organizations integrate AI into everyday processes, governance, and management practices.

Before examining its impact on the legal sector, it is useful to assess the broader landscape of AI adoption across industries. Recent data show a decisive acceleration. According to the 2025 McKinsey Global Survey, 88% of organizations now use AI in at least one business function, up from 78% in 2024 and 55% the year before. However, the same report highlights that most remain in early scaling stages: nearly two-thirds are still piloting rather than deploying AI at enterprise level.

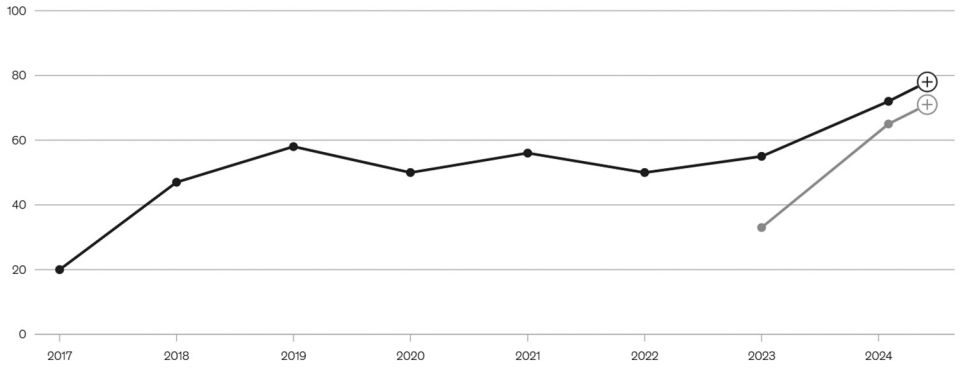


FIGURE 15.1 Organizations using AI in at least one business function (% of respondents)
Source: McKinsey & Company (2025), The State of AI

DOI: 10.4324/9781003632429-22
This chapter has been made available under a CC-BY-NC-ND 4.0 license

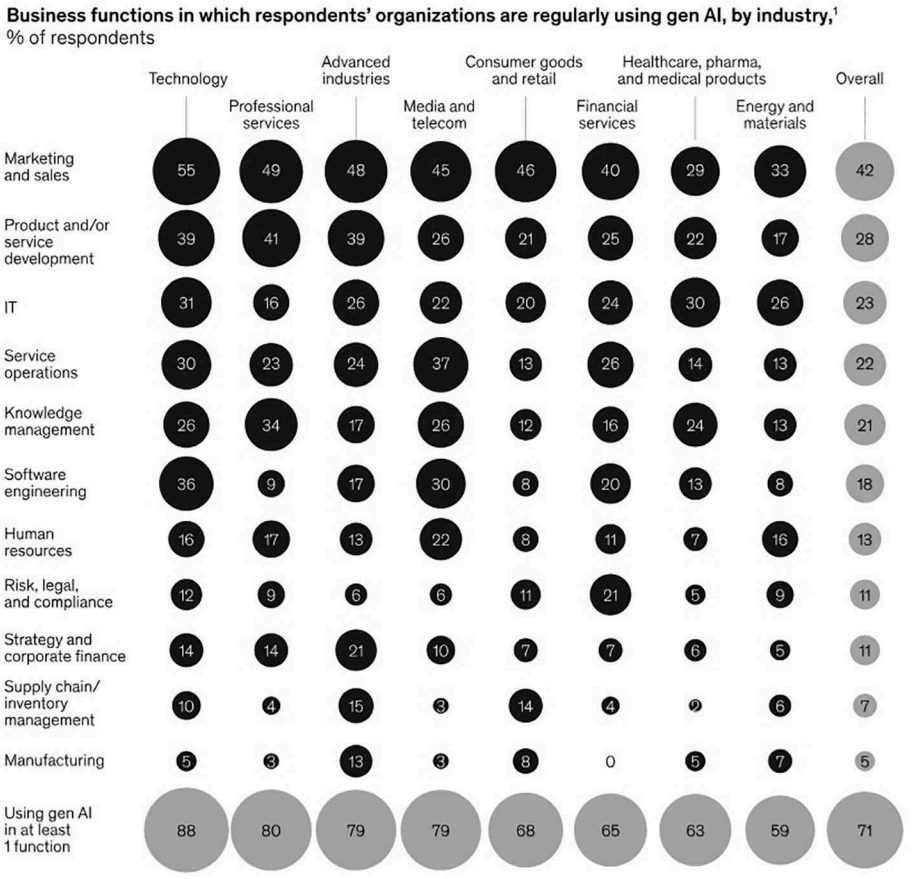


FIGURE 15.2 Business functions in which organizations regularly use gen AI, by industry
Source: Michael Chui, McKinsey Global Survey on AI (2025)

Generative AI (gen AI) has rapidly become part of daily work routines. 71% of McKinsey respondents report using gen AI tools regularly, often across multiple functions such as IT, marketing, and knowledge management. This shift marks a transition from curiosity to integration: AI is no longer a future capability but an active driver of efficiency, innovation, and new business models.

1.1 From General Adoption to Legal Transformation

Against this global backdrop, the legal profession is beginning its own transformation. While traditionally cautious toward technological change, law firms and in-house departments are now experimenting with AI for both operational and analytical tasks. A 2025 benchmarking survey by Counselwell and Spellbook, covering 256 in-house legal professionals (mainly in Canada and the United States), found that 38% of legal teams already use AI tools and another 50% are actively exploring implementation (Ambrogi, 2025).

To summarize the trend:

- **Adoption is expanding**—legal departments are moving from early pilots to regular use.
- **Use cases are concentrated**—contract work dominates (64%), followed by legal research (49%) and translation (38%).
- **Tool diversity is growing**—from general-purpose systems like ChatGPT (74% adoption) to legal-specific platforms such as Spellbook (22%) (Spellbook, 2025).

1.1.1 Key Adoption Barriers

Despite encouraging momentum, adoption remains uneven. Most users report moderate to strong effectiveness (97% cite some benefit), yet major challenges persist. Lack of trust in AI outputs (60%) and data-privacy concerns (57%) top the list of implementation barriers. Only 7% of departments formally measure return on investment, revealing a gap between operational enthusiasm and strategic oversight (Ambrogi, 2025).

These findings mirror broader business concerns identified by McKinsey: organizations are eager to innovate but often lack robust governance, performance tracking, and workflow redesign. In law, this tension is especially acute. Legal work depends on professional judgment, confidentiality, and accuracy—qualities that require both human expertise and reliable systems. Accordingly, 64% of legal professionals expect AI to automate certain tasks, but 66% see it primarily as a *decision-support tool*, not a substitute for legal reasoning.

1.2 From Efficiency to Ethics: The Emerging Tension

This growing adoption brings to the fore a central issue: while AI enhances efficiency and supports complex analytical work, its integration into legal practice raises critical questions of *trust, accountability*, and *ethical compliance*. The challenge for legal professionals is no longer whether to use AI, but how to use it responsibly—ensuring accuracy, safeguarding client data, and maintaining professional integrity.

Observing widespread uptake alone is insufficient. To create genuine value, legal organizations must cultivate transparency, build evaluative metrics, and strengthen governance frameworks. The subsequent sections of this chapter trace this evolution—from the broader trajectory of LegalTech to the design and testing of GiuriMatrix, and finally to the ethical, organizational, and strategic implications of AI in legal contexts.

These lessons extend beyond law, offering insights relevant to any domain where *expertise, trust, and compliance* define professional practice.

2 The Evolution of Legal Technology

Understanding AI's impact on law requires tracing legal technology's evolution. The shift from paper-based systems to AI-driven tools has been gradual, shaped by technological advances, client expectations, and the ongoing drive for efficiency. This

evolution has transformed workflows, professional thinking, and value delivery. In this section, we explore how technological progress and market dynamics converged to create today's AI-driven legal ecosystem.

2.1 From Digitization to AI-Driven Decision Support

LegalTech has evolved in stages. LegalTech 1.0 introduced basic automation, such as e-digitization. Paper archives and ledger books were replaced by document management systems and case-tracking software. Electronic filing and secure client communication became standard, especially during the COVID-19, when courts and firms shifted to remote work, reducing administrative burden without altering the quality of legal reasoning (Wagner, 2020). This stage marked a cultural turning point: digital literacy became a professional necessity rather than a technical curiosity.

LegalTech 2.0 began with e-discovery, automated workflows, and contract drafting. It enabled agreements through blockchain smart contracts, while policymakers introduced frameworks for ethical use and accountability. Automation improved efficiency but still required human review and interpretive skills (Hartung et al., 2018). Strategically, this phase revealed how digital transformation could enhance client service and market competitiveness — yet also underscored the need for clear governance over algorithmic decision-making.

With large language models (LLMs), we are now experiencing LegalTech 3.0. AI is assisting in applications such as statute interpretation, case law summarization, draft arguments generation and preliminary legal advice. Predictive analytics forecast litigation outcomes and assess compliance risks. Yet progress raises serious questions of fairness, transparency, and professional roles. These debates remind us that AI augments legal reasoning with new insight and efficiency, rather than replacing it (Goodenough, 2015). This stage reflects a deeper cultural shift—from viewing technology as a tool to seeing it as a strategic collaborator shaping how justice itself is delivered.

The three stages of LegalTech form a continuum driven by discovery, adoption, and openness to change. They show how swiftly the sector moved from minimal automation to AI, marking a decisive turning point. In the era of LegalTech 3.0, solutions abound, yet ethical and governance questions persist.

Early adopters have translated this rapid evolution into efficiency, new service models, and market advantage. This parallel of technological progress and economic growth sets the stage for examining the sector's expanding market and innovation ecosystem.

2.2 Market Growth and the Innovation Ecosystem

The evolution of LegalTech is mirrored by its rapid market expansion. A 2025 report values the global LegalTech industry at approximately $31.6 billion in 2024, with forecasts projecting growth to $63.6 billion by 2032 — an annual increase of roughly 9.4%. North America currently represents about 36% of the total market, but demand is rising globally as law firms, corporations, and governments adopt AI to manage research, compliance, and contracts more effectively (Fortune Business Insights, 2025, para. 1).

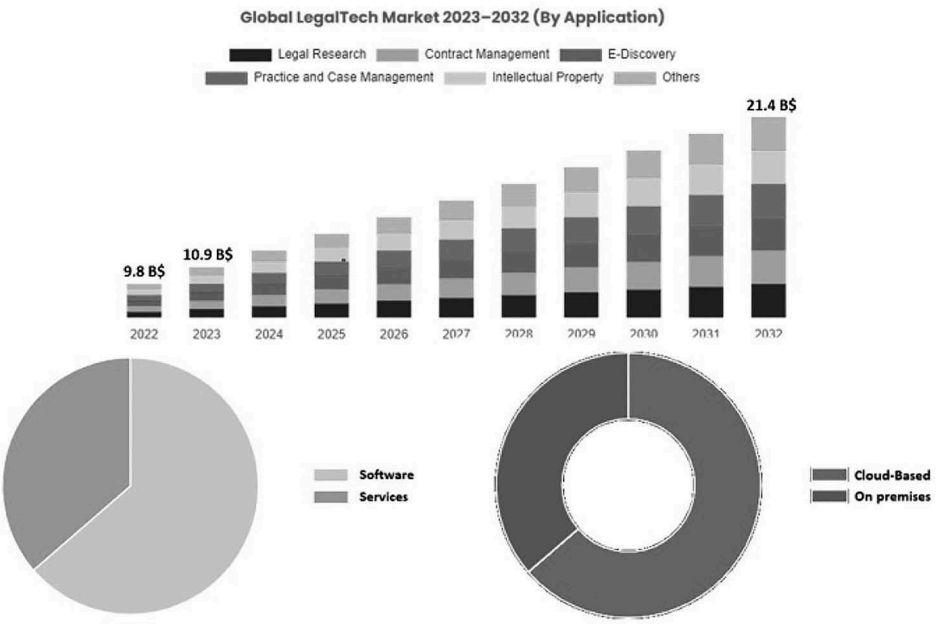


FIGURE 15.3 Global LegalTech market growth, 2023–2032.
Source: S. Miki, "Top Legal Tech Startups in 2024," Clio, 2024.

Generative AI is a central driver of this expansion. By learning from vast textual datasets, gen-AI tools draft documents, extract clauses, and summarize complex legal material. Some analysts estimate that nearly half of all legal tasks could be automated in some form, underscoring both the disruptive and transformative potential of these systems (Fortune Business Insights, 2025).

This transformation has stimulated an innovation ecosystem that now spans the full legal value chain. New start-ups and established players alike are redefining practice boundaries:

- **Paladin** (Paladin, 2025) connects lawyers with pro bono opportunities, expanding access to justice.
- **DoNotPay** (DoNotPay, 2025) empowers citizens to contest fines and negotiate claims without legal representation.
- **Evisort, Ironclad, LawGeex, Casetext** (Evisort, 2025; Ironclad, 2025; LawGeex, 2025; Thomson Reuters, 2023; 2025), and **Kira Systems** (Litera, 2025) automate contract review and compliance.
- **Relativity** is a tool that integrates AI into e-discovery, while **Clio, Litify**, and **HighQ** (Clio, 2025; Litify, 2025; Reuters, 2025) provide cloud-based platforms for firm management.

Together, these developments reveal more than a thriving market — they signal a **strategic and cultural realignment** of the legal profession. The contemporary lawyer operates at the intersection of law, data, and design; success depends as much on

ethical governance and digital fluency as on doctrinal mastery. LegalTech thus becomes a catalyst for redefining what it means to *practice law*: a shift from solitary expertise to networked intelligence, from precedent to prediction, and from reactive service to proactive problem-solving.

3 GiuriMatrix: A Case Study in Context-Aware Legal AI

GiuriMatrix (Casale et al., 2025) is an AI-powered platform supporting legal research, education, and public access in the Italian civil law system. By combining **retrieval-augmented generation (RAG)** with structured legal databases, it delivers contextually accurate answers while highlighting the operational, ethical, and strategic challenges of deploying legal AI.

3.1 Context and Motivation

Legal interpretation carries high stakes: misreading statutes or case law can materially affect outcomes. Ethical controversies like the COMPAS recidivism tool in U.S. courts (Corbett-Davies et al., 2016) show algorithms can systematically misclassify certain groups. GiuriMatrix addresses these risks by emphasizing **accuracy, transparency, and human oversight**, providing natural-language access to Italian law grounded in statutes, case law, and scholarly commentary.

Pro tip: Always combine AI outputs with human verification for critical decisions.

3.2 Data Collection and Preparation

High-quality data was central to GiuriMatrix. The team curated 757 question–answer pairs from Q&A sessions, legal forums, and law school materials, cleaning inconsistencies. Context passages were parsed from concatenated "###" delimiters into separate documents for accurate retrieval (Table 15.1, Figure 15.4). Careful data preparation echoes lessons from COMPAS: biased or incomplete input propagates errors.

In practice:

- Audit datasets for completeness and bias.
- Clean and structure context passages to reduce hallucinations.

3.3 Methodologies: RAG Assessment and Topic Modelling

GiuriMatrix uses RAG assessment and topic modelling to evaluate and understand user interactions. These approaches illuminate both the quality of AI-generated responses and the patterns of user demand.

3.3.1 Evaluating Retrieval-Augmented Generation (RAG)

RAG works in two steps:

1. A **retriever** finds the most relevant documents for a user question.
2. A **generator** uses these documents to craft an answer.

TABLE 15.1 Report on the cleaned and parsed data

Dataset statistics		*Variable Types*	
Number of variables	5	Text	4
Number of observations	757	Categorical	1
Missing cells	769		
Missing cells (%)	20.3%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	29.7 KiB		
Average record size in memory	40.2 B		

Source: Field research.

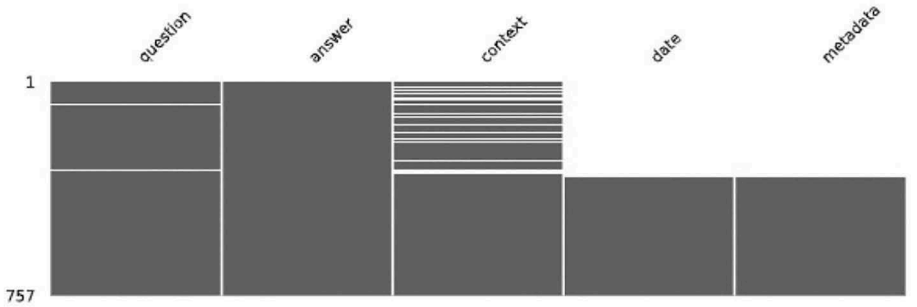


FIGURE 15.4 Parsed data missing values
Source: Field research.

Traditional metrics like BLEU or ROUGE (Santhosh, 2023 measure textual similarity but fail to capture contextual and legal accuracy. GiuriMatrix uses RAGA (retrieval-augmented generation assessment) metrics (Es et al., 2024; Roychowdhury et al., 2024), evaluating:

- **Faithfulness**—proportion of answer statements supported by retrieved context.
- **Answer relevance**—how well the answer addresses the question.
- **Context relevance**—whether retrieved passages contain sufficient information for an accurate answer.

