DEVELOPING RESPONSIBLE AI LEADERSHIP CAPABILITY: A FRAMEWORK FOR GOVERNANCE, ETHICS, AND AI LITERACY

by

Glenn Chilcott MBA MSc

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfilment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JANUARY 2026

DEVELOPING RESPONSIBLE AI LEADERSHIP CAPABILITY: A FRAMEWORK FOR GOVERNANCE, ETHICS, AND AI LITERACY

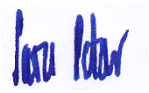
by

Glenn Chilcott MBA MSc

Supervised by

Dr Denis Buterin

APPROVED BY



Prof.dr.sc. Saša Petar, Ph.D., Dissertation chair

RECEIVED/APPROVED BY:

Renee Goldstein Osmic

Admissions Director

Dedication

Dedicated to the memory of my parents, John and Rosina Chilcott, whose example and values remain a lasting source of inspiration.

Artificial Intelligence (AI) Use Disclosure

Artificial intelligence (AI) tools were used in a limited, assistive capacity to support language clarity and editorial refinement. They were not used to generate research content, design the methodology, analyse data, interpret findings, or develop original arguments. All intellectual contributions, including the development of the TRAIL framework, are the independent work of the author, who retains full responsibility for the content of this dissertation.

ABSTRACT

DEVELOPING RESPONSIBLE AI LEADERSHIP CAPABILITY: A FRAMEWORK FOR GOVERNANCE, ETHICS, AND AI LITERACY

Glenn Chilcott MBA MSc
2026

AI has emerged as a general-purpose technology with the capacity to reshape economic activity, organisational structures, and decision-making processes on a global scale. Widely cited economic forecasts suggest that AI could contribute up to $15.7 trillion to global economic output by 2030, positioning it as a central driver of productivity and innovation. However, the acceleration of AI adoption has been accompanied by significant ethical, legal, and societal risks, including algorithmic bias, opacity, data misuse, and the erosion of human agency. These risks challenge existing models of leadership, governance, and accountability.

This research identifies a critical leadership readiness gap at the heart of responsible AI adoption. While an overwhelming majority of professionals recognise AI literacy as an essential leadership competency, empirical findings reveal markedly lower confidence in navigating the ethical, regulatory, and governance dimensions of AI in practice. The results expose a persistent disconnect between aspirational commitments to fairness, transparency, and human dignity and the organisational capabilities required to operationalise these principles effectively.

Adopting a sequential explanatory mixed-methods research design, this study integrates quantitative survey data with qualitative insights to examine how external regulatory pressures and internal organisational practices shape AI implementation. The findings

demonstrate that responsibility for AI governance is frequently fragmented and delegated to technical or compliance functions, rather than embedded within strategic leadership and board-level oversight. This fragmentation contributes to inconsistent decision-making, delayed adoption, and heightened organisational risk.

In response, the research introduces the TRAIL (Trustworthy and Responsible AI Leadership) framework, an empirically derived model designed to support leadership capability development for responsible AI. The framework comprises four interdependent pillars: Strategic and Transformational Leadership; Governance and Regulatory Fluency; Ethical and Risk Literacy; and Ethical Data Management. Together, these pillars provide a structured pathway for translating ethical intent and regulatory awareness into practical leadership competence across the AI lifecycle.

The study further extends Porter's Five Forces framework by proposing Responsible AI Capability and Stewardship as a sixth strategic force, arguing that ethical integrity and governance maturity now constitute critical determinants of organisational resilience and sustainable competitive advantage. This reconceptualization positions responsible AI not as a compliance obligation, but as a strategic leadership capability integral to long-term value creation and stakeholder trust.

The research concludes that the trajectory of AI adoption will be shaped less by technological sophistication than by the quality of leadership guiding its design, deployment, and use. By advancing an integrative leadership framework grounded in empirical evidence, this study contributes to both theory and practice, offering organisations a practical and human-centred approach to developing AI-literate leaders capable of governing AI responsibly in complex and evolving regulatory environments.

TABLE OF CONTENTS

LIST OF TABLES

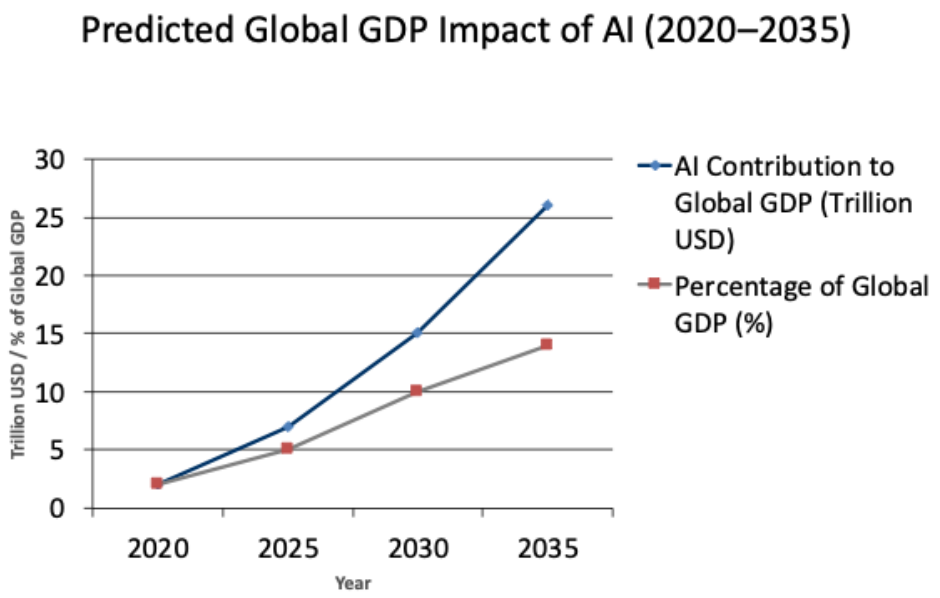
LIST OF FIGURES

CHAPTER I

INTRODUCTION

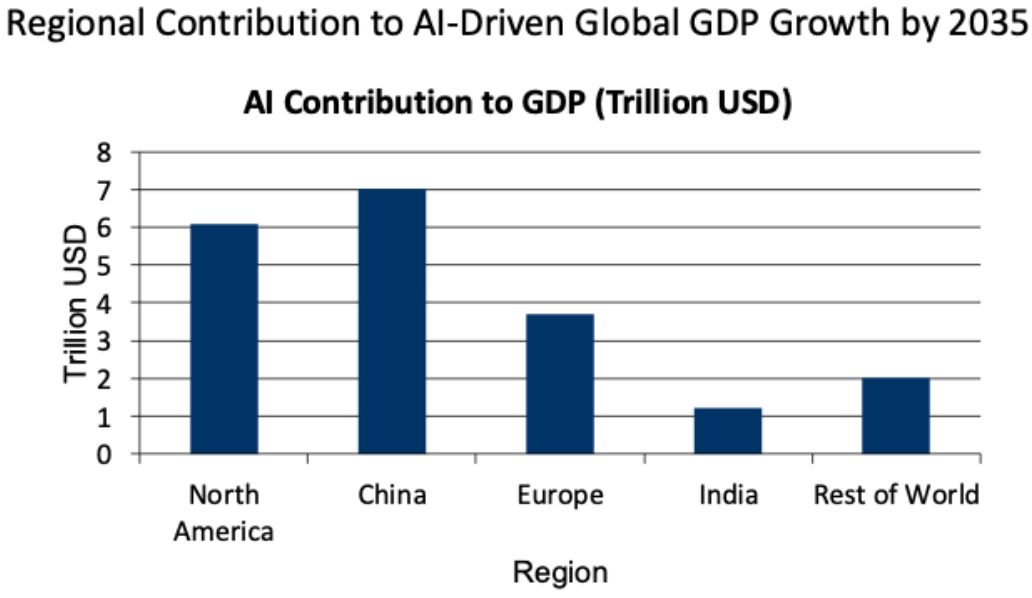
1.1 Introduction

Artificial intelligence (AI) represents one of the most transformative technological developments of the modern era. Its potential to enhance productivity, stimulate innovation, and generate new forms of economic and social value has been widely documented. Global forecasts suggest that AI could contribute up to $15.7 trillion to the world economy by 2030 (PwC, 2027), positioning it as a general-purpose technology comparable in impact to electricity or the internet. Analysts predict that its economic and societal impact could exceed that of previous industrial revolutions, reshaping how organisations operate, innovate, and compete. The economic potential extends beyond sectoral productivity. Goldman Sachs (2023) forecasts that generative AI could boost annual global GDP growth by 1.5 percentage points over the next decade, while the International Monetary Fund (2025) projects an average 0.5 percent annual lift between 2025 and 2030. Such forecasts, if realised, could represent a structural shift in global economic performance, comparable to the transformative effects of electrification or mass digitisation. McKinsey & Company (2023) estimates that AI could generate between $2.6 trillion and $4.4 trillion annually in economic value. Adoption is already widespread, with 78 per cent of organisations reporting use of AI in at least one business function in early 2025 (McKinsey & Company, 2025).



Sources:
PwC (2023) 'Sizing the Prize';
McKinsey Global Institute (2023) 'The Economic Potential of Generative AI'; Goldman Sachs (2024) 'The AI Economy'

Figure 1.1: Predicted Global GDP Impact of AI 2020-2035



Sources:
PwC (2023) 'Sizing the Prize';
McKinsey (2024) 'The Economic Potential of Generative AI'

Figure 1.2: Regional Contribution to AI-Driven GDP Growth by 2025

Figure 1.2 highlights the uneven regional distribution of projected AI-driven contributions to global GDP growth by 2035, with North America and China expected to capture the largest economic gains due to advanced digital infrastructure, investment capacity, and concentration of AI capabilities. Europe's comparatively lower projected contribution, alongside more modest gains across emerging economies, underscores the asymmetric economic impact of AI as a general-purpose technology shaped by differing regulatory, institutional, and leadership conditions (PwC, 2023; McKinsey & Company, 2024).

Yet the same forces that drive innovation also give rise to profound risks concerning fairness, accountability, privacy, and societal well-being. The dual nature of AI, as both opportunity and threat, defines the central challenge of this century: ensuring that technological progress aligns with ethical values, human rights, and long-term sustainability.

1.2 Opportunities, Risks and the Data Foundation of AI

AI systems are already delivering substantial benefits across diverse sectors. AI plays a transformative role in facilitating economic growth by enhancing productivity, promoting innovation, and generating new market opportunities. In healthcare, AI supports diagnostics and drug discovery; in manufacturing, it enables predictive maintenance and process optimisation; and in financial services, it enhances fraud detection, risk management, and customer insight. Furthermore, AI fosters the creation of novel products, services, and business models, thereby facilitating the emergence of entirely new markets. Governments and businesses increasingly view AI as a driver of growth and a tool for addressing complex global challenges such as climate change and resource management (OECD, 2021; World Economic Forum, 2022). The opportunities

are broad and transformative, and these advances illustrate the capacity of AI to augment human decision making and create novel markets.

However, the realisation of these benefits is inseparable from a spectrum of risks. Bias embedded within training data can reproduce or amplify existing inequalities, while opacity in algorithmic decision making threatens accountability and transparency. Privacy violations, environmental costs from intensive energy use, and security vulnerabilities in data infrastructure further complicate the picture. At a systemic level, concentration of technological and data power within a small number of corporations raises concerns about competition, democratic control, and equity. Labour market disruption adds another layer of complexity: while AI may enhance productivity, it also risks displacing workers in roles that are routine or cognitively repetitive. These risks underscore the need for a structured and ethically informed approach to AI adoption, one that acknowledges both its transformative potential and the societal responsibilities it entails. The challenge is not simply to accelerate AI adoption, but to ensure that its development and deployment are safe, ethical, and aligned with societal values.

Consequently, AI can be considered comparable to safety-critical systems. If the gravity of risk is not sufficiently recognized, then this could result in substantial financial penalties, sanctions against executives, significant brand damage, decreased stock valuations, or, in extreme cases, pose an existential threat to organizations. The challenge, therefore, is not simply to accelerate AI adoption, but to ensure that its development and deployment are safe, ethical, and aligned with societal values. As the pace of innovation accelerates, the window for establishing effective safeguards narrows. Failure to address these risks in a timely and coordinated manner could erode public trust, stifle innovation, and ultimately diminish the long-term benefits AI promises to deliver.

Data is the essential substrate of AI. The quality, provenance, and governance of data directly influence the performance, reliability, and fairness of AI systems. As AI applications increasingly depend on vast datasets drawn from personal, organisational, and public sources, concerns regarding privacy, bias, and transparency have intensified. Failures in data stewardship can undermine public trust, expose organisations to regulatory sanctions, and compromise the integrity of AI outputs. Inadequate data stewardship undermines trust and transparency.

Effective data governance requires embedding ethical principles throughout the data lifecycle, from collection and storage to analysis, sharing, and deletion. AI systems routinely process vast amounts of sensitive personal data, thereby posing significant privacy risks. Practices such as anonymisation, encryption, and access control form part of a wider commitment to 'privacy by design,' ensuring that individual rights are preserved even as data driven innovation proceeds. Furthermore, bias remains a significant concern as AI systems frequently inherit societal prejudices embedded within their training datasets. Equally important are fairness audits, bias testing, and the inclusion of diverse perspectives in dataset construction. Addressing bias and representation issues is not a purely technical exercise; it reflects a moral responsibility to ensure that AI serves all members of society equitably. The rapid increase in incidents of data breaches and algorithmic misuse highlights the urgency of this issue. The integration of robust data ethics frameworks and transparent governance mechanisms is therefore fundamental to responsible AI.

1.3 Governance, Regulation and Ethical Frameworks

The governance of AI presents one of the most pressing policy challenges. Regulatory approaches differ across jurisdictions but share a common objective: to ensure that innovation does not outpace accountability and functions safely. The

European Union's AI Act represents the most comprehensive legislative effort to date, employing a risk-based framework that imposes stringent obligations on high-risk systems (European Union, 2024). The United States maintains a more decentralised, sector specific model grounded in existing consumer protection and civil rights laws, while the United Kingdom has adopted a more flexible, innovation first stance that relies on non-statutory guidance and cross sectoral principles.

This divergence has produced a fragmented regulatory landscape that complicates global alignment and increases compliance burdens for multinational organisations. Scholars such as Bradford (2023) describe this as a competition among 'digital empires': the Brussels effect, which exports European Union norms through market power; the Washington model, which privileges innovation and market freedom; and the Beijing model, which emphasises state control and strategic industrial advantage. The absence of harmonisation risks impeding international collaboration and weakening collective capacity to manage shared risks. The goal is to strike a balance between protective oversight and innovation enablement.

Beyond formal regulation, governance must be embedded at the organisational level. Ethical governance involves the translation of abstract principles, such as fairness, accountability, and transparency, into operational practice. This requires leadership commitment, cross functional collaboration, and cultural change. Ethical impact assessments, values canvases, the establishment of roles such as AI ethicists, and the enhancement of AI literacy are increasingly acknowledged as mechanisms for institutionalizing responsibility. Governance, therefore, is not simply a question of compliance; it is an expression of corporate culture and moral orientation.

The governance challenge is multidimensional. It encompasses technical issues such as safety, robustness, and transparency; social concerns including bias,

discrimination, and privacy; and economic factors such as market concentration and labour displacement. Addressing these issues requires not only formal regulation but also a broader culture of ethical responsibility that spans both public and private sectors.

Ethics remain central to the responsible development and deployment of AI. The core principles of fairness, transparency, and accountability must guide every phase of the AI lifecycle. Responsible AI can be defined as the practice of designing and implementing AI systems in ways that are fair, transparent, accountable, and consistent with societal values. Yet ethical AI extends beyond technical compliance: it demands recognition of the human agency that underpins all decision making. While AI systems may demonstrate autonomy in operation, moral responsibility ultimately resides with human actors, designers, developers, regulators, and leaders, who define the parameters of AI behaviour.

Ethical governance extends beyond compliance into the domain of organisational culture. Traits such as integrity, transparency, and fairness are difficult to codify yet essential for long-term trust. Leadership plays a pivotal role in modelling these values and ensuring that decision-making processes reflect them. Embedding ethics into performance metrics, incentives, and professional development can help translate abstract principles into day-to-day practices.

The literature advocates a ‘human approach’ to AI that prioritises human judgement, empathy, and critical reasoning. Such an approach empowers individuals to remain active participants rather than passive recipients of AI generated decisions. Explainability is increasingly recognised as a key component in AI governance frameworks. Regulations such as the General Data Protection Regulation (GDPR) (European Commission, 2018), the European Union Artificial Intelligence Act (European Union, 2024), and sector-specific AI guidelines emphasise transparency and

accountability in AI decision-making. However, while explainable AI (XAI) methods enhance interpretability, they do not inherently provide full normative justifications for AI-driven outcomes. Ethical concerns persist regarding the potential manipulation of explanations, misleading interpretations, and an over-reliance on approximate reasoning.

This perspective reinforces the notion that technology must serve humanity, not the reverse, and that AI should augment rather than replace human capabilities. Ethical AI cannot be achieved through technological means alone. It requires institutions and cultures that promote transparency, encourage whistleblowing, support continuous learning, and embed moral reasoning into strategic choices.

Despite the increasing awareness of ethical AI, the research results reveal that knowledge of governance and regulation remains a significant blind spot. Addressing this issue will be crucial for ensuring that responsible AI principles are not only comprehended but also legally and strategically integrated into practice.

1.4 Organisational and Cultural Implications

The societal implications of AI underscore the need for a more conscious and responsible approach to technological development. The potential of AI to perpetuate social inequalities, disseminate misinformation, and erode trust in institutions is a significant concern.

Fostering a responsible approach to AI has emerged as a defining imperative in the digital era, encompassing the ethical, legal, and societal considerations that govern the design, development, deployment, and use of intelligent systems. It represents a shift from viewing AI merely as a technological tool towards understanding it as a sociotechnical system embedded within organisational and societal contexts. The responsible use of AI requires transparency, accountability, fairness, and human oversight to ensure that innovation aligns with human values and public interest.

Responsible AI (RAI) offers a structured approach that aligns AI development and deployment with ethical, transparent, and human-centred values. Responsible AI can be defined as the practice of designing and implementing AI systems in ways that are fair, transparent, accountable, and consistent with societal values (Jobin et al., 2019). Robust data governance is essential to reduce bias and error by ensuring the quality, traceability, and representativeness of training data. Effective governance can reduce these disparities by embedding quality controls and continuous monitoring.

The concept extends beyond compliance with existing regulation to include proactive governance and ethical foresight. Frameworks such as the European Union AI Act (European Union, 2024), the OECD Principles on AI OECD (2021), and the NIST AI Risk Management Framework (Dotan, 2024) have collectively emphasised that the pursuit of innovation must be balanced with safeguards that mitigate harm, protect fundamental rights, and preserve trust. These frameworks increasingly serve as reference points for organisations seeking to operationalise responsible AI practices within corporate strategy and leadership culture.

From a leadership perspective, responsibility in AI demands more than technical literacy. It requires leaders to cultivate organisational cultures that prioritise ethical awareness, data integrity, and inclusive design. The ability to navigate complex trade-offs between efficiency, fairness, and accountability has become central to effective governance. As AI systems influence decision-making across sectors, the role of leaders in setting ethical direction, ensuring compliance, and maintaining stakeholder confidence has never been more critical.