Two generative models were tested: GPT-3.5 Turbo and GPT-4o mini, with experiments on prompt modifications. Experiments have shown that including explicit references to Italian statutory articles (e.g., "Articolo 2043 del Codice Civile") improved the performance in the answer. Evaluation scripts handled batched API calls, error logging, and result storage. The results, summarized in the draft's Table

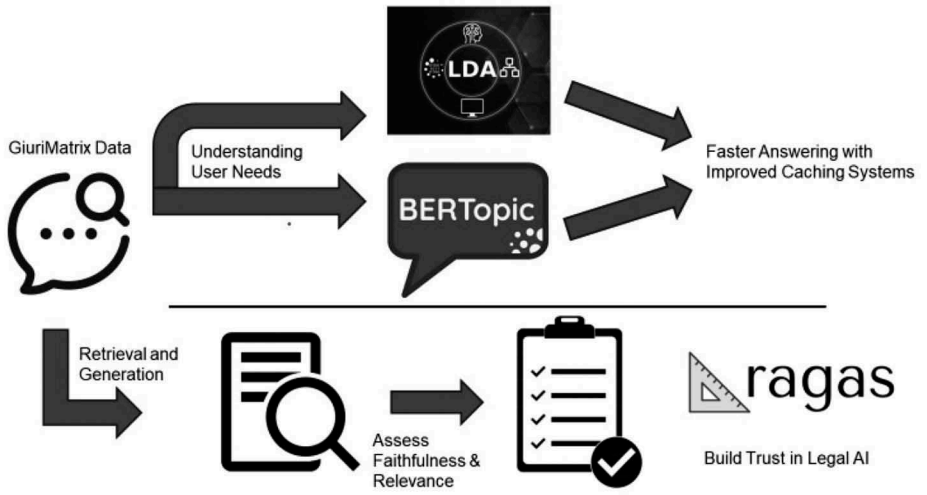


FIGURE 15.5 Schematic representation of the research objectives, illustrating two parallel paths: (1) topic modelling analysis of GiuriMatrix user queries using BERTopic and LDA in parallel; and (2) evaluation of the RAG chatbot with RAGAs metrics—Faithfulness, Answer Relevance, and Context Relevance.
Source: Designed by the authors.

15.2, showed that GPT-4-o-mini with tailored prompts achieved higher faithfulness (0.342) and similar answer relevance (0.669) compared with GPT-3.5 Turbo (faithfulness 0.189; relevance 0.600).

Figures 15.6 and 15.7 show bar charts comparing how well three different AI models perform in terms of giving accurate and relevant answers. The models tested are GPT-3.5 Turbo, GPT-4o mini, and a version of GPT-4o mini with an improved prompt.

These findings highlight the importance of prompt engineering and model choice. They also reveal that cost considerations matter: GPT-4o mini delivered better performance at lower cost due to its improved architecture and token pricing, demonstrating that organizations must weigh accuracy and trustworthiness against economic sustainability.

Pro tips:

- Include statutory references in prompts to improve legal accuracy.
- Evaluate models on both accuracy and cost efficiency.

In practice:

- Automate evaluation scripts for faithfulness, relevance, and context coverage.
- Ensure human review checkpoints for AI outputs.

TABLE 15.2 Summarized results of the RAGAs metrics for each tested model

Model	*Faithfulness*	*Answer Relevance*	*Context Relevance*	*Cost (USD)*
GPT-3.5 Turbo	0.189	0.600	0.018	$3.20
GPT-4o mini	0.336	0.675	Not evaluated	$0.51
GPT-4o mini + Prompt	0.342	0.669	Not evaluated	$0.87

Source: Authors' own elaboration, based on the field research findings.

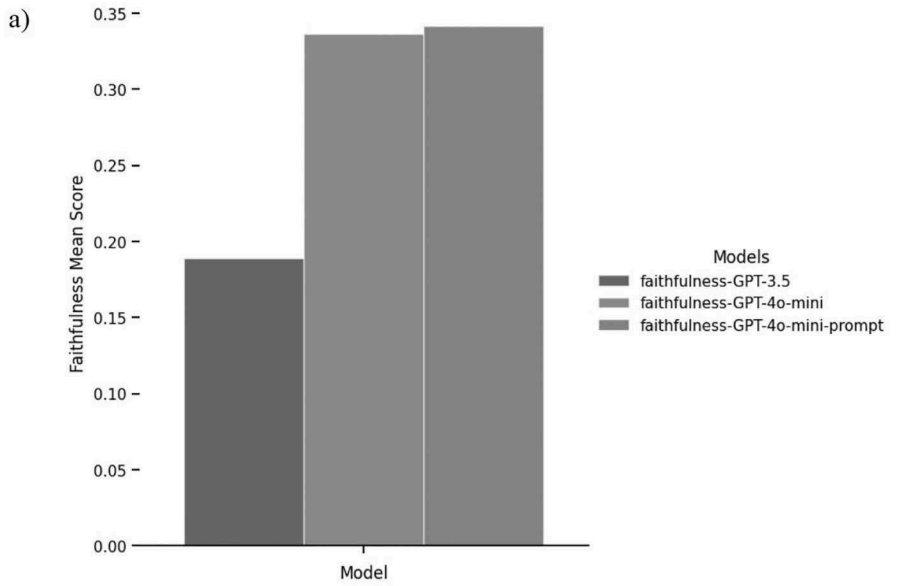


FIGURE 15.6 Faithfulness score comparison across models
Source: Field research.

3.3.2 Understanding User Demand through Topic Modelling

To understand the types of legal questions users are asking to the system, topic modelling using both latent Dirichlet allocation (LDA) and BERTopic has been considered.

LDA is a probabilistic model where each document is a mixture of topics, and each topic is a mixture of words (Silge & Robinson, 2021). In order to find the best clusters of topics, the team tuned hyperparameters to maximize coherence (how well the words in a topic fit together) and minimize perplexity (how well the model understands the text) (Texas Tech University, 2024).

The best configuration of LDA resulted in selecting six topics that captured core legal domains: intellectual property, civil procedure, family law, private property disputes, environmental obligations, and data protection.

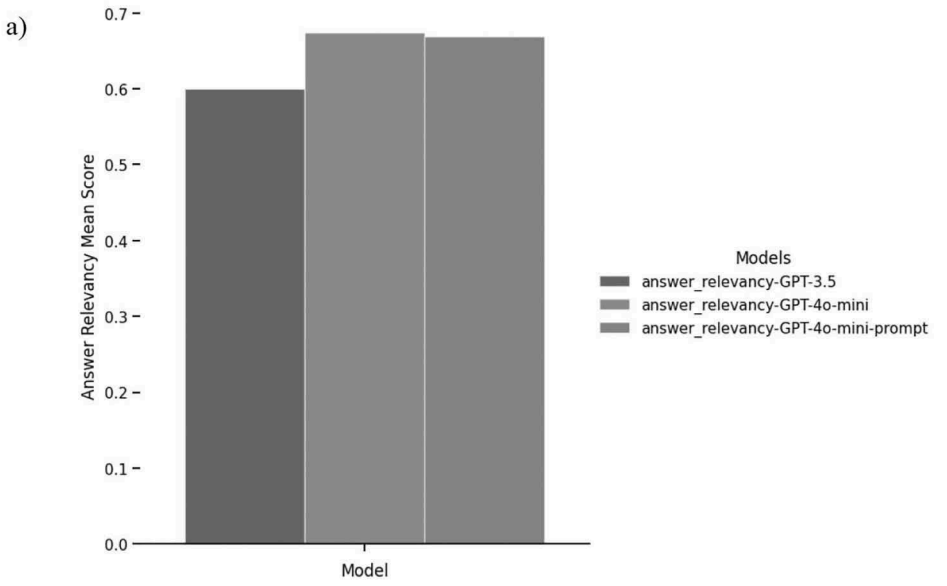


FIGURE 15.7 Answer relevance score comparison across models
Source: Field research.

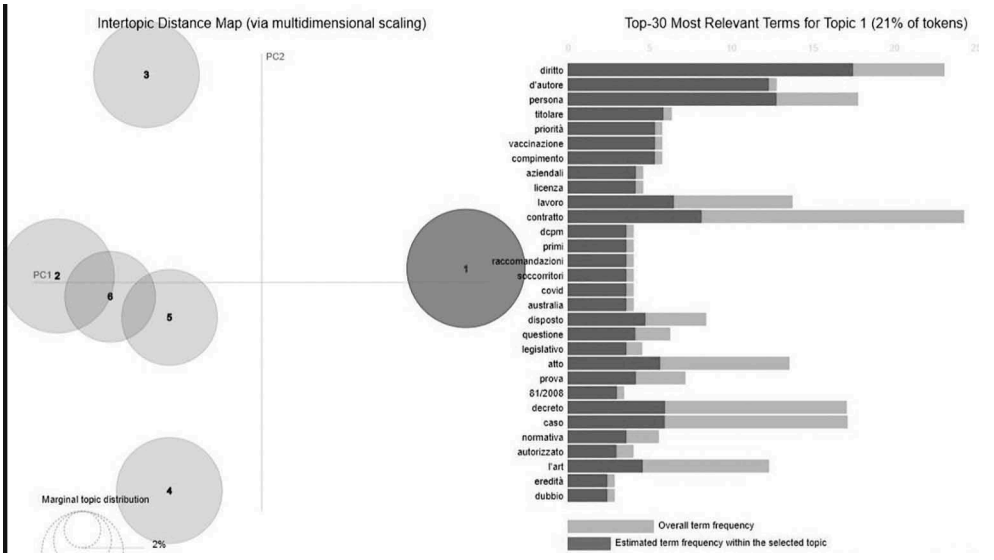


FIGURE 15.8 Visual representation of the topics via LDA inter-topic distance map
Source: Field research.

On the other hand, BERTopic is a modern method that groups similar questions by analyzing their content. It processes each question, reduces the complexity of the data, and then organizes them into clear, meaningful topics. Each step of the BERTopic pipeline is shown in Figure 15.8.

In this case, the team adjusted the hyperparameters related to how the data is grouped and the number of topics. This setup resulted in ten topics, with the model showing good consistency and variety. BERTopic grouped the text into topics like labor disputes, property donations, traffic incidents, environmental permits, and vaccine mandates.

By combining LDA and BERTopic, the team gained a broader view of user interests, as LDA provided interpretable topic distributions, while BERTopic highlighted emerging concerns and subtopics.

Pro tips:

- Use multiple topic modelling methods to capture both established and emerging user needs.
- Visual maps of topics help quickly identify gaps in content coverage.

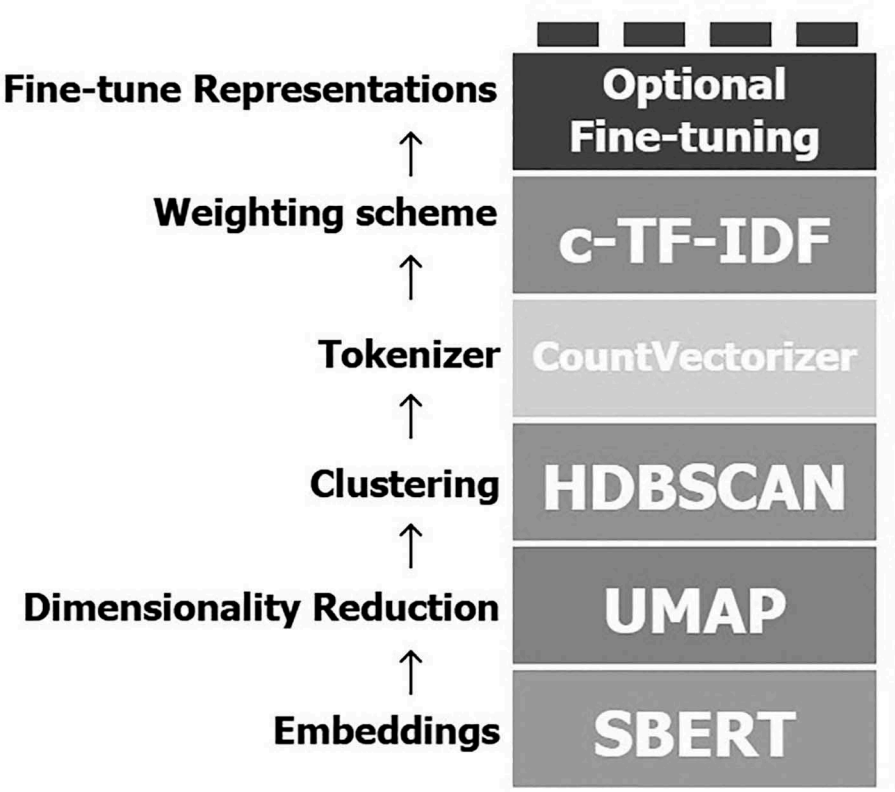


FIGURE 15.9 Five steps BERTopic tower
Source: By courtesy of Maarten Grootendorst. Image available at: https://maartengr.github.io/BERTopic/algorithm/algorithm.html

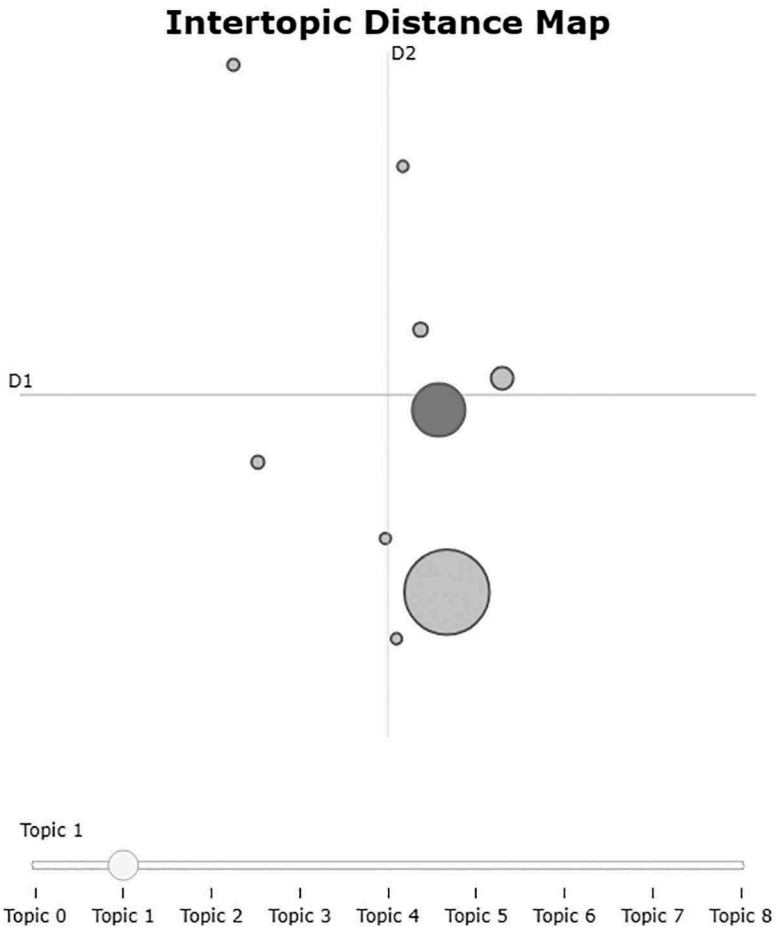


FIGURE 15.10 BERTopic inter-topic distance map
Source: Field research.

- Always check if topics make sense to domain experts—AI may group unrelated items together.

In practice:

- If launching a legal AI platform, analyze user queries regularly to adapt content and improve coverage.
- Combine interpretable methods (like LDA) with flexible AI methods (like BERTopic) to cover broad and niche topics.
- Use topic insights to guide prompt design, training data updates, and user guidance materials.

3.4 Insights Gained

The case study yielded several insights. First, prompt engineering and model selection have a significant influence on answer quality. Using prompts with clear statutory references increased faithfulness and relevance, and GPT-4o mini outperformed GPT-3.5 Turbo while being more cost-effective.

Second, topic modelling showed users ask about both classic and emerging legal issues—indicating that any legal AI system must be prepared to address a broad range of topics. Third, the project underlined how important data quality is: meaningful evaluation simply isn't possible without carefully cleaned and well-structured data. Fourth, a specific framework like RAGAs highlighted whether answers were grounded in context, which generic tools often miss (Roychowdhury et al., 2024). Finally, even with powerful models and thoughtful system design, human oversight remained essential, as retrieval errors, ambiguous questions, and limitations of generative models pushed lawyers to review AI outputs before relying on them.

To further address the lessons learned and to contextualize GiuriMatrix, it is possible to compare it with other platforms along data, evaluation, human oversight, and ethical considerations.

This comparison illustrates strategic and ethical trade-offs: GiuriMatrix emphasizes contextual fidelity and human oversight, whereas other platforms prioritize scale or accessibility, sometimes at the expense of rigorous evaluation.

All in all, these lessons demonstrate that achieving the balance between governance and algorithms is paramount in building trustworthy legal AI. GiuriMatrix shows that retrieval-augmented generation can produce useful answers when the underlying data are reliable, prompts are thoughtfully designed, and outputs are rigorously evaluated.