Responsible AI therefore represents not only a governance obligation but also a strategic opportunity. Organisations that integrate ethical and regulatory considerations

into their AI strategies are better positioned to build trust, attract talent, and sustain competitive advantage.

Corporate culture and leadership competency are critical determinants of whether organisations can effectively develop and deploy responsible and ethical AI. Culture shapes the norms, values and behaviours that guide how technology is adopted and applied, while leadership competency provides the vision, skills and accountability ‘to align AI initiatives’ with ‘ethical and risk-aware practices’ (Schein, 2017; Bass and Avolio, 1994).

A culture that prioritises transparency, inclusivity and accountability creates the conditions for responsible AI, ensuring that ethical principles such as fairness, human oversight and privacy are embedded into decision-making (Floridi et al., 2018). In contrast, weak cultural foundations can exacerbate risks, including algorithmic bias, opaque decision-making and reputational harm. Leadership plays a pivotal role in shaping this culture by setting expectations, fostering cross-disciplinary collaboration, and embedding responsible AI standards into governance structures (Wilson and Daugherty, 2018).

1.5 Leadership Education and the Readiness Gap

The integration of AI into organisational decision making has elevated ethical literacy and governance fluency to the status of core leadership competencies. The findings from this research revealed a striking paradox: while more than ninety per cent of leaders recognise AI literacy as essential, fewer than one in five feel confident navigating its governance and ethical dimensions. This disparity between awareness and preparedness represents a critical readiness gap with significant implications for organisational performance and public trust. The question is not merely 'what capabilities

does the technology possess?' but rather 'are we prepared to assume the responsibility that accompanies it?'.

Leadership in the AI era extends beyond technical understanding. It encompasses the capacity to pose the right questions, to anticipate ethical risks, and to cultivate cultures of accountability. The results from this research indicate that many leaders lack the structured guidance and cross disciplinary training needed to operationalise responsible AI principles. The consequences of this deficiency are visible in inconsistent policy implementation, fragmented oversight, and uncertainty regarding regulatory compliance. Furthermore, the evidence points to a pressing need for structured, actionable approaches to responsible AI.

Leadership education, therefore, must evolve. It should integrate AI literacy, ethical reasoning, and governance frameworks into executive development programmes, equipping decision makers with the knowledge and confidence to guide AI adoption responsibly. Embedding these dimensions into training, policy, and leadership structures will be vital to advancing AI maturity and mitigating risk. Organisational success will increasingly depend on leaders who can align business objectives with societal values, balancing innovation with duty of care.

Overall, the research findings call for urgent investment in AI-focused leadership training and cross-disciplinary approaches that integrate business strategy with governance and ethics. Organisational success will increasingly depend not only on technical implementation but on leadership able to navigate the ethical, regulatory, and practical dimensions of AI with integrity and accountability.

1.6 The TRAIL Framework

This research responds to the identified readiness gap through the development of the TRAIL (Trustworthy Responsible AI Leadership) framework. TRAIL provides a comprehensive and practical model designed to support senior leaders in developing, deploying, and governing AI responsibly. Embedding responsibility at the core of AI leadership enables innovation that is both credible and sustainable, laying the foundation for frameworks such as the TRAIL model that seek to prepare leaders for the demands of the AI-driven future. TRAIL integrates insights from governance, regulation, ethics, and leadership development literature with empirical findings from this study, which surveyed over 300 international professionals across various sectors. The framework is designed to address the critical gaps identified in leadership competency, AI literacy, and organisational governance.

The framework consists of four interrelated pillars: Strategic and Transformational Leadership, which emphasizes the cultivation of vision, adaptability, and the capacity to integrate AI into strategic decision-making while ensuring human oversight. Governance and Regulatory Fluency, which ensures that leaders comprehend and adhere to evolving national and international AI standards. Ethical and Risk Literacy, which incorporates moral reasoning, bias awareness, and risk mitigation into leadership practices. Ethical Data Management, which advocates for integrity, transparency, and accountability in the stewardship of data throughout the AI lifecycle.

Together, these pillars establish the foundation for responsible AI leadership. They position ethical competence and governance fluency as critical dimensions of modern executive capability, emphasising that responsible AI is not merely a technological or compliance challenge but a question of leadership responsibility.

1.7 Chapter Summary

AI stands at a crossroads between extraordinary promise and profound risk. Its capacity to enhance human welfare, drive sustainable economic growth, and solve complex global challenges is unparalleled. Yet without effective governance, ethical safeguards, and informed leadership, it could exacerbate inequality, erode trust, and destabilise social institutions.

The next decade will be decisive, determining whether AI's trajectory is defined by its capacity to drive inclusive growth and human flourishing or by the failure to manage its risks responsibly. Achieving the former requires collaboration between policymakers, industry, academia, and civil society to establish governance systems that are adaptive, transparent, and human centred. It also demands leaders capable of navigating complexity with wisdom, foresight, and ethical integrity. The TRAIL framework aims to address the fundamental question of preparedness by equipping leaders with the necessary competencies, values, and structures to ensure that AI serves society in a manner that is trustworthy, responsible, and aligned with the collective good.

The future will not be decided by algorithms and technological capability, but by the choices and competence of the leaders who guide them. because ultimately, responsible AI is not a question of technological possibility, but of leadership responsibility.

The preceding discussion has outlined the dual nature of AI as both a transformative opportunity and a source of complex ethical, regulatory, and governance challenges. To situate this study within the wider scholarly context, the following Literature Review examines the theoretical and empirical foundations that inform responsible AI leadership.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

The integration of AI into commercial and societal frameworks has emerged as a transformative force, reshaping industries and redefining human interaction with technology. This pivotal shift in the modern economy could, according to consultancy PwC, contribute up to $15.7 trillion to the global economy by 2030 (PwC, 2027), representing the largest commercial opportunity in today's evolving landscape. Scharre (2023) compares this to a new industrial revolution, reflecting the impact of previous technological advances.

The rapid proliferation of AI technologies offers numerous opportunities, including substantial enhancement of operational efficiency through automation, leading to increased productivity and cost reduction (Schönberger, 2023). AI can optimise logistics and production processes, enhancing supply chain management (Guo, 2023), while advanced data analytics facilitate enhanced decision-making capabilities and personalisation by harnessing vast amounts of data for accurate predictions (Duan et al., 2019). Furthermore, AI has the potential to revolutionise customer experience by enabling personalised services that enhance satisfaction and loyalty (Ameen et al., 2021), and it can drive innovation in service development, providing a competitive advantage (Huang & Rust, 2018).

While the transformative potential of AI is widely recognised, its accelerating adoption introduces a complex range of risks to commercial standards, organisational integrity, and societal wellbeing. These risks encompass ethical and regulatory challenges, as well as the potential erosion of human agency. Balancing innovation with effective ethical oversight and regulatory compliance therefore underscores the

interdependence of technical, organisational, and societal considerations in the deployment of AI.

This literature review aims to provide a comprehensive analysis of the existing discourse concerning Responsible AI, regulation, governance, and the role of leadership in mitigating associated risks. By integrating diverse perspectives, from technical risks to organizational change theory, the review seeks to establish a coherent framework for understanding the essential capabilities required of leadership in the AI era.

2.2 AI as Organisational Value: Opportunities and Strategic Drivers

The adoption, implementation, and utilization of AI systems in commercial organizations present numerous opportunities that can significantly enhance operational efficiency, customer engagement, and decision-making processes.

One of the most significant opportunities presented by AI is the enhancement of operational efficiency through automation. Schönberger highlights that AI integration can lead to increased productivity, cost reduction, and improved decision-making capabilities (Schönberger, 2023). By automating routine tasks, organizations can allocate human resources to more strategic activities, thereby fostering innovation and enhancing overall competitiveness. This perspective is corroborated by Guo, who discusses how AI can optimize supply chain management, improve logistics and production processes while ensuring product quality (Guo, 2023). The capacity to streamline operations through AI can reduce costs while enhancing service delivery, outcomes that are increasingly critical in highly competitive environments.

AI offers significant advancements in customer experience. Ameen et al. emphasize that AI has the potential to revolutionize customer interactions by enabling personalized services and experiences (Ameen et al., 2021). Through the utilization of data analytics and machine learning, organizations can tailor their offerings to meet

individual customer preferences, thereby enhancing satisfaction and loyalty. This personalization is particularly relevant in sectors such as retail and finance, where customer expectations are continually evolving. Flavián et al. further support this notion, noting that customers' technological readiness and awareness play a crucial role in the adoption of AI in service sectors, which can lead to improved customer engagement (Flavián et al., 2021).

Moreover, AI systems facilitate enhanced decision-making capabilities through advanced data analytics. The ability to analyse complex datasets expeditiously allows organizations to respond to market changes and customer needs more effectively. Duan et al. highlight that AI enables organizations to harness vast amounts of data to make informed predictions and decisions, significantly reducing the costs associated with traditional data analysis methods (Duan et al., 2019). This capability is particularly beneficial in sectors such as healthcare, where timely and accurate decision-making can have profound implications for patient outcomes.

AI also presents opportunities for innovation in product and service development. Huang and Rust discuss how AI can drive innovation in service delivery by enabling organizations to create new service models that were previously unattainable (Huang & Rust, 2018). This innovation is not limited to existing products but extends to the development of entirely new offerings that leverage AI capabilities, thereby providing organizations with a competitive advantage.

Furthermore, the ethical deployment of AI can enhance corporate reputation and stakeholder trust. Jobin and Ienca emphasize the importance of developing ethical guidelines for AI usage, which can assist organizations in navigating the complexities of AI implementation while maintaining public confidence (Jobin & Ienca, 2019). By

committing to ethical AI practices, organizations can differentiate themselves in the marketplace and foster stronger relationships with customers and stakeholders.