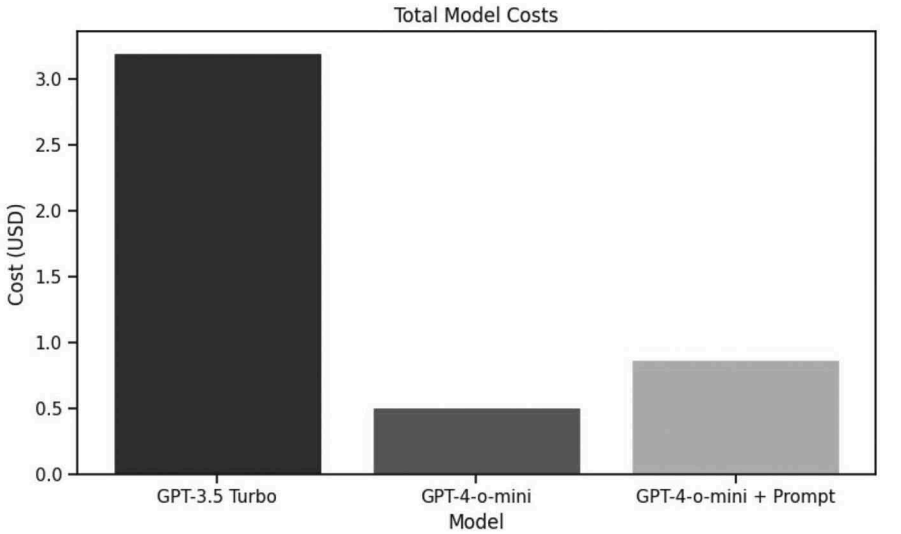


FIGURE 15.11 Bar chart—total model costs
Source: Field research.

TABLE 15.3 Case Comparison Table contrasting GiuriMatrix with other LegalTech platforms

Platform	*Data Quality Processes*	*Evaluation Framework*	*Human Oversight*	*Reported Accuracy Gains*	*Ethical/Trust Considerations*
GiuriMatrix	Curated 757 QA pairs; cleaned & split contexts	RAGA metrics (faithfulness, relevance, context)	Mandatory human review; prompt engineering	GPT-4o mini + Prompt: 0.342 faithfulness	Context grounded in statutes; transparency emphasized; learns from COMPAS lessons on bias
Casetext	Large corpus of legal documents; automated indexing	BLEU/ ROUGE, some relevance checks	Human verification limited; AI-assisted drafting	High for contract clauses	Primarily designed for U. S. law; fairness less explicitly addressed
DoNotPay	Aggregated legal templates; minimal verification	User feedback; limited automated evaluation	Minimal oversight; no human-in-the-loop	Minimal	Misled users as "robot lawyer", no mitigation of hallucinations

Source: Authors' own elaboration, based on the field research findings.

It also suggests that human–AI collaboration, rather than automation alone, will be the hallmark of effective legal-AI services.

In practice—key takeaways:

- Audit your dataset for completeness and bias.
- Apply context-grounded evaluation metrics (RAGAs).
- Establish human checkpoints to verify AI-generated advice.
- Consider cost–performance trade-offs in model selection.

4 Understanding the Ethical, Legal, and Governance Landscape

The deployment of AI in legal contexts raises complex ethical, regulatory, and governance challenges. Practitioners must navigate algorithmic fairness, data privacy, transparency, and human oversight. Establishing clear frameworks ensures that legal AI systems are trustworthy, accountable, and aligned with professional standards.

4.1 Algorithmic Fairness and Bias

AI systems can unintentionally reinforce societal inequalities if not carefully designed and monitored. A prominent example is the COMPAS recidivism tool, which classified black defendants who did not reoffend as medium or high risk more than twice as often as white defendants (Corbett-Davies et al., 2016). Although the system was calibrated to optimize predictive accuracy within groups, no algorithm can satisfy all

fairness criteria: achieving equal false-positive rates and predictive accuracy simultaneously is generally impossible.

Legal AI developers must make trade-offs: which fairness metrics to prioritize and how to balance them against accuracy. GiuriMatrix avoids high-stakes decision-making, focusing on providing information grounded in statutory text. Bias can still arise from uneven data coverage or prompt design. Mitigation strategies include bias audits, monitoring outputs by demographic group, involving diverse development teams, and documenting fairness objectives. Transparency about these choices is crucial for sustaining trust.

Pro tip: Regularly review outputs against known benchmarks to detect subtle bias trends.

4.2 Data Privacy and Security

Legal AI systems handle sensitive information such as client data, contracts, and court documents. In the EU, the General Data Protection Regulation (GDPR) mandates transparency, anonymization, and data minimization. AI systems must prevent unauthorized disclosure of personal data.

Techniques like federated learning—training models on decentralized datasets—and differential privacy, which adds controlled noise to outputs, enhance data protection. Encryption for data at rest and in transit is also essential. GiuriMatrix uses only public legal texts and anonymized queries; however, operators must secure logs and prevent storing or sharing users' personal data.

4.3 Trust, Transparency, and Human Oversight

Trust in AI extends beyond accuracy. It includes transparency, interpretability, and accountability. The DoNotPay case illustrates risks of neglecting these principles: in 2024, the U.S. Federal Trade Commission sanctioned the company for marketing its chatbot as "the world's first robot lawyer," misleading users to believe it could replace a human lawyer (DoNotPay, 2025; FTC, 2024). The company paid $193,000 in penalties and had to inform users of limitations.

This highlights the importance of human oversight, responsible communication, and regulatory enforcement. Unlike DoNotPay, RAG systems like GiuriMatrix enhance transparency by displaying source documents. Developers can improve trust by:

- Highlighting relevant passages and providing citations.
- Allowing users to explore underlying sources.
- Enabling feedback mechanisms to rate answers or flag mistakes.

Ultimately, professional ethics demand a "human-in-the-loop" approach: lawyers review AI outputs before reliance. Training programs that enhance AI literacy ensure responsible system use.

4.4 Regulatory Compliance and Responsible AI Frameworks

Users of AI systems are starting to demand transparency and accuracy, prompting governments and professional bodies to develop AI regulations, particularly in domains like law. The EU AI Act (European Parliament and Council, 2024) classifies the use of AI in the application of justice as a high-risk application, requiring human supervision, conformity assessments, and transparency regarding training data. Bar associations are drafting field specific guidelines and good practices on the ethical use of AI on explainability, robustness, data governance, and accountability. Field-specific guidelines recommend:

- **Ethical frameworks** promoting fairness, transparency, and accountability.
- **Data governance** ensuring lawful acquisition, anonymization, and security.
- **Model governance** documenting training data, hyperparameters, and evaluation results.
- **Stakeholder engagement** with legal professionals, clients, and affected communities.
- **Continuous auditing** to monitor performance and detect unintended consequences.

These practices support compliance while fostering trust.

In practice—key takeaways:

- Conduct regular bias audits and monitor outputs across demographics.
- Apply GDPR-compliant anonymization, encryption, and secure logging.
- Use "human-in-the-loop" processes; review AI outputs before legal reliance.
- Highlight sources, provide citations, and allow users to verify answers.
- Follow AI regulations, field-specific ethical guidelines, and continuous auditing protocols.
- Foster AI literacy for legal professionals to ensure responsible adoption.

5 Business Implications and Change Management

AI adoption rewrites the economic and organizational structure of companies adopting it. This section examines the economic impact, how can a company understand its readiness for adopting AI strategies and technologies, and how can it migrate into a new complete mindset by integrating AI workflows while managing change.

Economic Impact and Return on Investment: The innovations of AI reduce routine tasks, increase throughput, and improve quality. Document review systems speed up e-discovery and cut litigation costs, contract-analysis software helps close deals faster by automating clause checks, and research-assistants free lawyers for strategy. Although legal departments are not typically revenue centers, they can still drive value by reducing costs, mitigating risks, and improving client satisfaction. Nevertheless, measuring return on investment (ROI) of legal AI is challenging: only 7% of legal departments currently use specific key performance indicators (KPIs), and 40% are unsure whether their tools justify the investment. To address this, organizations should aim for data-driven evaluation methods structured with metrics like time saved, error reduction,

throughput, and satisfaction; cost–benefit analyses must include fees, integration, training, and risks (such as data breaches or incorrect advice).

Assessing Readiness and Building Capabilities: AI implementation requires readiness across multiple dimensions. The ITU AI Readiness Framework and similar models emphasize data availability, infrastructure, talent, culture and governance (ITU, 2024). For legal departments, these factors translate into the following matrix:

TABLE 15.4 AI Readiness Assessment Framework for Legal Implementation

Readiness Area	*What It Means*	*Where We Are*	*Where We Want to Be*	*What to Do Next*
Data	Do we have clean, organized legal documents and case data?	Some data is scattered and inconsistent	All key legal data is structured, accessible, and up-to-date	Audit data, organize into standard formats, create pipelines for new documents
Tools & Tech	Do we have the platforms, software, and security to run AI safely?	Basic tools, limited integration	Fully integrated, secure AI-ready systems	Deploy cloud/ vector DB, integrate with legal software, ensure security controls
People & Skills	Do our staff understand and can use AI effectively?	Limited AI knowledge; few experts	Teams trained in AI; collaboration across law & tech	Upskill lawyers, hire AI specialists, encourage teamwork between legal and tech
Governance & Ethics	Are there rules, oversight, and human checks on AI use?	Policies informal or missing	Clear policies, oversight committees, mandatory review steps	Define AI rules, create committees, track AI outputs, enforce human review

Source: Developed by the authors.

Pro tip: Remember the 10–20–70 rule when adopting AI: 10% of success comes from algorithms, 20% from technology, and 70% from people and processes. Focus most of your effort on communication, training, incentives, and feedback to secure buy-in and reduce resistance.

5.1 Integrating AI into Legal Workflows

Integrating AI into legal operations is an iterative process. Organizations typically begin with high-value, low-risk use cases, such as contract analysis and document review. These tasks have clear outputs and well-defined success criteria, making them suitable for pilot projects with baseline measurements, control groups where possible, and transparent reporting to stakeholders.

Once pilots succeed, AI can be embedded into broader workflows by integrating tools with document and practice management suites. Continuous monitoring ensures performance feedback, while frameworks such as RAGAs track faithfulness and

relevance. Usage metrics demonstrate how effectively tools are employed, complemented by human oversight in areas where technology cannot provide judgment.

Organizations should define mandatory human review points, for example final legal opinions. Staff should be trained to override AI suggestions, when necessary, with documentation of model versions, data sources, and evaluation results to ensure regulatory compliance and auditability.

In practice—key takeaways:

- Start small, scale strategically: Pilot AI on high-value, low-risk processes such as contract review before broader deployment.
- Measure and monitor ROI: Define clear KPIs including time saved, error reduction, and client satisfaction. Include indirect benefits such as risk mitigation and process efficiency.
- Invest in talent and cross-functional collaboration: Upskill legal teams, hire AI specialists, and encourage collaboration across technical and legal functions.
- Establish governance and human oversight: Define ethical and operational responsibilities, monitor AI outputs, and enforce mandatory review points.
- Manage cultural change: Focus on communication, incentives, and training to secure buy-in and integrate AI into everyday workflows effectively.

6 Conclusions

AI technologies are rapidly transforming the legal landscape and there is more to come.

Generative AI technologies are now handling not only text but also code, images, and audio generation. The next steps in legal AI go beyond drafting. Agentic AI isn't just about producing text; it's about acting. We can think of tools that drafts contracts, notice something's missing, circulates it for review, gathers comments, and then adapts the draft. However, when systems are given that autonomy, errors multiply. One mistake early in the chain can spread through the whole process. That's why firms need to start testing controlled pilots, keep a lawyer's eyes on the outputs, and build safety rails before trusting the tech with bigger decisions. Moreover, law is more than text. It's also voice recordings, scribbled notes, photos, and video evidence. That's where multimodal AI matters: systems that extract meaning out of mixed data instead of ignoring half the evidence. Finally, the language itself is a considerable problem in legal settings. In the European Union, one legal question may appear in five or six languages. Multilingual systems—built with translation, embeddings, and context sensitivity—are essential for real legal work. But this won't come from tech alone. Lawyers, linguists, and engineers must work together, an often uncomfortable collaboration.

AI is no longer knocking at the door of the legal profession—it's already inside and it will penetrate further. Lawyers and law assistants have gone from scanning documents to decision-support systems and now semi-autonomous agents. The GiuriMatrix study shows what can be done when technology is carefully designed and evaluated: retrieval, topic modelling, prompt engineering. These methodologies ensure real impact if used wisely. Still, none of this works without trust. Surveys show people

are cautious, adoption is slow, and ROI uncertain. Yet firms that integrate AI well already see higher revenue and returns. The signal is clear: ignoring AI isn't an option, but rushing in blindly is worse.

The road forward isn't just coding better models; it's building systems that are fair, transparent, and accountable. It's about leadership, governance, and putting humans where they matter. The next wave—generative, agentic, multimodal, multilingual—will only raise the stakes. If technical skill, ethics, and business sense are balanced, AI won't just make law faster; it could make it fairer, more accessible, and more just.

Next wave checklist:

- **Pilot controlled agentic workflows:** Start with low-risk processes, measure performance, and expand gradually.
- **Integrate multimodal and multilingual capabilities:** Ensure AI systems can process diverse evidence types and multiple languages for comprehensive legal coverage.
- **Maintain human oversight:** Define mandatory review points, document model outputs, and train staff to intervene when necessary.
- **Foster cross-disciplinary collaboration:** Encourage teamwork between lawyers, data scientists, linguists, and engineers to align technology with real-world practice.
- **Establish governance and audit mechanisms:** Implement ethical, transparent, and accountable processes to monitor AI performance and manage risks.

References

Ambrogi, R. (2025, June). Legal departments show growing AI adoption but implementation challenges remain, new survey finds. *LawNext*. https://www.lawnext.com/2025/06/legal-departments-show-growing-ai-adoption-but-implementation-challenges-remain-new-survey-finds.html.

BCG. (2024, December). Executive perspectives AI unlocked: Value creation with AI (including generative AI). https://media-publications.bcg.com/BCG-Executive-Perspectives-Value-Creation-with-AI-EP10-17Dec2024.pdf.

Casale, P., Viola, L., Filippelli, M., & Cozza, F. (2025). GiuriMatrix. https://www.giurimatrix.it/.

Clio. (2025). Clio | The industry's #1 legal software | Try it for free! https://www.clio.com/.

Corbett-Davies, S., Pierson, E., Feller, A., & Goel, S. (2016, October 17). A computer program used for bail and sentencing decisions was labeled biased against Blacks. It's actually not that clear. *The Washington Post (Monkey Cage Blog)*. https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.

DoNotPay. (2025). The AI that fights for you. https://donotpay.com/.

Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2024). RAGAS: Automated evaluation of retrieval augmented generation. *EACL 2024–18th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of System Demonstrations*, 150–158.

European Parliament and Council. (2024, July 12). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain regulations and directives (Artificial Intelligence Act). *Official Journal of the European Union, L 1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Evisort. (2025). The unparalleled leader in contract AI. https://www.evisort.com.

Federal Trade Commission. (2024). DoNotPay. https://www.ftc.gov/legal-library/browse/cases-proceedings/donotpay.

Fortune Business Insights. (2025). Legal technology market size, share & industry analysis, 2025–2032. https://www.fortunebusinessinsights.com/legal-technology-market-109527.

Goodenough, O. R. (2015, April 8). Legal technology 3.0. *HuffPost*. https://www.huffpost.com/entry/legal-technology-30_b_6603658.
Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. doi:10.5281/zenodo.4430182.
Hartung, M., Bues, M.-M., & Halblieb, G. (2018). Legal tech, how technology is changing the legal world (p. 5).
Ironclad. (2025). AI-powered contract management software: Unblock contracts, unlock potential. https://ironcladapp.com.
ITU. (2024). AI ready – Analysis towards a standardized readiness framework (Issue September). International Telecommunication Union. https://s41721.pcdn.co/wp-content/uploads/2025/03/T-AI4G-AI4GOOD-2024-2-PDF-E-1.pdf.
LawGeex. (2025). LawGeex – Conquer your contracts. https://www.lawgeex.com/.
Litera. (2025). Kira + GenAI. https://www.litera.com/products/kira/gen-ai.
Litify, L. L. C. (2025). Litify: The legal platform designed for high performance. https://www.litify.com/.
Miki, S. (2024, June). Top Legal Tech Startups in 2024. Clio Blog. https://www.clio.com/blog/legal-tech-startups/.
Paladin. (2025). Pro bono software for the legal industry. https://www.joinpaladin.com/.
Reuters Thomson (2025). HighQ: Effortless collaboration tools for legal professionals. https://legal.thomsonreuters.com/en/products/highq.
Roychowdhury, S., Soman, S., Ranjani, H. G., Gunda, N., Chhabra, V., & Bala, S. K. (2024). Evaluation of RAG metrics for question answering in the telecom domain. http://arxiv.org/abs/2407.12873.
Santhosh, S. (2023). Understanding BLEU and ROUGE score for NLP evaluation. *Medium*. https://medium.com/@sthanikamsanthosh1994/understanding-bleu-and-rouge-score-for-nlp-evaluation-1ab334ecadcb.
Silge, J., & Robinson, D. (2021). Topic modeling. In Text mining with R. https://www.tidytextmining.com/topicmodeling.
Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2025). The state of AI: How organizations are rewiring to capture value. McKinsey Global Survey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.
Spellbook. (2025). https://www.spellbook.legal.
Texas Tech University. (2024). Perplexity and burstiness. https://guides.library.ttu.edu/ai-literacy/evaluate-pb.
Thomson Reuters. (2023). Thomson Reuters completes acquisition of Casetext, Inc. https://www.thomsonreuters.com/en/press-releases/2023/august/thomson-reuters-completes-acquisition-of-casetext.html.
Thomson Reuters Corporation. (2025). Thomson Reuters. https://www.thomsonreuters.com/en.
Wagner, J. (2020). *Legal tech und legal robots: Der wandel im rechtswesen durch neue technologien und künstliche intelligenz*. Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-28555-5.

16

AI STRATEGY BLUEPRINT

A Leadership Mandate in the Age of Artificial Intelligence

Zorina Alliata

Introduction

Even the most advanced organizations face a paradox: more than 80% of artificial intelligence (AI) projects fail to accomplish their objectives or never reach production, nearly doubling the historical failure rate of traditional IT projects. This harsh fact is not a technical footnote, but a rallying cry for executive action. AI, with its promising potential and intimidating complexity, is ultimately a leadership issue. It requires a new type of coordination that seamlessly merges business vision, rigorous execution, and relentless operational scale.

As a business leader, whether from the CEO's office, the program management command center, or the perspective of a technology executive, your responsibility is for every AI output, whether it succeeds or fails. Leadership is what distinguishes the world's most effective transformations: aligning people, processes, and technology, gaining buy-in, and pulling the proper change levers when ambiguity or risk threatens progress.

In this chapter, we present your blueprint: a practical guide to developing a plan that is not only ambitious, but also feasible, responsible, and scalable. We will explain why so many AI projects fail at the "pilot" stage and outline a strategy for transforming isolated attempts into enterprise-wide, competitive realities.

AI projects are about twice as likely to fail as non-AI technology projects: a staggeringly persistent and expensive gap.

Part 1: People—Establishing the Foundation via Talent and Culture

83% of businesses now view AI as a strategic necessity (McKinsey & Company, 2025a), emphasizing its importance to corporate success.

Your AI goals begin with your employees, not just data scientists, but everyone involved in the data-driven revolution. As a leader, you set the tone and expectations: you must promote data and AI literacy as key business skills, knowing that future competitiveness depends on a company-wide proficiency in analytics, ethical data use, and a willingness to challenge the current status quo.

DOI: 10.4324/9781003632429-23

This chapter has been made available under a CC-BY-NC-ND 4.0 license

TABLE 16.1 Executive AI landscape: by the numbers

Metric	*Statistic*
AI Project Failure Rate	70–85%, "Over 80%" (S&P Global, RAND, Gartner, Informatica) (Beam.ai, 2025) (Brainvire, 2025) (BCG, 2025) (Exploding Topics, 2025a)
Non-AI Technology Project Failure Rate	30–40% (Standish Group, Gartner, HBR) (CIO Dive, 2025)
AI Proofs of Concept Fail Before Production	46% (S&P Global, Informatica, CIO Dive)(Beam.ai, 2025) (Brainvire, 2025) (Exploding Topics, 2025a) (Exploding Topics, 2025b)
AI Projects Never Move Beyond POC	70–90% (CIO, S&P Global, Beam.AI, industry surveys) (Exploding Topics, 2025a) (HAI Stanford, 2025b)
Companies Implementing AI at Scale Across Teams	~13% (industry consensus, leadership studies) (Exploding Topics, 2025c)
Scaling AI vs. POC ROI	3x higher returns (case studies, 2025 reports) (Glide Apps, 2025)
Global AI Market Value (2025)	$184B–$420.5B (Grand View, LogMeIn Resolve) (Grand View Research, 2025) (HAI Stanford, 2025a)
Global AI Market Value (2030)	$826.7B–$1.8T (Grand View, LogMeIn, Exploding Topics) (FAIR, 2025) (HAI Stanford, 2025a) (Grand View Research, 2025)

Source: Compiled from multiple industry reports including Beam.ai (2025), BCG (2025), Exploding Topics (2025), HAI Stanford (2025), and Grand View Research (2025).

For CEOs, this means investing in learning programs that go beyond IT. Board members and senior management require specialized training to understand AI's promise and limitations, the distinction between machine learning and traditional analytics, the systemic risks of bias, and the leadership role in governance. Encourage managers at all levels to take responsibility for AI knowledge growth by combining cross-functional workshops, job rotations, and structured mentoring.

Almost every firm that has successfully used AI has a recognizable strong base of data literacy and analytical curiosity. It would be a mistake to assume that such talents are always limited to technical teams. Instead, the most effective organizations promote data and AI literacy at all levels. Business leaders, for example, do not need to be experts in neural network mathematics to be effective sponsors; however, understanding fundamental concepts such as supervised versus unsupervised learning, the value of high-quality data, and the challenges of bias and explainability enables them to make practical and forward-thinking decisions.

For companies in its early stages, the construction of this literate environment can be similar to the formation of a new language and set of conventions. Leadership is critical in fostering an experimental mindset while remaining grounded in practical commercial goals. Effective firms treat AI as an essential part of their operational model, which is reflected in their training priorities. Employees at all levels are given onboarding paths that focus not only on new capabilities, but also on the business challenges those tools are intended to solve and the ethical environment in which such solutions must exist.

Consider the onboarding process for a new data scientist. From the start, she is involved in the domain's key concerns, such as how AI can streamline the supply chain, customize product recommendations, and increase customer responsiveness. By ensuring that her technical orientation is directly tied to fundamental company strategy, her abilities are naturally guided towards relevant outcomes.

It is important to have more than just technical skills. One overlooked fact is the critical requirement for communicators and translators: people who can bridge technical depth with business relevance. Product managers, for example, who understand both the potential and limitations of AI, can have a significant influence by focusing efforts on areas that are likely to produce meaningful, long-term value rather than novelty. Similarly, the development of "AI translators," professionals with hybrid backgrounds in business and analytics, often serves as the glue that keeps projects from falling into obscurity or being stopped by technical misunderstandings.

A restricted understanding of skill development cannot be afforded by a company that seeks long-term success. Cross-functional training and job rotations promote an environment of empathy and mutual understanding. In mature environments, data scientists work in sales or supply chain departments, whereas business leaders attend workshops where they build a rudimentary model from scratch, not for production, but to appreciate the craft. Many firms have discovered that by weaving these threads of experience together, their AI teams can work as integrated partners within the business rather than in isolation.

Along with skill acquisition, leadership development is critical. Senior decision-makers, those in charge of strategy, must have access to customized executive education that allows them to explore in depth the developing difficulties of trust, governance, and competition in a more algorithm-driven world. This executive training is not a one-time event, but rather a regular element of the organization's learning environment, reflecting the rapid pace at which the AI frontier is shifting.

FACT BOX

While 83% of worldwide firms consider AI to be a key priority, only 28% of non-technical staff believe they are "prepared" to work with it. (McKinsey & Company, 2025a)

Acquiring specialized talent has become a competitive sport for corporations looking to embark on ambitious AI initiatives. However, even the most attractive recruitment pitch is ineffective unless it is supported by an environment that fosters meaningful, high-impact work. The fight is not only for technical expertise, but also for hearts and minds that are sufficiently engaged to persevere through adversity and increase the organization's AI maturity over time.

Structured onboarding in this context is more than simply an introduction to corporate culture; it is a thorough understanding of what distinguishes the organization's approach to AI. New hires go through immersive experiences, such as workshops that simulate common AI project issues, introductions to company-specific data ethical standards, and early exposure to potential mentors. Mentorship is more than just a

social courtesy; it speeds the transmission of explicit and tacit information, helping new team members to climb the learning curve more quickly and with fewer mistakes.

Beyond onboarding, retention is an exercise in both community building and career path clarity. AI experts like firms that promote intellectual curiosity and foster a culture of learning at all levels. Teams that view milestones and even unsuccessful experiments as learning opportunities are more likely to succeed. Companies that document and communicate their significant lessons, sometimes as "lunch and learn" sessions, sometimes as written best-practice guides, have better employee loyalty because they believe their effort is recognized and meaningful.

A sense of advancement is critical for technical talent, many of whom would want to grow their careers rather than pursue management. Forward-thinking firms provide dual career ladders, allowing individual contributors to advance in status and salary alongside their managerial colleagues. These organizations are also deliberate in their efforts to create an inclusive workplace in which diversity, in background, opinion, and experience, is actively promoted as an asset rather than a mere window dressing. This inclusivity extends to project design and decision-making, increasing not only staff involvement but also the possibility of producing strong, unbiased AI solutions.

One of the most overlooked aspects of successful AI transformation is the systematic acquisition and dissemination of organizational knowledge. In traditional IT environments, this may imply documentation and version control; but, in AI contexts, the stakes are often higher, since insights into data idiosyncrasies, modeling techniques, and recurring issues might mean the difference between one-time success and serial achievement.

Statistics show that companies that provide structured mentorship and knowledge transfer retain 50% more AI talent than those that do not.

To that purpose, the best companies institutionalize their findings and mistakes. They produce model and code libraries, which are repositories of reusable materials, annotated datasets, pre-trained models, and domain-specific code that are more valuable than an individual's memory. These libraries benefit from the collective intellect of the organization's AI community, which meets on a regular basis to discuss, question, and reinforce lessons learned. After-action reviews become ritualized, and departing employees are expected to "carry the torch" by formally transferring expertise, which is commonly accomplished through video walkthroughs, written documentation, and peer shadowing.

Alignment with larger learning management systems guarantees that AI knowledge does not sit in isolation. Instead, it becomes one stream in the river of organizational capability development, accessible to anyone who may need to learn or teach in the future. This method converts what would otherwise be a collection of isolated trials into a compounding source of value that becomes more valuable with each project and new employee.

FACT BOX

An alarming 4.2 million AI posts remain unfilled worldwide, with hiring processes taking an average of 142 days. The average annual cost of recruiting delays for AI talent is $2.8 million. (TrustArc, 2025)

Part 2: Process—Managing Governance, Risk, and the Operating Model

Here's a structured glossary sidebar for your AI strategy blueprint, split into two categories: technical terms and regulatory terms. This format is ideal for sidebars, callouts, or appendices in your document or presentation.

From your position in the C-suite or as a program leader, you are responsible not only for vision but also for execution at scale. Governance transforms enthusiasm for AI into responsible, long-term value.

If people are the heartbeat of the AI-driven organization, process serves as the skeleton and nerves: the connective tissue that shapes efforts, channels energy, and builds resilience. A strong procedural architecture ensures that AI programs do not lose momentum or fall foul of supervision, regulatory, or ethical constraints. In the next sections, we will look at how organizations establish governance, operational models, compliance infrastructure, and risk mitigation measures to support long-term AI transformation.

The excitement at the start of an AI journey can lead to chaos if not controlled by clear lines of accountability and a systematic strategy for oversight. The governance of AI begins with unambiguous leadership. At its best, AI governance brings together

TABLE 16.2 Regulatory Glossary: Compliance and Ethics

Term	*Definition*
GDPR (General Data Protection Regulation)	EU regulation that governs data privacy and protection, requiring transparency, consent, and rights to access and deletion.
EU AI Act	Proposed European legislation to regulate AI systems based on risk categories, mandating transparency, oversight, and safety for high-risk applications.
Privacy by Design	A principle requiring privacy to be embedded into system architecture from the outset, not added as an afterthought.
Responsible AI	A governance framework that ensures AI systems are ethical, transparent, fair, and accountable throughout their lifecycle.
Impact Assessment	A formal evaluation of potential risks and harms associated with an AI system, often required before deployment.
Compliance Sprint	A collaborative review process where legal and technical teams assess AI projects for regulatory and ethical alignment.
Incident Response Plan	A predefined protocol for managing AI-related failures or breaches, including communication, remediation, and accountability steps.
Transparency Report	A public-facing document that outlines how AI systems are used, governed, and audited, often shared with stakeholders or regulators.

Source: Author's own compilation based on GDPR (2016), EU AI Act (2024), and established AI governance frameworks.

voices from throughout the organization, including CEOs in charge of corporate strategy, technical leads who understand model complexities, legal advisers familiar with changing regulatory frameworks, and subject matter experts concerned with end-user effect.

In reality, this typically requires the formation of a steering committee or board-level AI council. Their first step is to describe a vision that matches the organization's goals and constraints: Are we looking to automate back-office operations, reinvent the customer experience, or discover totally new business models? Governance bodies determine who owns the data, who has the authority to approve models for deployment, how risks will be disclosed, and how model outcomes will be conveyed both internally and publicly.

Once the governance structure is formed, a pattern of transparent decision-making emerges. The logic for selecting use cases, prioritizing projects, and allocating money is documented, resulting in a public record that can be examined and improved. Regular meetings and reviews bring to light new discoveries and issues, while periodic, systematic audits assess the ongoing effectiveness, fairness, and compliance of operationalized AI models.

These bodies must remain vigilant regarding the social bond between business and society. Stakeholder participation is not a bureaucratic roadblock, but rather a source of credibility. By including external voices, customers, suppliers, and even regulators, organizations may guarantee that their AI efforts anticipate public expectations and avoid the danger of controversy or backlash. This emphasis on open conversation and adaptability distinguishes truly mature AI governance from simple protocol for its own sake.

The decision of where and how to incorporate AI capabilities is highly strategic, influencing not only short-term project performance but also the organization's long-term ability to scale. Some businesses start with a centralized approach, combining AI knowledge into a dedicated team that acts as an inside consultant for business units. This model enables knowledge concentration, consistency of standards, and control over pipeline prioritizing.

As AI goals evolve, many organizations employ embedded or hybrid models, distributing AI specialists across business units, where they gain domain expertise and become trusted partners in ongoing change. The hybrid model, which is possibly the most adaptable, combines centralized governance (and thus risk management and compliance capacity) with distributed delivery, fostering both innovation and consistency.

Regardless of the model adopted, clear mechanisms for project identification, prioritization, and resource allocation are critical. The operating model must address problems such as how projects are funded, who owns the business, and how input from production deployments is routed back into the central AI knowledge base. These approaches incorporate iteration and adaptability, similar to the agility found in high-performing software engineering shops. Both formal and informal feedback techniques facilitate speedy learning and prevent the repetition of previous mistakes.

The efficient AI operating model fosters long-term relationships across business, technology, and compliance functions. These cross-functional teams are not formed and maintained at random, but rather on purpose, with the sole goal of bringing

multiple viewpoints to bear on issue resolution. In the most modern firms, collaboration is codified: it is included in performance appraisals, career promotion, and even compensation structures.

Given the rising speed of AI regulation, such as GDPR, the EU AI Act, and sector-specific rules in the United States, proactive compliance is more than just avoiding penalties; it is also a source of competitive advantage. A reputation for trustworthy, explainable AI leads to opportunities with both enterprise partners and consumer customers.

Set a high standard: require privacy-by-design in data ingestion, detailed model documentation at each handoff, and an incident response strategy in case of any AI-related errors or ethical breaches. Leadership signals, such as quarterly AI ethics updates to the board and public transparency reports, can help position your business ahead of regulation, rather than behind it.

FACT BOX

Companies that invest early in AI compliance and responsible practices have 15% higher customer trust and are 47% less likely to experience operational delays or regulatory attention than their "wait and see" competitors. (Unframe.AI, 2025)

AI applications frequently begin with data, thus privacy-by-design becomes the norm. This principle requires that privacy considerations be integrated across all stages of the AI lifecycle, from data gathering to model deployment and eventual decommissioning. Documentation of data sources, model parameters, and decision logic is kept not only for repeatability purposes, but also in case regulators or unhappy stakeholders require a full accounting.

Integrating compliance into the development lifecycle is a continuous issue. Organizations frequently discover that legal and technical teams must learn one other's languages, sometimes through combined training or "compliance sprints" in which new projects are collaboratively assessed for conformance to both external and internal criteria. The ability to generate evidence, documentation, audit trails, communications, becomes just as vital as the innovation itself.

AI has enormous potential, but it also has significant hazards. These hazards are classified as technical (model drift, insufficient data), operational (lack of maintainability, outages), reputational (unintended bias, lack of transparency, negative social impact), and governance (regulatory fines, model misuse). The discipline of responsible AI stems from an awareness of these risks and a determination to mitigate them proactively rather than reactively.

A responsible AI framework is often based on specified principles, such as justice, transparency, privacy, and accountability. Some organizations have defined these concepts in an AI ethics charter, which is publicly published and used in day-to-day operations. Others incorporate these norms into operational policy, referencing company ideals and legal mandates.

Prior to the construction of any model, risk identification begins. A project intake process may require thorough impact evaluations that map potential technological

failure modes, ethical hazards, and business risks. During development, teams conduct bias and fairness checks using both automatic and human methods. Importantly, the process does not end with deployment: continual monitoring and frequent audits establish a feedback loop that might detect drift or unintentional damage.

Organizations that are most committed to responsible AI plan for the inevitable occurrence of errors. Incident response plans are created in advance, outlining who needs to be contacted, how decisions will be taken, and how the organization will communicate with affected parties and the general public. In the event of a breach or failure, this planning is vital, and it frequently distinguishes those firms that can weather reputational storms from those that cannot.