2.3 Risk, Sanctions and Accountability in AI Systems

The adoption, implementation, and utilization of AI systems in commercial organizations present a range of risks that can significantly impact operational efficiency, ethical standards, and overall business integrity.

One of the primary risks associated with AI adoption is the potential for misinformation and misunderstanding regarding AI capabilities. Ganapathi and Duggal highlight that misconceptions about AI, particularly the belief that AI could entirely supplant human roles, can lead to anxiety and resistance among employees Ganapathi & Duggal (2023). This outlook not only affects employee morale but can also impede the successful implementation of AI systems, as staff may be less inclined to engage with technologies they perceive as threats to their job security.

The inherent complexity of AI systems presents significant challenges for risk assessment and management. Nagbøl et al. (2021) emphasise the importance of robust processes for identifying, analysing, and prioritising AI related risks prior to deployment. In the absence of systematic risk assessment, organisations may fail to detect critical vulnerabilities, increasing the likelihood of operational failures or ethical breaches.

In the financial sector, the integration of AI into risk management and fraud detection processes is becoming increasingly prevalent. Majumder notes that while AI can enhance these capabilities, it also introduces new risks, including algorithmic bias and the potential for excessive reliance on automated systems (Majumder, 2024). These risks require a balanced approach that leverages AI for efficiency while maintaining appropriate human oversight to support ethical decision making and accountability.

Privacy concerns also represent a significant risk associated with AI implementation, particularly in industries that handle sensitive data. Price and Cohen discuss the ‘ethical implications of utilizing AI in conjunction with big data’, highlighting the potential for privacy violations and the need for stringent data protection measures (Price & Cohen, 2019). Organizations must navigate these challenges to maintain trust with their stakeholders and comply with regulatory requirements.

Concerns regarding fairness and bias present substantial ethical and practical challenges. Bias in AI systems frequently originates from training data that mirror historical or societal inequalities, resulting in discriminatory outcomes (Alzghoul and Alrawahna, 2025). For instance, facial recognition systems have exhibited disproportionate error rates for marginalized groups (Shukla, 2025). Similarly, automated hiring tools have been found to perpetuate gender and racial biases (Sackett and Lievens, 2025).

Fairness in AI is a multifaceted concept, encompassing various definitions such as equal opportunity, demographic parity, and individual fairness, which complicates its operationalization (Vidal and Russo, 2025). In the absence of clear standards, developers frequently resort to ad hoc mitigation techniques that may prove ineffective or potentially introduce new forms of unfairness (Panchal and Panchal, 2025). Explainability represents a related concern, as opaque models may conceal biased decision making, thereby undermining trust and accountability (Kate, 2025).

The regulatory landscape is beginning to adapt. The EU AI Act seeks to categorize systems based on risk and to enforce transparency and fairness in high-risk applications (Ayeola and Okunola, 2025). However, critics contend that implementation is lagging behind technological advancement, particularly in the Global South (Shahid et

al., 2025). Consequently, multidisciplinary approaches that integrate technical, legal, and sociological perspectives are crucial for mitigating bias and promoting equitable AI.

The absence of standardized documentation and evaluation metrics for AI systems can exacerbate risks. Mitchell et al. point out that the lack of standardized procedures for reporting the performance characteristics of AI models can lead to misinformed decisions regarding their deployment (Mitchell et al., 2019). This deficiency in transparency can undermine the reliability of AI systems and erode stakeholder confidence.

The operationalization of AI in commercial settings also faces challenges related to the integration of AI solutions into existing workflows. Homeyer et al. note that while many AI applications show promise, the transition from prototype to product is fraught with difficulties, including technical, business, and regulatory hurdles (Homeyer et al., 2021). These challenges can delay the benefits of AI adoption and lead to inefficient resource allocation if not managed effectively.

In conclusion, the risks associated with the adoption, implementation, and utilization of AI systems in commercial organizations are multifaceted and necessitate careful consideration. Organizations must address misconceptions regarding AI, implement robust risk assessment procedures, ensure privacy protections, and establish standardized evaluation metrics to mitigate these risks. Through these measures, they can enhance the ethical deployment of AI technologies and foster a culture of responsible innovation.

The Facebook and Cambridge Analytica scandal serves as a pivotal case study on the ethical implications of AI and data privacy. This 2018 incident, involving the unauthorized extraction of personal data from millions of Facebook users to influence political campaigns, raised profound ethical concerns regarding data privacy and the manipulation of information for targeted advertising. Yazdani (2023) emphasizes the

critical importance of ethical considerations in AI deployment, particularly in the utilization of consumer data. This scandal exemplifies the unethical potential of AI, resulting in diminished consumer trust and highlighting the necessity of robust ethical frameworks.

A primary ethical issue arising from this case is the lack of transparency and informed consent in data practices. Mancosu and Vegetti (2020) note that Facebook's response, involving stricter access to its Application Programming Interface (API), underscores the importance of user awareness concerning data collection and usage. The case emphasizes the risks when profit motives supersede ethical imperatives, leading to unethical corporate behaviour.

The scandal also precipitated a broader discourse on the ethical obligations of technology firms. Akbar et al. (2023) posit that organizations must adhere to principles of transparency, fairness, and accountability when addressing AI's ethical challenges. This perspective aligns with Jobin and Ienca's (2019) observation that the proliferation of AI ethics guidelines signifies a recognition of the necessity for clear ethical governance in AI practices. The Cambridge Analytica incident provides a salient cautionary example of the consequences that arise when ethical principles are neglected in the deployment of data driven and AI enabled systems.

Furthermore, the incident has precipitated discussion on digital ethics, particularly within the domain of social media and data utilization. Kazim and Koshiyama (2020) propose a comprehensive conceptualization of AI ethics that encompasses digital ethics, acknowledging the broader societal implications of AI technologies. They contend that effective ethical guidelines necessitate interdisciplinary collaboration among stakeholders, including policymakers, researchers, and technology companies.

The regulatory landscape, exemplified by frameworks such as the General Data Protection Regulation (GDPR), has implemented stringent sanctions for data protection violations. Presthus and Sønslien (2021) highlight cases such as Google's €50 million fine in 2019 by the French data protection authority CNIL for transparency failures, illustrating the financial ramifications of non-compliance.

The California Consumer Privacy Act (CCPA) has implemented comparable penalties for data misuse, as evidenced by the 2019 Equifax settlement of $700 million (FTC, 2019b) resulting from a substantial data breach affecting 147 million consumers. This incident not only revealed vulnerabilities in data protection practices but also underscored the ethical obligation of corporations to safeguard consumer information.

Taken with Facebook's $5 billion fine by the U.S. Federal Trade Commission for failing to safeguard user data and for deceptive data practices (FTC, 2019a), these instances underscore the enforcement of ethical standards and the monetary penalties associated with unethical data handling practices. They also highlight systemic weaknesses in corporate data governance and reinforced concerns about the commodification of human experience raised by Mejias and Couldry (2019) and further illustrate Hasselbalch's (2022) warning about governance by opaque digital infrastructures and Marcus's (2022) critique of the ethical risks tied to data-centric AI models.

Regulatory scrutiny of AI misuse has intensified. In 2020, the U.S. Federal Trade Commission (FTC) issued a warning to organizations regarding discriminatory AI practices, emphasizing potential legal ramifications for unethical AI applications (FTC, 2020). This increased focus underscores the ethical imperative to mitigate AI-driven biases and safeguard vulnerable populations.

The Cambridge Analytica case, alongside other instances AI-driven surveillance, exemplifies the ethical risks associated with the misuse of data analytics and underscores the necessity for more stringent guidelines governing AI technology, particularly in data-sensitive environments. The ramifications of ethical transgressions extend beyond monetary penalties to reputational damage and erosion of consumer trust, as explained by Teixeira et al. (2019). Adherence to ethical standards not only mitigates legal risk but also strengthens organisational reputation and consumer trust, underscoring that ethical conduct in AI and data governance constitutes both a regulatory requirement and a source of strategic advantage.

In conclusion, the spectrum of sanctions and regulatory measures for data protection infractions reflects an increasing recognition of the ethical imperatives organizations face in managing consumer data. With regulations such as GDPR and CCPA shaping the compliance landscape, entities must proactively incorporate ethical practices into their data management and AI strategies to engender trust and uphold responsible innovation. The adverse consequences of breaches in data related regulation underscore the need for robust ethical frameworks and accountability mechanisms in the commercial deployment of AI.

Enterprise Risk Management (ERM) has emerged as a critical component of corporate governance, particularly in commercial contexts and in relation to AI deployment. Demidenko and McNutt (2010) expand on the ethical dimensions of ERM, emphasizing that effective governance extends beyond compliance to fostering a culture of integrity. They posit that ethical governance can enhance decision-making processes and promote sustainable business practices (Demidenko & McNutt, 2010). Arora & Sharma (2016) demonstrate that firms with weaker governance structures experience greater agency problems, which can negatively impact their profitability. This correlation

underscores the significance of governance mechanisms in aligning the interests of various stakeholders and promoting organizational performance.

The effectiveness of corporate governance practices is influenced by the regulatory environment in which organizations operate. Xu et al. (2016) further explain this by examining how internal and external governance structures interact to influence corporate performance, suggesting that a well-defined governance framework can enhance organizational resilience in volatile markets.

A primary concern in AI governance is the accountability of AI systems, particularly when they produce unintended consequences. In the marketing sector, Naz highlights the dual nature of AI's impact, wherein it can enhance consumer insights but also perpetuate biases and privacy violations (Naz, 2024). This duality underscores the necessity for ethical frameworks that guide AI applications to prevent harm and ensure equitable outcomes.