To turn AI potential into concrete commercial benefit, product and program management must be rigorous. AI initiatives are notorious for scope creep, technical delays, and uncertain value realization. The notion of user-centricity is fundamental to successful management: initiatives should not begin with the query, "What can our technology do?" but rather, "Whose problem are we solving, and how do we assess success?"

In AI, effective product management prioritizes user understanding. Whether the end-user is a customer, an internal stakeholder, or an automated process, the approach starts with extensive discovery: interviews, data analysis, and prototyping. Value-focused key performance indicators (KPIs) define projects by tracking not only technical measures like model correctness, but also business-level results like new revenue, lower churn, and higher satisfaction.

The lifecycle of an AI product is ongoing. Models must be maintained, updated, and finally retired if they become obsolete or hazardous. Product managers embrace this end-to-end approach, ensuring that maintenance budgets, support methods, and learning programs are established from the start.

Above the individual product, program management directs the AI portfolio. A combination of short-term victories and long-term bets is developed, balancing certainty with ambition. Cross-functional teams, which combine technical depth, business acumen, and compliance knowledge, are the norm, not the exception. Open and constant communication is prioritized: stakeholders are kept informed and engaged, ensuring course corrections are timely and projects stay on track.

Part 3: Technology—Building a Pragmatic, Future-Ready AI Architecture

If the main takeaway from the last decade has been the high failure rate of AI projects, the basic cause is similarly obvious most firms excel at pilots but fail catastrophically at scaling those pilots into production systems that provide meaningful value. The "proof-of-concept trap" is costly and widespread, as we saw in the statistics cited at the beginning of this chapter.

The root issues are largely operational integrating AI into legacy systems, managing data quality at scale, reducing costs, and ensuring compliance. Many businesses dramatically underestimate these obstacles as distinct from standard technology initiatives, resulting in the dual problems of "pilot purgatory" and paralyzing architectural sprawl.

TABLE 16.3 Technical Glossary: AI & ML Operations

Term	*Definition*
MLOps (Machine Learning Operations)	A set of practices that combines machine learning, DevOps, and data engineering to automate and scale the lifecycle of ML models—from development to deployment and monitoring.
LLMOps (Large Language Model Operations).	Specialized workflows and tools for managing large language models (LLMs), focusing on fine tuning, prompt engineering, safety, and deployment at scale.
FMOps (Foundation Model Operations)	Operational frameworks for deploying and managing foundation models (e.g., GPT, BERT) across multiple taks and domains, emphasizing governance, adaptation, and resuse.
Model Drift	A phenomenon where a model's performance degrades over time due to change in data patterns or external conditions.
Human-in-the-Loop (HITL)	A design approach where human oversight is integrated into AI systems to ensure accountability, accuracy, and ethical decision-making.
Agentic Frameworks	Architectures where autonomous AI agents interact with environments and each other to complete tasks, often with embedded reasoning and feedback loops.
Data Lineage	The tracking of data's origin, transformations, and usage across systems, crucial for compliance and reproducibility.
Bias and Fairness Checks	Techniques used to detect and mitigate discriminatory patterns in AI models, ensuring equitable outcomes across demographic groups.
Explainability	The ability of an AI system to make its decisions understandable to humans, often through interpretable models or post hoc analysis.
Audit Trail	A chronological record of model development, deployment, and usage decisions, used for compliance and accountability.

Source: Author's own compilation based on established AI/ML industry terminology and practice.

FACT BOX

Despite the vast market potential, only 13% of firms have successfully deployed numerous AI systems at actual scale. The divide is not technological, but rather operational, cultural, and strategic (UnitX Labs, 2025).

TABLE 16.4 Top obstacles to AI success (Informatica CDO Insights 2025)

Obstacle	*Percentage of Respondents*
Data quality and readiness	43%
Lack of technical maturity	43%
Shortage of skills	35%

Source: Informatica CDO Insights (2025).

The truth for most firms beginning on AI programs is that data is chaotic, inconsistent, and frequently incomplete. AI is famously hungry for large amounts of high-quality, diverse data, thus the first technological issue is to build effective pipelines for data intake, cleaning, and enrichment.

A comprehensive data pipeline automates the process of collecting data from a variety of sources, including internal transactional systems, customer touchpoints, third-party providers, and many more. The quantity of available data is less significant than its usefulness and purity. Early pipeline stages concentrate on validation and de-duplication, attempting to detect and rectify problems before they propagate into models. Once data has been standardized, it is enriched, combined with external datasets, annotated, and classified in a way that will aid eventual model explainability and auditability.

Metadata management is a common theme. The capacity to document data lineage, how it was sourced, modified, and used, is critical in practically every aspect of compliance, governance, and technical reproducibility. Organizations use data catalogs and lineage tracing systems not only to keep organized, but also to ensure that every insight provided by an AI model can be traced back to its ancestral origins, thereby enabling both trust and regulatory compliance.

To make this data available at scale, businesses invest in platforms that can handle huge, concurrent access. Data lakes and warehouses serve as repositories, and standardized APIs control access, allowing internal teams and permitted partners to securely and efficiently access data resources.

Once data pipelines are in place, the focus shifts to managing models throughout their lives. The initial joy of a successful prototype is swiftly replaced by the knowledge that models must evolve: business contexts move, data patterns change, and regulatory constraints become more stringent.

Model management is the discipline that organizes this complexity. The development process is linked with version control techniques, which capture not just code specifics, but also data versions and parameters that influence model outcomes. When a model is ready for deployment, automation helps it move quickly from development to production settings, reducing manual errors, and increasing speed.

Once in operation, ongoing monitoring is required. Models can drift, worsening over time without any external indicators until flaws become visible and significant. Regular evaluation based on business-relevant and technical parameters permits quick retraining, adjustment, or withdrawal. Audit trails are kept, documenting who deployed which version of the model, on which data, and with what outcomes. This record of activities is invaluable for compliance audits and incident investigations.

Access control is a key component of governance. Not everyone should be allowed to deploy or alter models; role-based permissions protect against unintended or malicious modifications. Mature firms reinforce these controls with frequent penetration testing and vulnerability assessments, thereby improving both their compliance posture and operational resilience.

The most advanced enterprises now arrange ecosystems of collaborating AI agents, each tackling specific tasks while feeding into larger business activities.

Encourage your technical leaders to look into agentic frameworks, which enable scalable automation with human oversight. AI systems must be explainable in industries that rely heavily on trust or tight regulations. "Human-in-the-loop" design is criticalAI should enhance, not replace, human decision-making.

FACT BOX

Economic analyses suggest AI could boost global industrial productivity by 40% by 2035, with significant knock-on value for adjacent sectors and supply chains, especially if systems are designed for scale, adaptability, and ethical governance. (Rheodata, 2025; WeForum, 2025)

In practice, agents are digital actors who can see their surroundings, make judgments, and take actions to achieve certain goals. A recommendation engine, for example, acts as an agent when it automatically selects material for a user based on previous interactions, present context, and expected preferences. When such systems interact (communicate, negotiate, or even compete), they create multi-agent environments with emergent behaviors.

Agentic techniques' value to businesses stems from their capacity to automate complex, context-sensitive procedures. Logistics may be optimized with fleets of delivery workers negotiating routes in real time; customer service can be redesigned with chatbots that learn from each encounter and escalate to human agents only as needed. Critically, the most effective agentic systems do not replace human agency, but rather incorporate human judgment where necessary, known as "human-in-the-loop" design.

Strategically, the implementation of agentic frameworks requires not just technological complexity, but also a reengineering of business processes so that software and people operate in tandem. This paradigm shift in design enables new levels of efficiency, responsiveness, and personalization, provided it is directed by a strong sense of purpose and a mature grasp of potential hazards.

Measuring AI systems' effectiveness, fairness, and business impact is a combination of art and science. Too often, organizations focus solely on technical metrics (accuracy, recall, mean squared error) without considering the implications for business objectives or societal trust.

Comprehensive examination includes numerous domains. Standard technical measurements remain useful because they provide early indicators of underperformance or instability. However, business measurements are considered equally important. An AI system that performs well in experimental conditions but fails to enhance sales or

customer retention is ultimately of limited use. Organizations use dashboards to bridge this gap, providing visibility into genuine business impactcost savings, enhanced conversion rates, or better decision-making consistency.

CASE STUDY

Goldman Sachs—GS AI Platform:

- Approach: Build a proprietary AI platform to centralize and govern all AI development.
- Success factors:

a Supports multiple foundation models (OpenAI, Google Gemini, Meta Llama, etc.).
b Ensures compliance, security, and model interoperability.

- Impact: Accelerated internal AI adoption while maintaining regulatory rigor. (Rheodata, 2025)

In the age of responsible AI, various modes of evaluation have gained popularity. Fairness metrics assess demographic parity, disproportionate impact, or explainability scores in order to ensure that models do not systematically favor or disadvantage any particular group. Automated testing suites compare models to chosen test sets, looking for outliers, performance deterioration, or vulnerability to adversarial attacks. Visualization technologies bring performance trends to life, allowing non-technical audiences to evaluate and act on results.

These review methods are not limited to model development, but rather occur throughout the lifetime. Following deployment, real-world feedback is gathered, sometimes directly from users, sometimes from downstream systems, to fuel continuous improvement cycles that maintain performance while minimizing harm.

AI operationalization, the transition from experimental success to enterprise-scale production, is a significant problem. The rise of MLOps (machine learning operations), LLMOps (large language model operations), and FMOPs (foundation model operations) solves this issue by combining best practices from DevOps, data engineering, and information security.

MLOps adds rigor to the ML lifecycle by automating phases when appropriate and introducing friction for review and auditing. Continuous integration and deployment (CI/CD) pipelines automate code and model testing, validation, and release. Monitoring and alerting detect performance decline early and initiate retraining or rollback procedures as appropriate. Resource management keeps computational overheads in check, which is an important consideration as models and data develop exponentially.

LLMOps tailors these approaches to the specific needs of big language models, emphasizing scalability, customization, and rapid administration. FMOPs, on the other hand, are concerned with the management of enormous, pre-trained foundation models, which requires specific methods to versioning, security, and performance tuning. In all cases, the emphasis is on robustness, repeatability, and controlled agility.

Organizations that develop these operational capabilities discover that they not only enhance efficiency, but also provide the auditable, transparent processes required by regulators and increasingly vigilant customers.

The third component of the technological AI approach is scalability: the ability to progress from a promising pilot to an organizational-wide transformation. To this purpose, infrastructural decisions dominate early discussions. Cloud systems provide elastic, pay-as-you-go computing capacity, whereas on-premises solutions provide control and data sovereignty, which is typically required in highly regulated businesses. Hybrid setups, which combine the two, enable enterprises to optimize flexibility while adhering to cost and compliance constraints.

However, scalability is not limited to hardware. The architecture of AI systems frequently determines their long-term viability. Modular design, which favors loosely connected, reusable components, allows teams to upgrade or adapt systems without requiring complete rewrites, lowering the technological debt that can stifle innovation. Reusable components become institutional learning repositories that may be used across multiple projects and teams.

As adoption increases, corporate change management must keep up. The culture that supports a single AI team is not always suitable for enterprise-wide implementation. New teams must be trained in successful values and processes, as well as technical abilities. Federated learning, the technique of training models on decentralized data, frequently for privacy or data sovereignty reasons, presents new potential and concerns, requiring appropriate governance and cross-functional coordination.

Scalability is thus presented as both a technological and a human barrier. It is solved with forethought, architectural insight, continual training, and, most importantly, a commitment to learn and change as the AI journey progresses.

CASE STUDY

Bloomberg—BloombergGPT:

- Approach: Created a domain-specific foundation model and platform tailored for financial services.
- Success factors:

a Integrated into Bloomberg Terminal for tasks like summarizing earnings reports and risk analysis.

b Offers APIs for developers to build on top of the model.

- Impact: Enhanced user experience and analytical capabilities for financial professionals (Rheodata, 2025).

Metrics and Model Evaluation: Developing Your Executive AI Scorecard

Demand more from your teams than just technical precision and speed. Your AI dashboards should include operational KPIs (uptime, error rates), ethical measurements (bias detection, privacy breaches), and business effect (revenue, cost savings, customer happiness).

Invest in solutions that offer real-time visual insights. Ensure that all essential models include summary explainability reports, business owner approval, and post-deployment monitoring plans.

A well-designed AI scorecard enables executives to track, assess, and enhance their company's AI projects. This section explains how to create, deploy, and exploit a scorecard based on the previously described executive-level questions and strategic domains. The goal is to move beyond discrete KPIs and ensure that every AI project is evaluated holistically, including technical, business, ethical, and operational pillars.

1 Define Scorecard Domains and Key Metrics

Begin by identifying the strategic pillars you want to analyze. Most organizations will have the following:

Technical: For generative AI solutions, we would assess output quality and dependability, monitor hallucinations/errors, ensure safe/traceable agent operation, and be mindful of operational and security threats.

Business: Return on investment (ROI), direct cost savings, increased customer happiness, and additional revenue attributed to AI.

Ethical/compliance considerations include bias detection and mitigation, privacy safeguards, explainability of AI conclusions, and ethical review procedures.

Operational requirements include incident response times, issue escalation protocols, failure recovery, and ownership clarity for AI events.

The core assessment items for each pillar should be 3–4 executive-level questions. Begin by answering the previous section's questions.

2 Score Each Metric Using a Simple Scale

For each question or statistic, ask responsible leaders to offer a score.

1 (Weak): No formal procedure, ad hoc or incomplete practice.
3 (Developing): There are some processes and monitoring in place, but they are not complete or proactive.
5 (Best Practice): Mature, completely integrated, and reviewed on a regular basis under unambiguous executive leadership. Demonstrates outcomes and audits.

Encourage honest, evidence-based self-assessment by reviewing procedures, dashboards, incident logs, and policy papers to support scores.

Sample Executive Questions for Every Domain

1 Technical Domain (e.g., accuracy, model drift, uptime)

Are our generative AI models providing material (text, photos, suggestions, etc.) that meets the quality, relevance, and reliability requirements of our business applications?

How do we check for "hallucinations," factual inaccuracies, or improper outputs from our GenAI models and agents, and what safeguards are in place to protect against damage?

Do our autonomous agents (e.g., chatbots, virtual assistants, or decision agents) follow clearly established behavioral and safety bounds, and is their decision-making traceable and auditable?

Have we assessed and documented where, why, and how our generative models or agents could provide additional operational, reputational, or security issues (for example, prompt injection, jailbreaking, or data leakage)?

2 Business Domain (e.g., ROI, Cost Savings, Customer Satisfaction)

Can we clearly tie revenue increases, cost savings, or productivity benefits to our AI projects, and is this documented?

Are we measuring how AI affects consumer happiness or experience?

Do our AI investments provide a demonstrable ROI that is comparable to our other digital initiatives?

3 Ethical/Compliance Domain (e.g., Bias, Privacy, and Explainability)

Have our AI models been evaluated for bias, and are there processes in place to conduct regular ethical reviews?

Are privacy controls included into data collection, processing, and model deployment?

Can we explain, on an executive level, how our AI models make judgments, particularly in sensitive or regulated areas?

4 Operational Domain (e.g., Incident Response Times)

When an AI system fails, do we have a clear incident response mechanism in place, and how quickly can we identify and resolve these issues?

Are operational risks from AI documented, monitored, and monetized in business continuity plans?

Is there a defined owner and protocol for AI incidents, from detection to resolution and communication with stakeholders?

Best practices in this context are often measured by:

what top organizations (peers, industry leaders, regulatory advice) are now doing;
standards from reputable organizations (e.g., ISO/IEC for AI risk, NIST for trustworthy AI, and responsible AI principles from governments and multinationals).

In a culture of continuous development, best practices are constantly evolving, but currently include:

All technological, business, ethical/compliance, and operational parameters are proactively evaluated and addressed.
The practices are well-documented, transparent, and periodically evaluated with C-suite oversight.
The organization can demonstrate to stakeholders and external parties that AI systems are trustworthy, explainable, reliable, and provide value based on external standards.

3 Aggregate Scores by Domain

To get the domain scores, average the scores of the underlying questions in each pillar. A Technical score, for example, would be the average of your ratings of output quality, hallucination control, agent safety, and risk management.

4 Visualize the Results

This sample speedometer chart (gauge chart) provides a clear, executive-friendly summary of AI maturity across your organization's four core strategic pillars. Each gauge shows preparedness on a scale of 0 to 100 for the following domains:

Technical (generative AI and agents): Assesses the quality and dependability of AI outputs, the capacity to detect hallucinations or errors, the safety and traceability of agentic systems, and knowledge of operational and security threats.

Business: Evaluates the ROI, direct cost reductions, increases in customer satisfaction, and the creation of new revenue streams ascribed to AI efforts.

Ethical/compliance considerations include bias detection and mitigation, privacy measures, explainability of AI conclusions, and the availability of ethical review systems.

Operational: Documents incident response times, clarity of issue escalation protocols, capacity to recover quickly from failures, and well-defined ownership of AI incidents.

The sample scores shown are:

- Technical: 75
- Business: 70
- Ethical/compliant: 65
- Operational: 80

This executive dashboard clearly shows which pillars are the strongest and where improvement efforts could be most effective.

How to utilize this chart: Review the gauges during board or leadership meetings to ensure that strategic discussions take into account business, ethical, and operational dimensions, rather than just technical performance.



FIGURE 16.1 Sample executive AI strategy scorecard: speedometer chart showing maturity scores for technical, business, ethical/compliance, and operational pillars
Source: Designed by the author.

5 Set Benchmarks and Best Practice Targets

A score of "5" in any category indicates complete adherence to current industry best practices, not perfection, but leadership maturity in comparison to industry norms. A score of "3" indicates that you have achieved significant progress, but you should focus on systematic, proactive changes to keep up with top firms.

6 Communicate and Act on the Scorecard

In the most productive firms, the AI scorecard is more than just an internal tracking tool; it is the focal point of executive meetings and boardroom discussions. Leadership ensures that the scorecard is periodically reviewed at the highest levels, ensuring that all important stakeholders have access to more than just technical performance metrics. Rather than focusing solely on metrics like uptime or accuracy, the scorecard illuminates the bigger picture: how AI initiatives affect business outcomes, whether ethical and compliance safeguards are truly embedded, and whether operational protocols keep up with the complexity of intelligent systems.

As the executive team examines the most recent scorecard, the findings translate into specific next initiatives. If the technical domain reveals flaws, perhaps evidence of unreliable generative outputs or emerging operational risks associated with agentic systems, leadership takes decisive action to invest in stronger monitoring frameworks, improved guardrails for autonomous agents, or more rigorous performance testing for generative AI models.

If the ethical or compliance components fall behind best practices, the organization does not wait for problems to arise. Leaders conduct rigorous bias and privacy assessments and develop targeted initiatives to train teams in explainability and responsible AI. The commitment is clear: ethical integrity and user trust are not afterthoughts but are required components of AI maturity.

Similarly, if the scorecard identifies deficiencies in business impact or operational resilience, the route ahead is clear. The executive focus switches to fine-tuning the measurement of value, delivering unmistakable clarity about ROI and tangible business benefits from AI implementations. Simultaneously, incident management rules are

reinforced, with a renewed emphasis on quick, coordinated response and unambiguous accountability to resolve any interruption caused by AI systems.

Across all of these areas, the scorecard serves not only as a diagnostic but also as an executive mandate for continuous improvement, encouraging action, ensuring accountability, and instilling long-term confidence in leaders that their organization's AI journey is still aligned with the highest standards of performance, integrity, and enterprise value.

7 Review Regularly and Benchmark Externally

Repeat this evaluation every quarter or after important AI deployment milestones. Compare your ratings to those of industry peers, published frameworks (e.g., NIST, ISO/IEC), and growing regulatory standards. Use progress monitoring to recognize accomplishments and sustain momentum. AI excellence is a journey, not a destination.

This scorecard transforms AI supervision from a technical "black box" to a transparent, actionable tool for executive leadership, ensuring that every AI project is carried out with rigor, responsibility, and business value in mind.

The Executive as Chief AI Strategist

In the age of artificial intelligence, executives wield unmatched power: they determine not only investments, but also the entire DNA of how their business thinks about, manages, and implements AI. Your leadership in developing talent, implementing rigorous processes, and demanding future-proof technology will determine whether AI delivers true business value or becomes another empty promise.

Make AI your business strategy. Set firm standards for learning, cooperation, and risk. Implement measures for transparency and accountability. Support and challenge your teams as they transition from pilot to platform. Above all, consider AI as the next great frontier of corporate transformation, a journey shaped and accelerated by your choices as a leader.

References

BCG. (2025). Closing the AI impact gap. https://www.bcg.com/publications/2025/closing-the-ai-impact-gap.

Beam.ai. (2025). Why 42% of AI projects show zero ROI (and how to be in the 58%). https://beam.ai/de/agentic-insights/why-42-of-ai-projects-show-zero-roi-(and-how-to-be-in-the-58-)

Brainvire. (2025). AI-led solutions for manufacturing industry. https://www.brainvire.com/blog/ai-led-solutions-for-manufacturing-industry/.

CIO Dive. (2025). AI project fail data – S&P Global. https://www.ciodive.com/news/AI-project-fail-data-SPGlobal/742590/.

Exploding Topics. (2025a). AI statistics. https://explodingtopics.com/blog/ai-statistics.

Exploding Topics. (2025b). Companies using AI. https://explodingtopics.com/blog/companies-using-ai.

Exploding Topics. (2025c). Startup failure stats. https://explodingtopics.com/blog/startup-failure-stats.

FAIR. (2025). The AI leaders: What sets the top 13% apart? [Report].

Forbes. (2025, August 21). Why 95% of AI pilots fail—and what business leaders should do instead. https://www.forbes.com/sites/andreahill/2025/08/21/why-95-of-ai-pilots-fail-and-what-business-leaders-should-do-instead/.
Fortune. (2025, August 18). MIT report: 95 percent generative AI pilots at companies failing. https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/.
Full Scale. (2025). AI developer shortage solutions. https://fullscale.io/blog/ai-developer-shortage-solutions/.
Glide Apps. (2025). AI stats 2025. https://www.glideapps.com/blog/ai-stats-2025.
Grand View Research. (2025). Artificial intelligence (AI) market. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market.
HAI Stanford. (2025a). AI Index Report. https://hai.stanford.edu/ai-index/2025-ai-index-report.
HAI Stanford. (2025b). What workers really want from artificial intelligence. https://hai.stanford.edu/news/what-workers-really-want-from-artificial-intelligence.
Hostinger. (2025). How many companies use AI. https://www.hostinger.com/tutorials/how-many-companies-use-ai.
Informatica. (2025). The surprising reason most AI projects fail—and how to avoid it at your enterprise. https://www.informatica.com/blogs/the-surprising-reason-most-ai-projects-fail-and-how-to-avoid-it-at-your-enterprise.html.
LinkedIn. (2025). Why 75% of AI projects fail to deliver ROI—and how to turn things around. https://www.linkedin.com/pulse/why-75-ai-projects-fail-deliver-roiand-how-can-turn-things-minett-jssac.
LogMeIn Resolve. (2025). Analyzing the AI global market. https://www.logmein.com/resources/analyzing-the-ai-global-market.
McKinsey & Company. (2025a). The state of AI. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.
McKinsey & Company. (2025b). Superagency in the workplace: Empowering people to unlock AI's full potential at work. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work.
Microsoft WorkLab. (2025). 2025: The year the frontier firm is born. https://www.microsoft.com/en-us/worklab/work-trend-index/2025-the-year-the-frontier-firm-is-born.
OPIT Faculty. (2025). AI strategy blueprint: A leadership mandate in the age of artificial intelligence. https://workos.com/blog/why-most-enterprise-ai-projects-fail-patterns-that-work.
PwC. (2025). AI predictions. https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html.
Raynovich, R. S. (2025, September 25). How enterprise AI is unfolding: A look at top use cases and platforms. Forbes. https://www.forbes.com/sites/rscottraynovich/2025/09/25/how-enterprise-ai-is-unfolding/.
Rheodata. (2025). AI failures stats. https://rheodata.com/ai-failures-stats/.
Rutgers Executive Education. (2025, January 28). How non-technical professionals can thrive in the age of artificial intelligence. https://execed.rutgers.edu/2025/01/28/how-non-technical-professionals-can-thrive-in-the-age-of-artificial-intelligence/.
Sourcing Innovation. (2024, October 25). Two and a half decades of project failure. https://sourcinginnovation.com/wordpress/2024/10/25/two-and-a-half-decades-of-project-failure/.
TrustArc. (2025). 2025 benchmarks report. https://trustarc.com/press/trustarc-2025-benchmarks-report/.
Unframe.AI. (2025). MIT reports: State of AI in business 2025. https://www.unframe.ai/blog/mit-reports-state-of-ai-in-business-2025.
UnitX Labs. (2025). Scalable vision AI machine vision systems. https://www.unitxlabs.com/resources/scalable-vision-ai-machine-vision-systems/.
WeForum. (2025, April). AI jobs – International Workers Day. https://www.weforum.org/stories/2025/04/ai-jobs-international-workers-day/.
Writer.com. (2025). ROI for generative AI. https://writer.com/blog/roi-for-generative-ai/.

INDEX

Note: *Italic* page number refer to *figures* and **Bold** page number refer to **tables**.

Aagaard, A. 160
absolute explainability 89
academic models 118
accenture 180
accessibility 22, 26, 42–43; regulations 43
accounting/accountability 43, 143–144, 189–191, 244, 328
accuracy 369; *vs.* fairness 88
acquisition efficiency 209
adaptability 117, 145
advanced AI: augmentation 54–55; for contextual relevance 54–55; statistical models 183; vector search for contextual 55, **56**
adversarial resilience 319
aerospace and defense (A&D) sector 317–340; assurance trilemma in 319–320, *320*; case studies 332–339; commercial applications 318; data collection and generation 321–322; data lifecycle stages 321–323, *322*; data retention and disposal 323–324; data sharing and dissemination 323; data storage and processing 322–323; integration of 318; landscape 317; limitations 340; methodology 320–321; regulatory and compliance landscape *324,* 324–326; secure data management and privacy 326–331; theoretical and managerial implications 339–340
"aesthetic loss of control" 300
agency, algorithms 300–301
agentic techniques **392,** 394
agent-to-agent (A2A) ecosystems 210
agile: development practices 58; financial system 189
agribusiness 96
AgroTIC (Colombia) 96
AI adoption/adoption of AI 22, 23, 295; accuracy *vs.* fairness 88; classification framework for 156–164; in cultural heritage 23; efficiency *vs.* engagement 88; innovation *vs.* regulation 89; and integration 280; and legal industry 364–366; maturity in 90; organizational changes during 278–295; paradox of 88–90, *90*; responsible 149–151; strategic challenges for 285–288; sustainability *vs.* computation 89; synthesis 89–90; transparency *vs.* complexity 89
"AI Bill of Materials" (AI BOM) 331
AI–blockchain convergence 7–18, 230–231; disruption notion 8; emerging business opportunities 10–14; factors and barriers **16**; methodology 10; notion of 8–9; practical recommendations for managers 16–18, *17*; rise of 7–8; terminal of truth (ToT) 7–8
AI Capabilities Matrix 119
AI Capability Assessment Model (AI-CAM) 119
AI-Core Value Providers 157
AI-digital transformation 280–285
AI-driven: applications 255; business model innovation **160**; client communication *268*; communication 243; communicative environment 252–253, *253*; decision support 367; hybrid communication model 255, *256,* 272; inventory optimization 195–215; KPIs

292–295; personalization 185; sustainability performance 230
AI-driven digital transformation 288–292; culture 288–289; elements for *289*; knowledge 290–291; leadership 290; policies 291; tools 291–292
AI Experimenters/Piloters 157
AI Innovators/Creators 158
AI maturity model *84,* 85, 87, 109; as cyclical learning framework *87*
AI-mediated communication 268–271, **274**
AI mindset 109
AI4People's framework 125
AI-Powered Risk Guardians 162
AI-powered talent acquisition 188
Airbnb 298–301, 304, **307,** 307–309, **310**
"AI readiness" 28–29, 109–153; adaptability and VUCA environments 117; AI-RMM framework 116, 126–129; assessment framework for legal implementation **380**; in businesses framework 155–172; as capacity and commitment 111; for client-oriented communication 257–260; client preparedness and change management 137–139; cluster analysis 129–149, **146–148**; and context-dependent risk 110–112; as dynamic and contextual 111–112; dynamic cycle *152*; ethical and regulatory readiness 133–136; existing models 121–122; frameworks 118, 156; functional analysis of *129*; governance, accountability, and assurance 143–144; human–AI interaction readiness 136–137; literature on 122; models 118–122; operating model *258*; organisational dimensions of 114–116; organisational-level readiness models 119; post-deployment and continuous adaptability 139–141; risk and security readiness 141–142; sector-specific readiness **120**; strategic and cultural readiness 130–132; technical and operational readiness 132–133; technology adoption models 112–114; trust and fairness in 116–117
AI Readiness and Adoption Assessment Framework 155, 164–172, *171*; AI adoption roadmap 168–169; assess 166–167; define strategic alignment 167; instruments and templates 170; methodology 170–172; metrics 167–168; pilot & scale 169; scope and objectives 164–166; structure **165**
AI Readiness and Maturity Framework 151, 153
AI Readiness Blueprint 24
AI Readiness & Management Framework (aiRMF) 111, 119
"AI revolution" 180–181
Airline Operator 115
AI-RMM framework 116, 121, 126–129; cluster analysis 129–144, 149; dynamic approach 145–149; eight interdependent clusters 126–129; feedback loop 150–151; responsible AI adoption 149–150; unified *128*
AI Transformers/Systemic Adopters 157–158
"AI Wall" 76, 77, 99–102; absorptive capacity and speed of assimilation 99–100; experimentation economics 100; geopolitics, data sovereignty, and strategic control 101; hybrid and edge architectures 101; organizational to physical constraints 100; reframing 101–102; structural constraints and global asymmetries *102*
Ajzen, I. 113
Alahakoon, D. 165, **165,** 171
algorithmic accountability 269
"algorithmic dehumanization" 300
Algorithmic Impact Assessments (AIAs) 126
algorithms: in business models 301; in corporate decision-making processes 298; as decision-making structures 300; ethical adoption of 302–304; fairness 377–378; long short-term memory (LSTM) 231; pricing 303; and relationships 300; sampling criteria **304**; scaling 86; trading 190
Allen, Rebecca 14
"allergy medicine" 199
Amazon 198; foundational model 199
ambiguity/emotional complexity 256
Americans with Disabilities Act (ADA) 43
Anadol, Refik 14
Analytical AI Adopters 159
Andreessen, Marc 7
Anthropic's Claude 3.5 251
anticipating demand 184
anti-money laundering (AML) 190
applicational convergence 9
application programming interface (API) 251
Ariffin, A. A. M., 299
ARIMA/LSTM baselines 200, 213
Arsel guidelines 304
artificial general intelligence (AGI) 212
artificial intelligence (AI) 1; A&D sector *see* aerospace and defense (A&D) sector; analytical strengths 10; artworks and 28–34; and blockchain 7, 9–10; business in cultural heritage 22–26; commercial viability in 50; communication theories to 245–247; and context-dependent risk 110–112; and cultural heritage 24; democratized market for 10; drawbacks 9; environmental sustainability in 92–93, *93*; financial institutions 13; as general-purpose technology 77; implementation in cultural

institutions 23; investments 2; landscape 3; leaders 34; maturity 79; maturity models *see* maturity models; as operating capability *80*; paradox *78*; readiness *see* readiness, AI; readiness and adoption challenges 28–29; recruitment and retention 286; rules *vs.* 205; strategy blueprint *see* strategy blueprint; technological convergence of 10; training 287; value creation 23, 251–254; in world of business 179–181
Artificial Intelligence Data Readiness and Integration Network (AIDRIN) 166
Artivive 22
ArtPI 28
"art tech" market 23
artworks: and AI 28–34; classification 26; contemporary 38; market 31; recognition to storytelling 28
asset-specific layer (model security) 330–331, **331,** 332, 333
assurance 143–144; trilemma 319–320, *320*
Athreye, S., 8
Attribute-Based Access Control (ABAC) 329
attribute-based classification 24
audit trails **392,** 393
augmented reality (AR) 33, 44
augmenting trust 351–361
authorship 43
automation 225–226; B2B & SaaS reorder 202–203; design framework for 358–361; implementation phases of trade registry **361**; in legal services 351; role of 361; to strategic engagement 251–254; technical bundles 201–202
Automation Ethics Charter 95
AutoML 201
automotive industry 230
awareness recurring stages 85
Ayrey, Andy 7