Mökander et al. emphasize the necessity of ethics-based auditing for automated decision-making systems, highlighting the need to develop guidelines to ensure ethical AI deployment (Mökander et al., 2021). These guidelines underscore the importance of transparency, fairness, and accountability, which are essential for fostering public trust in AI technologies. Similarly, Huriye discusses the complexities surrounding accountability in AI, noting that the opaque nature of AI systems complicates the identification of responsible parties for decisions made by these systems (Huriye, 2023). This complexity necessitates clear accountability structures to mitigate risks associated with AI deployment.

Due to the disparate regulatory frameworks across countries, the establishment of a global governance structure for AI is increasingly recognized as essential, addressing a significant gap in current approaches. It presently remains elusive and will necessitate

harmonizing efforts to create a cohesive global governance framework for AI, requiring collaboration, shared principles, and adaptive regulatory approaches.

While the existing literature effectively delineates the interconnections among governance, ethics, and risk management, it frequently presupposes a linear correlation between compliance structures and ethical performance. This perspective risks neglecting the more profound cultural and behavioural dynamics that influence the enactment of ethical principles within organizations (Kaptein, 2022). In the realm of AI, risk management frameworks may inadvertently emphasize procedural compliance, such as ethics audits or bias assessments, over the cultivation of moral responsibility and critical reflection among decision-makers (Binns, 2018). Furthermore, although global harmonization in AI governance is desirable, the assumption that shared regulatory principles will inherently ensure ethical outcomes fails to consider regional disparities in enforcement capacity, political will, and societal values (Jobin, Ienca, and Vayena, 2019). Consequently, effective AI risk governance necessitates not only harmonized frameworks but also the integration of ethical literacy and accountability throughout organizational culture and leadership practices.

The integration of AI into business practices has raised significant ethical concerns, particularly in areas of accountability and governance. As AI becomes a key component in organizational decision-making, the ethical implications of its application have gained increased attention. The literature reveals a consensus on the need for strong ethical frameworks to guide the responsible deployment of AI in business.

Oladele (2024) highlights the dual nature of AI systems in business, which present both opportunities and risks. This duality necessitates a comprehensive understanding of ethical considerations to ensure the responsible use of AI in critical business processes. The framework proposed by Oladele aligns with Häußermann and

Lütge (2021), who stress the importance of embedding business ethics into AI governance to achieve ethical objectives. They advocate for a pluralistic approach that acknowledges the interplay between AI technologies and existing business practices.

Furthermore, Hunkenschroer and Luetge (2022) identify an accountability gap when AI systems operate without human oversight, particularly in recruitment and selection. This raises critical questions regarding organizational responsibility for AI-driven decisions, a challenge also examined by Holzhausen (2024), who investigates the complexities of legal accountability for AI outcomes. The intersection of legal and ethical considerations is essential, as organizations must navigate the implications of AI deployment within regulatory constraints.

Ethical considerations extend to consumer trust and corporate responsibility. Olatoye (2024) emphasizes that ethical AI practices are vital in establishing consumer trust, particularly as organizations seek a competitive advantage through AI. This perspective is corroborated by Al (2023), who underscores the necessity for transparency in AI-based marketing to maintain consumer confidence. Clear communication regarding data usage and AI decision-making processes is crucial to integrating ethical considerations into organizational strategies.

Collectively, the literature indicates that ethical AI practices can enhance customer experience and support sustainable organisational growth. Tula (2024) emphasizes the significance of ethical standards in AI deployment strategies, advocating for continuous innovation aligned with ethical principles. This perspective aligns with Daza and Ilozumba (2022), who acknowledge the substantial benefits of AI while cautioning that these advantages are accompanied by significant ethical risks that necessitate effective management.

While the existing literature acknowledges the necessity of ethical frameworks in AI-enabled business environments, much of it remains normative rather than empirically substantiated. Claims that ethical AI enhances trust and performance frequently lack evidence of causal mechanisms or measurable outcomes (Martin, 2022). Furthermore, the focus on voluntary ethical adoption may exaggerate corporate willingness to self-regulate in the absence of external accountability, especially in competitive markets where short-term gains can supersede moral commitments (Treviño, den Nieuwenboer, and Kish-Gephart, 2014). Although framing ethics as a strategic asset can facilitate managerial buy-in, it risks instrumentalizing moral principles for reputational advantage rather than embedding them as intrinsic organizational values. Therefore, advancing business ethics in the AI era necessitates a shift from aspirational statements to demonstrable practices, independent audits, and leadership accountability mechanisms that operationalize ethical intent.

In conclusion, the literature on ethical AI in business underscores the necessity for comprehensive ethical frameworks that address accountability, transparency, and consumer trust. As AI technologies continue to evolve, businesses must prioritize ethical considerations to responsibly navigate the complexities of AI integration.

The opaque nature of AI systems complicates the identification of responsible parties for automated decisions, thereby creating ambiguities in accountability. Leaders must employ frameworks that foster adaptive governance and sense-making, particularly when confronting the uncertainty and non-linear consequences characteristic of complex and chaotic environments.

In this context, ethical leadership encompasses the ability to cultivate dialogic accountability, whereby leaders facilitate dialogue, self-reflection, and collective meaning-making, particularly pertinent when human values must remain central to

technological advancements. The strategic leader uses this proficiency to ensure that AI governance frameworks, such as harmonized regulatory and voluntary standards, are consistently applied across all organizational levels, thereby guaranteeing that ethical innovation is both accountable and resilient. In the absence of this focused leadership capability, ethical frameworks risk functioning merely as legitimizing tools rather than as transformative instruments for responsible innovation.

2.4 Data Governance and the Data Lifecycle

The commercial application of AI is fundamentally dependent on the data lifecycle, which encompasses processes ranging from collection to deletion. Within private sector enterprises, this lifecycle is integral to the development, deployment, and monitoring of AI systems. Consequently, effective data governance and lifecycle management have become crucial for responsible AI innovation and business performance.

Global data creation is projected to reach ‘181 zettabytes by 2025’, driven largely by commercial use of AI, which ‘increasingly relies on real-time data streams and cloud-based analytics’ (IDC, 2022; De Silva and Alahakoon, 2022). Data collection in private firms typically involves aggregating information from diverse sources, including sensors, user interactions, and enterprise databases (De Silva and Alahakoon, 2022). This process has heightened scrutiny of the ethical and legal implications regarding consent and data ownership (Chen et al., 2021).

Once acquired, data storage necessitates robust infrastructure to ensure scalability and security. Cloud storage solutions, frequently employed by private firms, offer elasticity and cost efficiency but introduce new challenges in regulatory compliance and data sovereignty (Zeydan et al., 2024). Data lakes and warehouses are commonly utilized,

with access controls being critical to maintaining privacy standards (Gudivada et al., 2017).

Data cleansing is essential for ensuring the accuracy and utility of AI models. Erroneous, duplicated, or incomplete data can distort analytical outcomes, rendering data cleansing a critical preparatory process rather than a trivial technical task (Yerra, 2023). Advanced tools employ AI to automate cleaning tasks, thereby improving reliability and reducing manual labour (Stöger et al., 2021). However, improper cleansing can inadvertently erase minority patterns, raising concerns over algorithmic bias.

Subsequent analysis and processing involve structuring data for training and evaluating AI systems. This process includes data labelling, transformation, and selection, with a strong emphasis on traceability and auditability (Shahriar et al., 2023). The increasing focus on explainable AI (XAI) necessitates not only accurate outputs but also interpretability of model behaviour, influencing how data is structured and interpreted.

The visualization and utilization of data are crucial for operationalizing insights. Dashboards and real-time monitoring tools are prevalent in sectors such as logistics and retail, enabling AI to support decision-making (Ahmed et al., 2023). Businesses leverage these outputs to optimize processes, personalize services, and forecast trends.

Data deletion and retention policies conclude the lifecycle. Retention strategies must balance business value against compliance, particularly with regulations such as GDPR, which mandate the 'right to be forgotten' (Chundru and Mudunuri, 2025). Secure deletion, including data overwriting or cryptographic erasure, is increasingly standard to mitigate risk (Chukwurah et al., 2024).

Hasselbalch discusses the concept of a ‘big data society,’ characterized by data technologies' pervasive influence. In such a society, data fundamentally alters democratic

principles, individual rights, and institutional frameworks. Hasselbalch (2022) critiques big data's impact on governance, public values, and societal expectations, emphasizing that this transformation extends beyond technology to norms and politics. In a big data society, data is used to observe and predict reality. This predictive capability reshapes interactions between individuals and institutions, undermining fairness, transparency, and autonomy. Hasselbalch (2022) cautions that decisions traditionally made through democratic processes are increasingly determined by opaque algorithms. A central theme is the transition from governance by law to governance by data. Traditional legal safeguards cannot adequately regulate algorithmic power and surveillance. Hasselbalch advocates for digital ethics grounded in democratic values rather than technical standards or corporate policies and calls for a novel approach to data governance that empowers individuals, reinstates public oversight, and integrates ethical considerations into digital policymaking, proposing ‘normative anchoring,’ which embeds societal values into technology from inception. However, the notion of ‘governance by data’ could be critiqued as potentially deterministic. Are democratic institutions completely disempowered, or are there emerging checks and balances (e.g. AI audits, algorithmic impact assessments)?

Marcus (2022) offers a prominent critique of prevailing AI approaches, particularly deep learning systems that rely on large scale data sets. He argues that data alone is insufficient for the development of genuinely intelligent and trustworthy systems, noting persistent deficiencies in robustness, transparency, and reasoning capability. Marcus cautions that excessive dependence on statistical correlations, in the absence of causal understanding, produces opaque models that are prone to failure in unfamiliar contexts, thereby undermining trust and safety in high-risk applications. He advocates for hybrid approaches that combine symbolic reasoning with data driven methods to support

verifiable, interpretable, and human aligned AI. This perspective foregrounds the importance of ethical oversight, data quality, and systematic evaluation of potential harms within AI governance and design.