Barcelò Group 299, 300, **305,** 305–307, 310, **310**
baseline maturity 128
Baxter, P. 304
B2B & SaaS reorder automation 202–203
Bekar, C. T. 9
Berlo's SMCR model 255–256, 272
BERTopic 372, 374, *374,375*
bias 377–378, 398; detection 134; and fairness checks **392**; mitigation strategies 59; and privacy 400
BigQuery SQL queries 201
'black boxes' 318; phenomenon 234n2
blockchain 3, 7; artificial intelligence (AI) and 7, 9–10; core principles 11; implementation of 13; for 3D printing data file 14
Borowiecki, K. J., 25
Boston Consulting Group (BCG) 23, 34
Bostrom, N. 86, 91
BrainChip Akida processor 330
brand monitoring 186
brand purpose 298–311; Airbnb 298–301, 304, **307,** 307–309, **310**; Barcelò Group 299, 300, **305,** 305–307, 310, **310**; beyond storytelling 301–304; ethical adoption of algorithms 302–303; ethical perspective 303–304
Brazil: National AI Strategy 89; SENAI-AI initiative 83
bridging talent shortages 207
Brooklyn microgrid 14
Bröring, S. 8
Brown, T. 51
building trust 58–59
bullwhip effect 197, *197*
business: in cultural heritage 22–26; domain 398; implications 379–381; leadership 384, 385; marketing 185; metrics 24; model 30; perspective 41; value 17
business functions *365*; challenges and ethical factors of 191; contemporary business landscape 180–181; foundational pillars of 181–184; optimizing operations 186–187; pervasive impact 184–191; revolutionizing human resources 188–189; revolutionizing marketing 185–186; transformative imperative of 179–181; transforming finance and accounting 189–191
business models: algorithms in 301; differentiation 33
business-oriented research 25
business-to-business (B2B) 199
business-to-consumer (B2C) contexts 242
business transformation, unified framework for 109–153; adaptability and VUCA environments 117; AI and context-dependent risk 110–112; organisational dimensions of readiness 114–116; technology adoption models 112–114; trust and fairness in AI readiness 116–117
Buyl, T. 88

cadence 200
Cakravastia, A. 230
California with its Consumer Privacy Act (CCPA) 325
Call Me VasarAI 21–47; accessibility and inclusion 42–43; AI business in cultural heritage 22–26; with augmented and virtual reality 33; CLIP embeddings 27; competitive advantage 31; competitive landscape and market positioning 29–33; dataset selection and preparation 35–37; design philosophy 42; development and refinement of 41; educational and cultural

impact 41–42; educational mission 31; framework 34–39, *45*; future development priorities **46**; GPT-Neo 1.3B 38; hybrid model 33; interplay between art and AI 28–34; limitations 43–44; objectives and description 26–28; prompt enrichment and multimodal fusion 38; system architecture overview 34–35; technical performance **40,** 40–41; user interface and deployment 38–39; value proposition 41
Capability Assessment Model 156
capability-based models 118
capacity, readiness as 111
carbon footprint of intelligence 90–91
Carlaw, K. I., 9
Carrefour 12, 157
Casetext 368, **377**
central message 244
Cetinic, E., 24
Chakravorti, B. 79, 82, 89, 91
change management 2, 137–139, 337, 379–381
chatbots 243, 247
Chesky, Brian 307, 309
CIDOC Conceptual Reference Model 23
circular design 91
circular economy 229–231
circular sustainability 98
Cisco 189
classical adoption models 114
classical resource-based theory 78
classic communication model *255*
client-facing communication systems 260
client-oriented communication 242–245, 257–260, 273; AI–human interface in **273**
client preparedness 137–139
client value construction, in digital age 245–247
Clio 368
Clio's Legal Trends Report 352
cloud-based AI logistics platforms 82
cloud systems 396
cluster analysis 129–144; synthesis of 144–149
cluster-by-cluster analysis 130
clustering 204
Cohen, W. M., 99
collaborative ecosystems 98–99
collaborative resilience 226
collaborative robots 83
Colombia's AgroTIC 96
commercialization, of AI solutions 74
commercial viability, in artificial intelligence (AI) 50
commitment, readiness as 111
communication models 255
communication theories 245–247
communicative AI 268–271
communicative environment 252, *253*
"community commitment" 307, 309
companies AI adoption, classification framework for 156–164; AI-Core Value Providers 157; AI Experimenters/Piloters 157; AI Innovators/Creators 158; AI Transformers/Systemic Adopters 157–158; Customer-Facing AI Adopters 158–159; Decision Support and Analytical AI Adopters 159; efficiency and cost-reduction seekers 161; industry sector 162–164; Operations and Process AI adopters 159; Product/Service Core AI Adopters 159–161; revenue growth and innovation drivers 161; risk mitigation and compliance focus 162
COMPAS recidivism tool 377
competitive advantage 2, 23; sustainable *see* sustainable competitive advantage
competitive differentiation 52
competitive landscape 29–33
complement readiness frameworks 122–123
compliance 134, 190; and ethics **388**
computed tomography (CT) 9
computer vision art classification 23
consumer behavior prediction 184
contemporary digital ecosystems 251
content generation 186
context: of artificial intelligence 111; relevance 370
context-aware: legal AI 369–377; text generation 25
context-dependent risk 110–112
continuous adaptability 139–141
continuous auditing 329, 379
Continuous Integration, Continuous Delivery (CI/CD) 58
continuous-learning mindset 140
continuous monitoring 183, 338
Contract Intelligence (COIN) 157
Contrastive Language–Image Pretraining (CLIP) 24–25, 27, 36; multimodal embeddings *39*; Multimodal Enhancement component 35; semantic embeddings with 37
Controlled Unclassified Information (CUI) 325
control-oriented machine 77
conventional approach 54
convergence: analysis 8–9; conditions for 14–16; as four-step process 9; notion of 8–9; rise of 7–8; success factors and barriers **16**
convolutional neural networks (CNNs) 24, 221, 227
cooperative mechanisms 89
core marketing database 54
core security 328–329

corporate communication 242
cost-reduction seekers 161
credit scoring 190
CRM Integration Playbook 204–205
cross-cultural governance models 104
cross-functional teams 389, 391
cross-functional training 386
cross-sector partnerships 25
cryptocurrency wallets 7
crystal-clear value proposition 51
C-suite assurance toolkit 338–339
cultural heritage: AI adoption in 23; applications 30; business in 22–26; intersection of artificial intelligence (AI) and 24, 26; key performance indicators 24; technical metrics 24
culture 124, 288–289; dimension 130; institutions 23, 25; and leadership 115; market 22; readiness 120; storytelling 25
curation 186
customer-centric innovation 77, 81–82
customer communication, AI applications in 247–251
customer engagement 284
Customer Experience Theory (CET) 245
Customer-Facing AI Adopters 158–159
customer lifetime value (CLV) expectation 52, 57
customer segmentation 203–204, 221, **310**
customer-service chatbots, 85
cutting-edge models 163
Cybersecurity Maturity Model Certification (CMMC) program 325

data 258; accessibility 53; cleaning and curation 36; governance 36–37, 328, 379; lineage **392**; maturity 335; and model integrity 330; privacy 378; quality assurance 53; sharing 323; silos 208; sovereignty 101; storage and processing 322–323
data collection: and generation 321–322; and preparation 183
DATALAND 22
data management 52–53; and privacy 317–340, *327*; secure 326–331, *327*; system 281, 283; traditional 319
Data Protection Impact Assessment (DPIA) 271
data readiness, for AI solutions 52–53, 114, 124, 206; assessment 64–68, **65–67**; data management 52–53; strategies for 53
data retention, and disposal 323–324
dataset-agnostic design 36
Davenport, T. H., 77
Day, G. S., 16
DCC4B + AI framework 96, 98–99, *104*; circular sustainability 98; collaborative ecosystems 98–99; digital efficiency 98; integrative mechanism 99; strength of 99
De Bie, E. 88
decentralized energy trading 13
decision-making: foresight and strategic 183–184; processes 179
decision-support: AI adopters 159; tool 366
DeepFleet foundation model 164
DeepInspect 164
deep learning 182; models 35, 163
deep networks 182
defense-specific cybersecurity mandates 325
dehumanization 300
delivering tailored experiences 57
delta detection 55
demand: forecasting 183–184; sensing 198–199
demonstrated maturity 128
deployment 38–39, 183
de-risking investment 51–52
descriptive analytics 183
descriptive storytelling 26
design thinking 51
diagnostic analytics 183
differential privacy 323
differentiated approach 54–55
digital, circular, collaborative for business (DCC4B) model 77
digital/digitalization 278, 279, 281; advertising expenditure 185; age, client value construction in 245–247; to AI-driven decision support 367; assets 7; communication systems 248; divide 83; economy 244; efficiency 98; finance 12; guides/chatbots 28; mediated services 245; transformation 288–292; value chain **282**
digitally native organizations 300
digital twins 225–226, 230
disruption, notion of 8
disruptive technologies 8
dissemination 323
distribution service 55
Diversity Recruiting App 308
DoNotPay 368, **377,** 378
dynamic capabilities view (DCV) 226, 227
dynamic pricing 203, 209, **310**; issue of 303; transparency 214
dynamic readiness cycle 145–150, *152*

ecological accountability 92
economic potential 23
economics of experimentation 100
ecosystem orchestration 86
edge-AI deployment 81
edge computing 82
education 234

efficiency 186–187; and cost-reduction seekers 161; to ethics 366; *vs.* engagement 88
EfficientNet 35, 37
Elasticsearch-enabled UX 54, 55
electronic health records (EHRs) 13
EleutherAI 38
Elgammal, A. 24
email validation algorithms 55
embedded governance 143, 144
emerging business opportunities 10–14; energy and sustainability 13–14; finance and trade 12–13; healthcare applications 13; intellectual property 14; supply chain management 11–12
emerging computational paradigms 330
emerging tension 366
emotional intelligence 241, 254, 272
emotional readiness 138
emotional value 246
empathy-driven interactions 242
employee retention 188
empowerment 233–234
energy 13–14; geography 92
enhanced accessibility 43
enterprise SaaS lead prioritization 205
environmental performance 230
environmental, social, and governance (ESG) 82, 86, 198; integration 77; monitoring systems 94
environmental sustainability 25, 90–92; artificial intelligence (AI) and 92–93, *93*; carbon footprint of intelligence 90–91; circular design and lifecycle analytics 91; green AI, governance for 91–92; hybrid and edge architectures 91; social and geopolitical dimensions 92; sustainable intelligence 92
ERP-AI integration 225–226
error, high cost of 54
E-S-QUAL model 245
ethical and regulatory readiness 133–136
ethical/compliance domain 398
ethics/ethical AI 122, 258; AI materiality 92; checklist 68, **68–73,** 74; compliance and **388**; and compliance readiness checklist **135**; efficiency to 366; emerging AI-specific governance and 326; frameworks 379; governance and 123; governance, standards, and education 234; implementation matrix **233**; implementing 125–126; and sustainable design 27; of sustainable supply chain management (SSCM) 232–233; as "systemic responsibility" 133
ethics-integrated maturity frameworks 124–125
EU General Data Protection Regulation (GDPR) 269, 271, 325, 378, **388,** 390
European Accessibility Act 43
European Commission report 2021 15
European Union Aviation Safety Agency (EASA) 326
European Union's AI Act 3, 91, 129, 140, 141, 145, 150, 326, 379, **388,** 390
Everledger's blockchain 12
evidence-based synthesis 219
Evisort 368
existing AI readiness models 121–122, 155
experimentation: economics 100; and pilots stage 85–86
explainability tools 89, **392,** 398
explainable AI (XAI) 162, 228, 267, 269; methods 231; with participatory design practices 271; techniques 59
exploration recurring stages 85
exponential organizations (ExO) 300, 311
Export Administration Regulations (EAR) 326
external consulting, in-house AI development *vs.* 207

facial recognition 182
fairness: functions 116–117; testing 134
faithfulness score comparison across models 370, *372*
family enterprises 96
fatigue monitoring 334–335
federated learning 378, 396; for collaborative predictive maintenance 333–334
feedback loops 126, 230
Fenwick, A. 287–289
finance 12–13
financial institutions 184
financial resources 8
financial services firm 114, 162–163
financial-services organisation 149
first-order disruptions 8
flexibility 26
forecasting demand 221–224
forecasting errors 197
forecast quality 209
foresight and strategic decision-making 183–184
forward-looking enterprises 92
forward-looking organizations 198
forward-thinking firms 387
foundational layer (governance) 328, **331,** 332, 333
foundational pillars, of AI in business 181–184; algorithmic prowess 181–182; machine learning (ML) 181–182; natural language processing (NLP) 182–183; predictive analytics (PA) 183–184
foundation model operations (FMOps) **392,** 395
fragmented customer identity 54

Frangos, P. 287–289
fraud: detection 12, 189–191; prevention 12; transactions 12
functional roadmaps 2
functional value 246

Gemini app 30
Gemini 2.5 Flash Image model 30
Gemini 3 Pro 30
General Data Protection Regulation (GDPR) *see* EU General Data Protection Regulation (GDPR)
general-purpose generative AI models 248
general purpose technology (GPT) 9, 77
generative adversarial networks (GANs) 285
generative AI (GenAI) 26, 33, 161, 163, 186, 203, 228, 285, 291, 365, 368, 381, 398
generative large language models 218
generative simulation 210
generic AI adoption 25
geocoding APIs 55
geopolitics 101
Getty Research Institute's Art & Architecture Thesaurus 23
GiuriMatrix 366, 369–377; context and motivation 369; data collection and preparation 369; insights gained 376–377; with LegalTech platforms **377**; RAG assessment and topic modelling 369–375; retrieval-augmented generation (RAG) 369–372, *371,***372,** 376
Globalstar 16
Goertzel, B. 10
Google Arts & Culture 30; Art & Selfie 22
Google's Gemini 2.0 251
Google's Gemini 2.0 Flash 248
go-to-market strategy 360–361
governance 78, 79, 82, 143–144, 208, 328, **331,** 332, 388–391; data maturity and 335; embedded 143, 144; ethical AI frameworks 234; and ethics 122; frameworks 2; for green AI 91–92; measurement and 208–209; and mindset 115; oversight in AI 114; value and 301
"governed productivity" 81
GPT-Neo 27, 35, 36, 38, *39*
GPT-Neo 1.3B 38
GPT-4o 248, 262, 370
GPT-4o mini 370, 371, **372,** 376
GPT-3.5 Turbo 370, 371, **372,** 376
Gradio interface *39*
Grand View Research (2024) projects 76
graphics processing units (GPUs) 182
graph neural networks (GNNs) 232
Grashof, N. 81
Green AI 37, 38, 45; governance for 45, 91–92; practices 25, 29, 45
green supply chain metrics 230
greenwashing by algorithm 91
Grossman, G. M. 83
grounding AI 51

Habermas, J. 254
Haenlein, M. 300
hardware security modules (HSMs) 329
hazards 390
healthcare 209; applications 13; and life sciences 163
HighQ 368
high-tech and digital natives 162
high-ticket B2C 209
Hoffman, R. 78, 98
HSBC 12
human–AI interaction: collaboration 83; readiness 136–137
human-centered approach 51
human-centric AI 287–288; for empowerment 233–234
human communication strategies 242
human–computer interaction 252
human-in-the-loop (HITL) 57, 228, 260, 271, 378, **392,** 394
humanizing AI 286–287
human network analysis 189
human oversight 254–257, 369, 378
human resource management (HRM) system 287–288
human resources 188–189
hybrid advantage 207
hybrid AI systems 163
hybrid and edge AI architectures 91, 98, 101
hybrid communication models 241, 255, 256, *256,* 257, 272–273

Iansiti, M., 77, 81, 86
IBM Food Trust 11
image classification 27
ImageNet 27, 40
image recognition, with ResNet18 37
incident response plan **388,** 398
inclusivity 233–234
Industrial AI Foundation Framework 164
industrial convergence 9
Industry 5.0 frameworks 83
industry leadership 86–87
industry sector 162–164; financial services 162–163; healthcare and life sciences 163; high-tech and digital natives 162; manufacturing and automotive 164; retail and consumer packaged goods (CPG) 163–164
influential communication models 254

information governance 328
Ing, L. Y., 83
in-house AI development *vs.* external consulting 207
Innovation Diffusion Theory (IDT) 112
innovations 8, 16; drivers 161; ecosystem 367–369; *vs.* regulation 89
Institutional and Legitimacy Theory 232
integrating behaviour 113
integrative mechanism 99
intellectual property (IP) protection 7, 14, 331
intelligence, surveillance, and reconnaissance (ISR) 318
intelligent automation 187
intelligent document processing (IDP) 161
intercultural negotiation 156
interdependent clusters 126–129
internal *vs.* external 207
Internet of Things 12
interpretive variables 254
interviews 10, 13, 132, 133; distribution of *11*; iterative process 305
inventory management 284–285
inventory optimization 187
inventory waste 197
investment optimization 189–191
Ironclad 368

Jack, S. 304
Jakobson, R. 254
JavaScript Object Notation (JSON) 15
Johann, M. 130
JPMorgan Chase 157

Kabbage 95
Kaplan, A. 300
Keeble, D. 8
Kelm, K., 130
key performance indicators (KPIs) 24, 98, 157, 199, 230, 267, 280, 288–292, 379, 391; for business and social impact **42**; for culture 289; for leadership 290, 291; operational 338–339; for policies 291; smart 292, **293,294**; for tools 292; traditional 292, **294**
kickstarting transformation 207
Kimberley process 12
Kira Systems 368
Klingemann, Mario 14
knowledge 124, 290–291; convergence 9; readiness 114
Kopka, A. 81
Krippendorff's guidelines 305
Kurniawati, D. 230