While Marcus's critique of deep learning is valid, there is an increasing body of research investigating the integration of symbolic and statistical AI. Are his criticisms perhaps overly concentrated on historical limitations? Furthermore, could it be argued that in certain applications, such as image recognition, deep learning has demonstrated remarkable generalization with minimal symbolic intervention?

Mejias and Couldry conceptualise data colonialism as a contemporary system of dominance in which human life is appropriated and commodified through large scale data extraction. They argue that digital infrastructures routinely capture everyday activities and convert them into economic value, often without meaningful consent, thereby extending historical patterns of expropriation within the digital economy (Couldry and Mejias, 2019). Enabled by platform-controlled infrastructures, this process reinforces global inequalities, with data extraction frequently occurring in the Global South while economic benefits accrue largely to corporations in the Global North (Mejias and Couldry, 2024).

Although the use of colonial terminology has been critiqued for potentially overstating equivalence with coercive land appropriation in regulated democratic contexts, the framework remains valuable in exposing the structural and geopolitical dimensions of data exploitation. Its central contribution lies in demonstrating how contemporary data relations reproduce asymmetrical power dynamics and inform debates on digital governance and ethical responsibility.

Collectively, these critiques raise substantive questions about the capacity of private sector organisations to reconcile innovation with ethical accountability. Where

data driven AI models risk reinforcing inequality or undermining democratic values, the extent to which corporate data governance can mitigate such harms remains contested. The private sector's approach to data in AI is therefore characterised by a persistent tension between innovation and regulation. Effective data lifecycle management is central not only to model performance but also to the maintenance of trust, legitimacy, and ethical responsibility within commercial contexts.

2.5 Regulatory Frameworks and International Developments

The regulation of AI has become a critical cross sectoral concern, driven by rapid technological advancement and the ethical, legal, and societal implications of its deployment.

Predominant themes within AI regulation literature encompass governance strategies, ethical frameworks, international cooperation, and sector-specific concerns. AI regulatory approaches are primarily dichotomized between a cautionary, risk-averse stance that aims to mitigate AI-associated risks, and a more innovation-oriented approach that emphasizes the potential of AI to contribute to economic growth and efficiency.

One of the primary challenges in regulating AI is the absence of standardized frameworks that address the ethical implications of AI technologies. Shreve et al. (2022) emphasize the necessity for robust regulatory and legal frameworks to mitigate issues such as 'data standardization, inherent biases in training datasets, and the integration of AI into existing workflows'. These concerns are corroborated by Hamon (2024), who discusses the complexities of ensuring cybersecurity compliance within AI systems, particularly in light of the European Union's AI Act. The interplay between cybersecurity and regulatory compliance is crucial, as AI systems are increasingly susceptible to cyberattacks, necessitating a proactive approach to regulation that encompasses both ethical and security considerations.

Additionally, the ethical implications of AI extend beyond technical compliance to encompass broader societal impacts. The General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) are ‘pivotal in establishing requirements for data protection and accountability in AI applications’ (Reddy et al, 2020). These regulations aim to uphold individuals' privacy rights while promoting ethical conduct. The need for transparency and accountability in AI systems is further supported by the findings of Naik et al. (Naik et al., 2022), who highlight the absence of well-defined regulations, underscoring the importance of algorithmic transparency and the protection of vulnerable populations.

The financial sector also faces unique regulatory challenges as AI technologies become more integrated into financial forecasting and reporting. Oyeniyi (2024) notes that while AI enhances reporting accuracy and efficiency, it also raises ethical concerns related to biases and regulatory compliance. The necessity for comprehensive regulatory frameworks that address these challenges is paramount, as highlighted by Ajiga (2024), who calls for a systematic review of AI's effectiveness in financial analysis, emphasizing the importance of regulatory oversight in fostering trust and accountability.

In the context of AI governance, the establishment of clear guidelines and frameworks is essential to balance innovation with ethical considerations. Gasser and Almeida (2017) propose a layered model for AI governance that incorporates ethical principles alongside existing legal frameworks. This model aims to create a structured approach to AI regulation, ensuring that stakeholders are held accountable for the implications of their technologies. Additionally, the work of Minkkinen et al. (2022) espouses the role of auditing in bridging the gap between AI ethics and responsible usage, advocating for mechanisms that extend beyond mere legal compliance.

The EU AI Act (European Union, 2024) represents a pioneering regulatory framework for AI in the European Union, constituting the world's first attempt at a comprehensive legislative approach to AI regulation. The legislation adopts a risk-based approach that imposes more stringent rules on higher-risk systems. However, Caroli (2024) posits that this approach may present a significant potential limitation, as the Act's focus on the ‘intended purpose’ of narrow, high-risk AI systems could be problematic. Providers may potentially circumvent responsibility by asserting that their AI's purpose does not align with regulated uses, which could potentially diminish the Act's efficacy.

The Act acknowledges that effective enforcement necessitates specialized expertise and recommends the establishment of supervisory authorities with proficiency in fundamental rights and technical AI matters. In contrast to GDPR, which encountered enforcement challenges due to centralized data processing, the AI Act will utilize national market surveillance authorities, potentially facilitating more efficient enforcement distribution. Nevertheless, concerns persist regarding these authorities' capacity to conduct AI-specific technical assessments (Caroli, 2024).

A significant observation is that AI, which influences rights and societal values, cannot be regulated exclusively through technical requirements. The Act endeavours to address this by introducing AI-specific rights, such as the right to an explanation of AI-driven decisions, and by mandating fundamental rights impact assessments for high-risk AI systems.

Bashir and Frank (2024) assert that the EU is implementing excessive regulations, and that only ‘high-risk’ systems will be subject to stringent regulatory measures, thereby potentially impeding its competitiveness. They posit that, by its nature, any regulation imposed on an evolving technology such as AI will inevitably fail to encompass all possible use cases and impose unintended restrictions (Bashir and Frank, 2024).

Caroli (2024) argues that criticisms of the AI Act underestimate its adaptability and that direct comparisons with the GDPR are misplaced. Unlike the GDPR, enforcement under the AI Act is exercised by market surveillance authorities at the point of infringement rather than being tied to the location of the provider, thereby reducing the risk of enforcement concentration and regulatory overload.

While Smuha (2021) and Justo-Hanani (2022) underscore the European Union's methodical and structured approach to AI regulation, this perspective necessitates critical examination. The EU's endeavour to establish ‘trustworthy AI’ through legal and ethical frameworks illustrates a normative ambition; however, it has faced criticism for potentially hindering innovation and imposing compliance burdens that disproportionately impact smaller enterprises (Ebers, 2022). Furthermore, the regulatory focus on ex ante conformity assessments and high-risk categorizations may risk stifling technological advancement, as innovation frequently progresses more rapidly than legislative adaptation (Veale and Borgesius, 2021). Although the incremental strategy highlighted by Justo-Hanani (2022) reflects prudence and adaptability, it may also engender regulatory uncertainty, particularly for global firms navigating fragmented national interpretations of EU law. Consequently, while the EU's approach serves as a global benchmark for responsible AI governance, its effectiveness ultimately hinges on achieving a balance between precaution and flexibility, ensuring that regulation both upholds ethical values and sustains technological competitiveness.

Critics (Caroli, 2024; Bashir and Frank, 2024) contend that the structure of the AI Act may be excessively rigid, given the rapid and dynamic evolution of AI. While the Act incorporates some degree of flexibility through standards and guidelines, concerns persist regarding its adaptability. As with most legislation, the feasibility and effectiveness of the Act can only be assessed retrospectively, at which point key lessons

must be drawn and implemented. Although the Act provides a foundational framework, subsequent amendments may be necessary to ensure its continued relevance and efficacy as AI technology advances.

The European Union has enacted the Digital Markets Act (DMA) to promote contestable and fair competition within digital markets, with particular obligations imposed on large technology firms designated as ‘gatekeepers’ (European Union, 2022a). Administered by The European Commission, the DMA aims to foster a more competitive digital landscape by implementing regulations that prevent dominant corporations from exploiting their market influence and impeding the expansion of smaller competitors (European Union, 2022a).

Under the DMA, gatekeepers are obligated to share the data they collect and ensure their primary platform services are interoperable with competing offerings. These entities are also prohibited from utilizing data gathered from third-party business users to directly compete with those enterprises.

Non-compliant firms may incur penalties of up to 10% of their worldwide annual revenue for initial infractions, and up to 20% for subsequent violations (European Union, 2022a). When deemed necessary, the Commission has the authority to enforce behavioural or structural measures to restore fair competition, such as mandating companies to divest certain business segments.

The European Union’s Digital Services Act (DSA) establishes a regulatory framework for online platforms and digital services aimed at enhancing safety, transparency, and accountability within the digital environment. Complementing the Digital Markets Act (DMA), the DSA focuses primarily on content moderation, platform transparency, and the protection of user rights rather than competition-related concerns.

Its core objectives include consumer protection, strengthened accountability, increased transparency, and the safeguarding of fundamental rights (European Union, 2022b).

The DSA's scope extends to all digital services that facilitate connections between consumers and goods, services, or content, with additional obligations imposed on very large online platforms (VLOPs) and very large online search engines (VLOSEs). Defined as platforms with a user base exceeding 45 million active EU users, VLOPs and VLOSEs are subject to more stringent requirements due to their substantial societal impact (European Union, 2022b).