Lakhani, K. R. 77, 81, 86
language 24
language translation 182
large language model operations (LLMOps) **392,** 395
large language models (LLMs) 27, 163, 211, 212, 367; for cultural storytelling 25
large-scale models 22; training 89
Lasswell, H. D. 255
last-mile delivery 225
latent Dirichlet allocation (LDA) 372, *373,* 374
LawGeex 368
leadership 29, 105, 286, 290–291, 384; and buy-in 337; commitment 115, 138; culture and 115; engagement 130; judgment 89; support 115, 131; transformative 115
lead scoring 209
least privilege 329
Leavitt, H. J. 280
Leavitt's Diamond model 280
legal sector: AI adoption and legal industry 364–366; AI readiness assessment framework for **380**; business implications and change management 379–381; context-aware legal AI 369–377; ethical and governance landscape 377–379; integrating AI into 380–381; legal technology's evolution 366–369; organizations using AI *364*
legal services: automation in 351; framing automation and artificial intelligence in 352–353; methodology 353; Romanian legal market 354–357; small law firms 358–361; technology–organization–environment (TOE) framework 358; use-case 354
legal technology's (LegalTech) evolution 366–369; digitization to AI-driven decision support 367; innovation ecosystem 367–369; market growth 367–369, *368*; stages of 367
legal transformation 365–366
leveraging digitalization 279
Levinthal, D. A. 99
lifecycle analytics 91
likelihood-to-buy score 204
Lipsey, R. G. 9
Litify 368
loan underwriting 190
logistics 161; planning 225
long short-term memory (LSTM) 221; algorithms 231; forecasting models 229
long-term value creation 57
loss of control 300–301
low-risk domains 110

machine learning (ML) 155, 181–182, 205, 245; algorithms 9, 183; approaches 24
machine learning operations (MLOps) **392,** 395

Maghzi, A. 299
Magnus 30–33
manufacturing and automotive 164
Mapping Generation module 35, 38
margin-safe discounts 203–204
marketable AI solutions 50–74; commercialization 74; cornerstone of marketability 51–52; data readiness for 52–53; de-risking investment 51–52; differentiated approach 54–55; ethical AI checklist 68, **68–73,** 74; ethical frameworks 58–59; fragmented customer identity 54; grounding AI 51; leap 56; practical toolkit for 60–74; technical curiosity to business imperative 56–57; transforming customer intelligence 54–57; value proposition template 60, **60–64,** 64; wider landscape of 57–58
market growth, of LegalTech 367–369, *368*
"Marketing Hub" 306
marketing optimization **310**
market positioning 29–33
market segmentation 267
massive transformative purpose (MTP) 300, 304, 309
master data management (MDM) 53, 54, **56**
Match & Merge engine 54
Mathematical Theory of Communication 254
maturity models 122–126; ethics-integrated maturity frameworks 124–125; organisational-level maturity models 123–124
McGee, F. 117
McKinsey & Company (2025) 23, 76, 86
meaningful human control (MHC) 328
mentorship 386
Mercado Livre 94–95
meta-communication layer 255
metadata management 393
methodological humility 85
metrics and model evaluation 397
Microsoft tool **250**
mission capability 319
MOBGEN 29
model: development 183; drift **392**; governance 379; integrity 330; management 393; provenance 331; security 332; validation and diagnosis 183
model-optimization techniques 82
modern lead scoring 204
Modified-Poisson Gamma (MPG) model 202, 213–214
Molnar, G. 287–289
Monet, Claude 36, 38
monitoring 183, 329
Montes, G. A., 10
MOTChallenge datasets 170
multi-agent systems 210
multimodal fusion 38
multimodal learning museums 23
multimodal machine learning 163
multinational firms 83
multi-object tracking methods 163, 170
multiple technologies 7
multi-source approach 320
MuseumNext 22

named entity recognition (NER) 55, 182
Nano Banana Pro 30
"narrative generation" process 35, 37, 38
natural language generation (NLG) 27, 182
natural language processing (NLP) 55, 179, 182–183, 245, 267; Generative AI tools 186
Navarrete, T. 25
negligible human intervention 181
Nestlé 12, 94, 98, 198
neuromorphic computing 330
new business models 81–82
next-order prediction 200–203, 209, 210, 212
non-AI technology projects 384
Nuveo (Brazil) 82, 95–96

objectives and key results (OKRs) 51
Octopus Organization perspective 78
ongoing monitoring 393
online face-to-face interviews 10
online transactions 10
ontologies/controlled vocabularies 25
OpenAI 37, 91; GPT-4o 248; tool **250**
open data 35
open innovation (OI) 227
open-source frameworks (PyTorch) 37
operating capability *80*
operating model 388–391
Operational AI adopters 159
operational constraints 319
operational domain 398–399
operational efficiency 77, 84, 179, 278; technology integration for 81
operational integration 86
operational KPIs 338–339
operational resilience 270
operations optimization 285
optimizing operations 186–187
organisational AI readiness 111, 114–116, 257–260
organisational communication 243
organisational-level: maturity models 123–124; readiness models 119, 123–124
organisational perspectives 113
organizational AI adoption 23
organizational changes 278–295; AI-digital transformation 280–285; AI-driven digital

transformation 288–292; AI-driven KPIs 292–295; customer engagement 284; digitalization 279; inventory management 284–285; operating model 279; operations optimization 285; staff resistance and humanizing AI 286–287; strategic challenges for 285–288
organizational culture 337
organizational literacy 85
outcome-oriented metrics 267
outsourced development team 360
overestimate demand 197

Paladin 368
participatory design practices 271
pattern recognition 2
PayPal & Visa 190–191
peer-to-peer insurance 13
people 258, 384–387; strategy blueprint 384–387
perceived authenticity 302
persana AI 204
personalization 57
personalized experiences and engagement 185
Picasso 36
pilot health 334–335
PlantAI 28
point-of-sale (POS) 198, 199
policies 291
portfolio optimization 190
post-deployment readiness 139–141
potential clients, AI value for 241–274, *242*; AI-mediated communication 268–271; classic communication model *255*; client-oriented communication 242–245, *246*; client value construction in digital age 245–247; communicative functions dynamics **249–250**; coordinated execution framework of *259*; customer communication, AI applications in 247–251, **248**; human oversight 254–257; incidents reported *265*; limits in communication 254–257; notable AI models *261–263*
practical recommendations, for managers 16–18, *17*
predictability 53, 58
predictive accuracy 81
predictive analytics (PA) models 183–184, 226
predictive inventory management 284, 295
predictive inventory optimization 198
predictive lead scoring 204–206
predictive maintenance 187, 333–334
predictive procurement 225
primary data collection 304
principled layer (security-by-design) 328–329, **331,** 332, 333
PRISMA framework 219–221, *222*
privacy 317–340, *327,* 398; bias and 400
privacy-by-design methodology 328, **388,** 390, 391
privacy-enhancing technologies (PETs) 329
privacy-personalization 303
privacy-preserving architectures 13
proactive threat intelligence 338
process 258; strategy blueprint 388–391
Process AI adopters 159
product mindset 51
Product/Service Core AI Adopters 159–161
professional judgement 361
programmatic purchasing 185
promotions 203–204; economics 209
prompt enrichment 38
"proof-of-concept trap" 391
provenance trust 11
public organizations 8
public–private ethics 89
public–private initiatives 99
public-sector interviewee 133
public-sector representative 140, 141
purpose-driven approach 303
"purpose orientation" 302
PwC's Responsible AI Framework 123, 131, 134, 143

quality assurance (QA) 182
quantum-resistant cryptography (QRC) 329

real-time fraud detection 190
recurrent neural networks (RNNs) 227
recursive learning system 87
regenerative (REGEN) purpose 304, 309
regional promotions 209
regional retailers 82
regulatory ambiguity 15
regulatory compliance 379
reinforcement learning 181, 224
Reinforcement Learning from Human Feedback (RLHF) 56–58
reinforcing ethical mindset 136
reinventing inventory control 224–225
Relativity tool 368
remote patient monitoring 13
Renaissance Florence 23
replenishment-driven categories 202
resilience 88, 186–187; adaptation for 338; growth framework 210, *211*; readiness to 335–339; supply chain 195–215, 226–228
reskilling investments 83
ResNet 27, 36
ResNet18 35, *39*; image recognition with 37
ResNet50 37

resource allocation 123–124
resource-based view (RBV) 226
resource readiness 114
responsible AI (RAI) 45, 59, 82, 124, *265,* 266, **388**; adoption 149–151; frameworks 264, 379, 390
Responsible AI Maturity Model (RAI MM) 124
retail and consumer packaged goods (CPG) 163–164
retention 387
retrieval-augmented generation (RAG) 369–377, *371,***372,** 376
return on investment (ROI) 51–52, 157, 379
revenue generation 42
revenue growth 161
revenue S&OP 207
review regularly 401
revolutionizing marketing 185–186
risk 388–391; mitigation 56, 162; predictive and proactive 227; and resilience mindset 142
risk management 184, 187, 189–191; of digital value chain **282**; ethical and responsible 228; and investment optimization 189–191; supply chain resilience and 226–228
robotic process automation (RPA) 81, 187
robust adversarial defense 330
robust testing 337
Rogers, E. M. 112
Role-Based Access Control (RBAC) 329
Romanian legal market 354–357
Ronanki, R. 77
Rosenberg, N. 8
Rouen Cathedral (Monet) 21
rules-based scoring 205
Ryan, M. 136

Sachs, J. D. 77, 79, 89, 92
Saleh, B., 24
Salesforce 267
satellite-based mobile telephony 16
satellite imagery and signals intelligence (SIGINT) 332
Satra**,** H. S. 79, 82, 86, 92
scalability 26, 53, 396; future-proofing with 58; and predictability 58
Schoemaker, P. J. 16
Schuelke-Leech, B. A. 8, 9
scorecard, communicate and act on 400–401
score comparison across models: answer relevance 370, *373*; context relevance 370; faithfulness 370, *372*
second-order disruptions 8
sector-specific readiness models 120, **120,** 121, 122
secure computing environments 329–330, **331,** 332
secure data management 326–331, *327*; AI security framework **331**; asset-specific layer 330–331, **331**; foundational layer 328, **331**; principled layer 328–329, **331**; technological layer 329–330, **331**
secured mobile enabled assisting device (SMEAD) 13
secure multi-party computation (SMPC) 329
secure supply chain management 337
security 378; readiness 142
security-by-design methodology 328–329, **331,** 332, 336–337
segmentation 210
self-determination theory 293
self-directed learning 41
self-reinforcing cycle 79, 84
semi-autonomous communicative agents 245
semi-structured interviews 10
sentiment analysis 182, 186
Sephora Virtual Artist 247
service-level agreements (SLAs) 358
service-management perspectives 155
Seurat, Georges 39
Shannon, C. E. 254, 255, 272
She, J. 24
Siemens 94, 98, 157, 187
Silva, D. D., 165, **165,** 171
Singapore's SkillsFuture 83
Six Sigma and Lean operations 279
small and medium enterprises (SMEs) 218, 233
small and mid-sized enterprises (SMEs) 81
small law firms 358–361
smart city applications 13
smart contracts 12, 13
Smartify 30–31, 33
smart KPIs 292, **293,** 294, **294**
smart segmentation, foundation of 204
social and environmental responsibility 79
social and geopolitical dimensions 92
social enterprise 29
socio-technical system 79
source-message-channel receiver (SMCR) model 255–256
specialist knowledge integration 201
stacking, transformers *vs.* 205–206, 214–215
staff resistance 286–287
stakeholders 13, 51; engagement 379; participation 389
Stanford HAI AI Index Report (2025) 23
Stark Reality of 2026 2
static/linear assessments 122
stewardship theory 328
The Storm (Turner) 21
storytelling, recognition to 28

strategic engagement, automation to 251–254
strategy 258; alignment 114, 123, 167; control 101; and cultural readiness 130–132; and cultural realignment 368; data hygiene 53; expansion 86; integration 336–337
strategy blueprint 384–401; aggregate scores by domain 399; benchmarks and best practice targets 400; business domain 398; ethical/compliance domain 398; executive AI landscape **385**; executive as chief 401; metrics and model evaluation 397; operational domain 398–399; people 384–387; process 388–391; review regularly and benchmark externally 401; scorecard, communicate and act on 400–401; technical domain 398; technology 391–396; visualizing 399–400
stress testing 58, 190
structural equation modelling (SEM) 226
structured metadata 25
SuperAGI 204
supervised learning 181
supplier evaluation 225
supplier's perspective 115
supply chain 161; B2B & SaaS reorder automation 202–203; customer segmentation 203–204; excellence 186–187; management 11–12; in manufacturing organization 279; margin-safe discounts 203–204; next-order prediction 200–203; promotions 203–204; resilience 187, 195–215
supply chain–marketing nexus 195–198
sustainability 13–14, 45, 229–231; engagement 183; intelligence 92; *vs.* computation 89
sustainable competitive advantage 76–105; AI maturity model *84,* 85, 87, *87*; as artificial intelligence 79; awareness and exploration stage 85; classical resource-based theory 78; core dimensions of 80; cross-case insights 96–97; DCC4B + AI framework 96, 98–99, *104*; digital divide and workforce transformation 83; environmental sustainability *see* environmental sustainability; ESG and responsible AI 82; experimentation and pilots stage 85–86; experimentation to institutional 85–87; family enterprises and agribusiness 96; governance 79; industry leadership 86–87; insight 94; integrative view 84; interdependent dimensions 80; invariant drivers of *97*; Kabbage 95; managerial implications 103; Mercado Livre 94–95; mid-market firms 94–95; Nestle's AI 94; new business models and customer-centric AI 81–82; operational efficiency, technology integration for 81; operational integration 86; paradox of AI adoption 88–90, *90*; reinforcing dimensions 77; Siemens 94; social and environmental responsibility 79; strategic expansion 86; theoretical foundations 77–80
sustainable supply chain management (SSCM) 217–234; automation and ERP–AI integration 225–226; and circular economy 229–231; digital twins 225–226; environmental and economic impacts of 218; ethics of 232–233, **233**; explainability and trust 231–232; human-centric AI for empowerment 233–234; included papers and thematic coupling **223–224**; logistics and last-mile delivery 225; operational research 218; performance and optimization 221–226; PRISMA framework 219–221, *222*; reinventing inventory control and warehouse management 224–225; reputation, risk, and paradox of transparency 233; resilience and risk management 226–228; smart optimization for 229; supplier evaluation and predictive procurement 225
SustAI–SCM framework 225, 229
symbolic value 246
synthesis 89–90
system architecture *34,* 34–35
"systemic responsibility" 133
systemic risk 352

talent availability 15
tangible business transformation 56
targeted advertising 185
targeted digital-literacy programs 83
technical and ethical integrity 133
technical and operational readiness 132–133
technical curiosity 56–57
technical domain 398
technical foundations 118
technical hazards 270
technical infrastructure 335
technical metrics 24, 41
technological convergence 9
technological diffusion 76
technological environment 17
technological implementation 18
technological layer (secure computing) 329–330, **331,** 332, 333
technology acceptance model (TAM) 112, 245
technology adoption models 112–114; innovation diffusion theory (IDT) 112; integrating behavioural and organisational perspectives 113; practical implications of 113–114; technology acceptance model

(TAM) 112; Theory of Planned Behaviour (TPB) 113; trust-enhanced technology models 113; Unified Theory of Acceptance and Use of Technology (UTAUT) 112
Technology Assistance Agreements (TAAs) 326
technology-first AI 51
technology–organization–environment (TOE) framework 358
technology & platforms 258
technology, strategy blueprint 391–396
Tegmark, M. 86
terminal of truth (ToT) 7–8; cryptocurrency wallets 7
text classification 182
text generation module 40
textiles 230
text-to-speech integration 43
Thai fish supply chain 12
theory-informed analytical strategy 353
Theory of Planned Behaviour (TPB) 112, 113
3D printing data file 14
Toro 198
trade 12–13
traditional KPIs 292, **294**
traditional segments 204
transformation, legal 365–366
transformative imperative, of artificial intelligence 179–181
transformative leadership 115
"transformative logic" 302
transformers *vs.* stacking 205–206, 214–215
transforming customer intelligence 54–57
transforming finance 189–191
transparency 94, 137, 331, 369, 378; AI–blockchain convergence for 230–231; *vs.* complexity 89; dynamic pricing 214; paradox of 232, 233; report **388**; trust and 210
triadic communication model 268
trial-and-error methodology 281
trust 361, 378; augmenting 351–361; explainability and 231–232; functions 116–117, 137; human values and 302; and transparency 210
trust-enhanced technology models 113
trust-enhanced versions of TAM (T-TAM) 113
Trustworthy AI 58
Tucci, C. 160

UiPath 95
underestimate demand 197
unified AI-RMM framework *128*
unified monitoring strategy 338
Unified Price, Promotion, and Markdown Optimization (UPPMO) 203
Unified Theory of Acceptance and Use of Technology (UTAUT) 112
United Nations Sustainable Development Goals 217
unsupervised learning 181
uplift modelling 203, 210
U.S. Department of Defense 326
user-centric design 51
user-centric product development 57
user engagement metrics 42
user experience (UX) 57
user interface 38–39

validation 337
value clarity 51–52
value creation 252; coordinated execution framework of *259*; *vs.* systemic risk *78*
value proposition template 60–64, **60–64,** 64
Van Gogh 36
Vasari, Giorgio 21
Verisart 14
Vijayakumar, H. 16
virtual assistants 247
virtual communication environments 254
virtual reality (VR) 33, 44
volatility, uncertainty, complexity, and ambiguity (VUCA) 117, 179

Walmart 157
warehouse management 224–225
waste management (commoditised sector) 121
wealth-management participant 138
Weaver, W. 254, 255, 272
web-based interface 38
Weiner, B. J. 110, 111, 139
well-designed user interface 39
WikiArt 27, 32, 35–37
workforce: optimization 188–189; planning 184, 188; transformation 77

XGBoost 202, 213–214
X-ray machinery 9

Yin, R. K. 304

zero-shot classification 25
zero trust architecture 329, 337



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com