McIntyre (2023) argues that the European Commission's centralized enforcement of the DMA risks imposing uniform standards across diverse digital markets, potentially inhibiting rather than fostering competition. Furthermore, Nurhayati (2023) suggests that the DMA's stringent requirements may inadvertently disadvantage smaller firms that lack the resources to comply.

Fletcher (2023) raises concerns regarding the feasibility of the DSA's transparency mandates in algorithmic decision-making, particularly for smaller platforms with limited technical capabilities.

Critics, including Colomo (2021), contend that both the DMA and DSA may encounter challenges in enforcement against large multinational corporations capable of contesting regulatory actions. Streel et al. (2023) advocate for a balanced approach, cautioning that regulatory measures should take into account digital market complexities to avoid disproportionate impacts on smaller entities.

While the DMA and DSA represent significant advancements in digital regulation, concerns regarding innovation, content moderation, and enforcement underscore the necessity for ongoing refinement to mitigate unintended consequences.

Gromova et al. (2021) contrasts regulatory developments in the BRICS countries with those of the EU, underscoring the diverse approaches based on socio-political and economic frameworks (Gromova et al., 2021). Similarly, Weaver (2018) provides a comparative analysis of AI regulations in the United States, highlighting a predominantly self-regulatory stance with selective government intervention, reflecting the country's emphasis on technological autonomy (Weaver, 2018).

Bradford (2023) delineates three principal ‘digital empires’: the United States, China, and the European Union. Each of these entities, which reflect diverse approaches to the interrelationship between markets, state control, and individual rights, influences the global digital landscape through distinct regulatory frameworks that reflect their respective ideological commitments.

The United States endorses a market-driven model that prioritizes innovation, freedom of expression, and minimal governmental intervention. According to Bradford (2024a), the U.S.'s global leadership in technology is attributed to systemic features such as robust capital markets, a supportive bankruptcy regime, and policies that attract global talent, rather than solely relying on a laissez-faire regulatory approach. However, critics underscore its inadequate protection of user privacy and the unrestrained influence of technology conglomerates, raising concerns regarding their impact on public discourse and societal dynamics. In contrast, China's state-driven model prioritizes governmental control over digital platforms and content, aligning technological advancements with national objectives. While this approach has facilitated rapid development, it has elicited criticism for enabling censorship, surveillance, and human rights infringements, as exemplified by instruments such as its social credit system (Bradford, 2023).

The EU adopts a rights-driven model, aiming to balance innovation with the protection of individual privacy and democratic values. While commended for its

commitment to user rights, it faces challenges, including allegations of impeding innovation and imposing burdensome regulation. However, Bradford (2024b) challenges this perspective and posits that there exists a complex relationship between regulation and innovation, suggesting that thoughtful regulation may even complement innovation by establishing standards that foster fair competition and trust. Furthermore, that it can stimulate ethical compliance and market certainty (Bradford, 2021).

These competing models not only reflect the values of their respective regions but also influence global approaches to digital governance, with significant implications for innovation, privacy, and societal structures.

While the comparative analyses conducted by Gromova et al. (2021), Weaver (2018), and Bradford (2023; 2024c) offer valuable insights into the divergent global regulatory philosophies, these typologies risk oversimplifying the fluid and contested nature of AI governance. Categorizing nations into discrete models can obscure significant internal variations, such as the increasing regulatory activism emerging within the United States at both state and federal levels, or the diverse degrees of enforcement and interpretation among EU member states (Cihon, Maas, and Kemp, 2021). Similarly, China's model, often characterized as wholly state-centric, exhibits elements of market dynamism and corporate innovation that complicate simplistic depictions of authoritarian control (Ding, 2021). Furthermore, the notion of three dominant ‘digital empires’ arguably neglects the growing influence of middle powers and regional alliances, such as the OECD, the African Union, and ASEAN, which are shaping global norms on data governance and AI ethics. Consequently, a more nuanced analysis recognizes that global AI regulation is neither static nor binary but constitutes an evolving ecosystem of overlapping legal, ethical, and geopolitical frameworks, in which convergence and contestation coexist.

AI is increasingly used in financial services for credit scoring, fraud detection, and algorithmic trading. The Equal Credit Opportunity Act and Fair Credit Reporting Act form the regulatory framework in the US, and the Consumer Financial Protection Bureau has confirmed these laws apply to AI-driven credit systems (Consumer Financial Protection Bureau, 2022; Streitz, 2023). Key requirements include non-discrimination obligations, requiring AI systems to avoid biased outcomes against protected groups; adverse action provisions, which oblige lenders to explain the principal factors underlying credit decisions regardless of model complexity; auditability, necessitating demonstrable explainability for regulatory review; and continuous monitoring, as credit models must be regularly tested and recalibrated to prevent discriminatory effects. This contrasts with the EU AI Act's approach to healthcare, which prioritises safety and human oversight, whereas financial regulation places greater emphasis on fairness and explainability (European Commission, 2021; Hacker, 2023). These sectoral distinctions illustrate how AI regulation is tailored to address context specific risks, notably patient safety in healthcare and consumer protection in financial services (Streitz, 2023).

The regulation of AI in employment, particularly concerning automated hiring systems, is increasingly emphasized to prevent discriminatory practices. In the United States, New York City has enacted Local Law 144 (New York City Government, 2023), which mandates bias audits of automated employment decision tools prior to their implementation. This legislation requires employers to disclose the use of AI to job candidates and to conduct independent audits to identify any discriminatory patterns (Raghavan et al, 2020). The primary objectives of this law are to ensure fairness, transparency, and the prevention of discrimination in workforce management (Bodie et al, 2020).

While these examples demonstrate the sectoral precision of AI regulation, they also reveal a growing fragmentation across jurisdictions and industries. The differentiation between fairness in finance, safety in healthcare, and bias prevention in employment illustrates regulatory responsiveness, yet it risks creating inconsistencies in enforcement and interpretability (Hacker, 2023). Moreover, such sector-specific measures may struggle to anticipate cross-domain challenges, such as the deployment of general-purpose AI systems that transcend traditional boundaries of regulation (Veale and Zuiderveen Borgesius, 2021).

Critics argue that the proliferation of localised or industry-specific rules could lead to compliance fatigue, particularly among multinational organisations, and hinder efforts to develop coherent global governance standards. Thus, while sectoral regulation advances ethical safeguards, it underscores the need for integrative frameworks that harmonise principles of fairness, transparency, and accountability across domains.

Wong (2021) examines ethical and self-regulatory measures, particularly in contexts where traditional regulations might impede technological progress. The paper discusses the shift towards company-driven ethical frameworks and corporate governance as a complementary measure to governmental regulation (Wong, 2021). In a similar vein, Hoffmann-Riem (2020) reviews the role of corporate self-regulation in the absence of strict governmental frameworks, identifying both the limitations and benefits of industry-led ethical initiatives (Hoffmann-Riem, 2020).

Wong (2021) and Hoffmann-Riem (2020) underscore the adaptive potential of corporate self-regulation within dynamic technological contexts. However, this reliance on voluntary governance raises substantial concerns regarding accountability and consistency. Ethical frameworks devised by private entities often lack enforceability and may prioritize reputational risk management over genuine public interest (Floridi, 2023).

Moreover, in the absence of external oversight, self-regulation risks generating fragmented ethical standards and inconsistent compliance across sectors and jurisdictions (Cath, 2018). Critics contend that such initiatives may serve as a form of 'ethics washing,' projecting responsibility without enacting substantive change (Bietti, 2020). Consequently, while self-regulation may temporarily address policy gaps, sustainable AI governance ultimately necessitates independent auditing, stakeholder participation, and integration with formal regulatory structures to ensure legitimacy and effectiveness.

Fortes et al. (2022) provide insights into the regulatory landscape concerning algorithmic risks in AI. Their analysis underscores the necessity of adaptive regulatory frameworks that can keep pace with rapid technological evolution. They propose an algorithmic governance model to address the unique challenges posed by automated decision-making systems (Fortes et al., 2022). Similarly, Buiten (2019) advocates for 'intelligent regulation' that aligns regulatory mechanisms with AI's potential unpredictability and rapid development cycles, suggesting an approach that is both precautionary and adaptable (Buiten, 2019).

While Fortes et al. (2022) and Buiten (2019) highlight the importance of adaptive regulation, excessive flexibility can weaken accountability and obscure responsibility when innovation outpaces oversight (Yeung, 2022). Moreover, algorithmic governance models often privilege technical risk assessment over ethical and social considerations, underscoring the need for regulatory mechanisms that integrate transparency, stakeholder participation, and normative evaluation (Ananny and Crawford, 2018).

Finocchiaro (2024a) examines the geopolitical context of the EU's proposed AI regulation, considering the global race in AI development. The paper highlights the regulatory competition between the EU and other leading AI economies, reflecting the

strategic implications of AI policy on global influence and market dynamics (Finocchiaro, 2024a).

While Finocchiaro (2024b) effectively positions AI regulation within a global power framework, interpreting the European Union's actions primarily as a competitive strategy may overlook its normative focus on ethics and human rights (Taddeo and Floridi, 2021). Furthermore, perceiving regulatory divergence as a contest for influence fails to account for the increasing interdependence between jurisdictions and the development of multilateral initiatives aimed at promoting interoperable and ethically aligned AI governance (Kiggins, 2023).

The landscape of AI regulation is characterized by a complex interplay between emerging technological advancements and the necessity for comprehensive governance frameworks. As AI systems become increasingly integrated into societal infrastructure, ethical considerations, legal implications, and practical regulatory measures have become paramount in discourse across regions and sectors.

This review illustrates a broad spectrum of regulatory considerations that continue to evolve as AI technology advances, showing a strong interdependence between national policies, international cooperation, and sector-specific requirements.

There exists a widespread consensus in the literature regarding the perception of regulation not as an impediment but as a potential enabler of sustainable innovation. Nevertheless, regulation-driven innovation necessitates further empirical investigation to explore its broader implications and to assist in mitigating against excessive regulation that may impede progress.

The impact of fragmented markets within regulatory regions, such as the EU, on the scaling of technology and innovation warrants a more thorough investigation to gain a comprehensive understanding of this dynamic. Furthermore, analysing the impact of

regulation on smaller technology firms in comparison to large enterprises could yield valuable insights into the innovation dynamics across firms of varying sizes and developmental stages.

In conclusion, the regulation of AI is a multifaceted issue that necessitates a comprehensive approach encompassing ethical, legal, and technical dimensions. The extant literature underscores the urgent need for standardized frameworks that address the complexities of AI technologies while promoting transparency, accountability, and ethical practices. As AI continues to evolve, ongoing research and dialogue will be imperative in shaping effective regulatory strategies that safeguard public interests and foster innovation.

2.6 Corporate Governance and Stakeholder Accountability

Governance refers to the structures and processes through which an organisation is directed, administered, and controlled. It encompasses the rules, practices, and accountability mechanisms that shape organisational behaviour and influence how an organisation is perceived by current and prospective stakeholders (Ullah et al., 2017).

Corporate governance frameworks delineate the allocation of rights and responsibilities among various corporate entities, including boards, managers, shareholders, and other stakeholders. These frameworks also establish the rules, procedures, and decision-making processes for corporate matters. In business organizations, governance is a multifaceted concept that encompasses the mechanisms, structures, and processes utilized to control and direct companies. It involves a complex balance of accountability, strategic guidance, decision-making, and engagement with stakeholders.

Chelliah, Boersma, and Klettner (2015) emphasize the significance of adaptive governance structures that address the specific requirements of commercial entities. They

posit that governance frameworks must evolve to maintain relevance in a rapidly changing business landscape, where market conditions and regulatory requirements are in constant flux. This adaptability is crucial for maintaining stakeholder trust and ensuring long-term sustainability (Chelliah et al., 2015).

A central theme in the literature is the relationship between governance and organizational performance. Research by Verschuere and Beddeleem (2013) suggests that effective governance can drive innovation and enhance organizational outcomes. They found that organizations with well-defined governance mechanisms are better positioned to foster creativity, ensure compliance, and manage risks effectively (Verschuere & Beddeleem, 2013). Similarly, Bart and Deal (2006) highlight the importance of a robust governance structure in strategic decision-making, noting that it enables boards to align corporate objectives with broader societal expectations (Bart & Deal, 2006).

Godenhjelm (2016), who explores the governance of project organizations, underscores the necessity for flexible governance mechanisms that can adapt to varying project requirements (Godenhjelm, 2016). Such flexibility is crucial within the rapidly evolving field of AI.

While the literature underscores adaptability and flexibility as crucial attributes of contemporary governance frameworks, these qualities can also create tensions between accountability and agility. Excessive flexibility may undermine the consistency and transparency of decision-making, particularly in complex technological contexts such as AI, where oversight demands are heightened (Kaptein, 2022). Furthermore, much of the existing research presumes that governance improvements directly lead to enhanced performance; however, empirical evidence indicates that this relationship is contingent upon contextual factors such as organizational culture, leadership capability, and ethical orientation (Aguilera and Crespi-Cladera, 2016). Consequently, the challenge for modern

governance lies not merely in structural adaptation but in embedding ethical and responsible principles within evolving organizational systems to ensure both legitimacy and resilience.

Stakeholder engagement is a recurring theme in governance literature. Effective governance necessitates clear communication channels and accountability to various stakeholders, including shareholders, employees, and the broader community. Sacchetti and Borzaga (2021) posit that governance frameworks should not solely focus on shareholder interests but also consider the broader societal implications of corporate decisions. They propose a model of inclusive governance that incorporates diverse stakeholder perspectives (Sacchetti & Borzaga, 2021). It is essential for organisations to understand the implications of these interactions and to develop frameworks that promote effective governance practices.

While Sacchetti and Borzaga (2021) present a persuasive case for inclusive governance, their model assumes a degree of stakeholder alignment that may prove challenging to realize in practice. The increasing complexity of global supply chains and digital ecosystems complicates efforts to reconcile the divergent priorities of investors, regulators, and affected communities (Crane and Matten, 2021). Moreover, although stakeholder engagement is frequently portrayed as inherently advantageous, critics argue that without explicit accountability mechanisms, it risks devolving into symbolic participation or reputational management rather than exerting substantive influence (Greenwood, 2022). In the realm of AI governance, these tensions are exacerbated, as stakeholders may lack the technical literacy necessary to critically assess algorithmic decisions. Consequently, while participatory approaches enhance legitimacy, their effectiveness ultimately hinges on transparency, capacity building, and institutionalized channels for stakeholder oversight.

Dotan usefully provides a streamlined AI governance framework aimed at unifying the EU AI Act and the NIST AI Risk Management Framework (AI RMF) (Dotan, 2024). Recognizing the complexities and challenges of AI governance due to a multitude of regulatory requirements, the framework condenses these into twelve questions, categorized into four pillars: Discovery, Measurement, Mitigation, and Accountability.

The framework's multi-tiered approach enables organizations to select their level of engagement, offering actionable guidance for those requiring comprehensive AI governance strategies (Dotan, 2024). Its emphasis on a structured governance process, ranging from ad hoc to adaptive activities, underscores the evolution necessary for organizations to attain a mature AI governance stance.

However, while ostensibly pragmatic, the extent to which this combined approach will be adopted remains uncertain.

Dotan's (2024) integrative model constitutes a significant advancement towards achieving coherence between regulatory and governance frameworks. However, it potentially overestimates the existing level of institutional capacity and cross-jurisdictional alignment. The harmonization of legislative instruments, such as the EU AI Act, with voluntary standards like the NIST AI RMF, assumes the presence of shared interpretive norms and consistent enforcement, which remain aspirational due to the varying legal cultures and policy objectives across different regions (Mökander and Floridi, 2023). Furthermore, while a tiered governance approach offers adaptability, it may inadvertently promote minimal compliance behaviors unless bolstered by external accountability mechanisms and sector-specific oversight (Cath, 2021). Therefore, although the framework effectively bridges compliance and governance, its ultimate

success hinges on the degree to which organizations internalize ethical and procedural standards beyond formal regulatory mandates.

The interplay between corporate governance structures and organisational performance underscores the significance of robust governance mechanisms. Corporate governance encompasses a network of relationships among stakeholders, shareholders, management, and the board of directors, necessitating clear guidelines for accountability, particularly in the context of emerging technologies such as AI. The literature emphasizes the necessity of collaborative frameworks, ethical standards, and human oversight to navigate the ethical complexities of AI systems, thereby fostering trust and transparency.

Furthermore, governance in commercial organizations constitutes an evolving and multifaceted domain. It necessitates adaptive structures, rigorous accountability, and active stakeholder participation to ensure ethical conduct, organizational resilience, and sustainable growth. The ongoing evolution of the business landscape highlights the imperative for governance systems that not only meet commercial objectives but also align with broader societal expectations, advocating for a comprehensive and integrated approach to governance.

2.7 Ethical Foundations and the Responsible AI Imperative

The history of ethics as a philosophical discipline has deep roots. Early ethical thought can be traced to ancient civilizations, wherein philosophers began to contemplate the nature of morality and the principles governing human behaviour. Seminal figures such as Socrates, Plato, and Aristotle established the foundation for ethical inquiry in the Western tradition.

The Milesian school, considered the origin of Western philosophy, focused on understanding the natural nonhuman world and humanity's role in the cosmos (Katz, 1991). Throughout history, philosophers have examined moral concepts and ideas,

emphasizing the significance of historical context in understanding ethical principles (Macintyre, 2003). Socrates emphasized the importance of self-examination and the pursuit of virtue, while Plato (2020) introduced the concept of ideal forms, positing that ethical truths exist beyond the physical realm. Aristotle (2020) further developed these ideas by proposing virtue ethics, which focuses on the character of the moral agent and the importance of achieving eudaimonia, or human flourishing, through virtuous actions.

During the medieval period, ethical thought was significantly influenced by religious doctrines, particularly within Christianity and Islam. The Enlightenment precipitated a paradigm shift towards secular ethics, with philosophers such as Immanuel Kant (2019) propounding deontological ethics, which emphasizes duty and adherence to moral laws. Kant's categorical imperative represents a foundational principle of modern ethical theory, asserting that moral action should be guided by maxims that one could rationally will to become universal laws.

This epoch also witnessed the emergence of utilitarianism, espoused by figures such as Jeremy Bentham and Mill (1998), which evaluates the morality of actions based on their consequences for aggregate well-being.

The evolution of ethical concepts such as respect for persons, beneficence, non-maleficence, truth-telling, and social responsibility can be traced 'across time, cultures, and professions' (Sinclair, 2012). Notably, while individualistic ethics became explicit with the Sophists and Socrates, environmental philosophy and ethics can be traced back to the very origins of Western philosophy itself (Katz, 1991).

There has been a resurgence of interest in ethics as a philosophical discipline, precipitated by the publication of John Rawls's A Theory of Justice and the necessity to address ethical issues arising from technological advancements (Drummond, 2002). The development of applied ethics fields such as bioethics and environmental ethics has

further contributed to this renewed focus on ethical inquiry (Löfström et al., 2021; Stanar, 2023).

Despite the extensive history of ethical discussion, several limitations can be identified in the approaches undertaken. A significant critique pertains to the tendency of ethical theories to be excessively prescriptive, often failing to account for the nuances and complexities inherent in real-world moral dilemmas. For instance, strict adherence to utilitarian principles may result in ethically questionable outcomes, such as justifying detrimental actions for the sake of greater aggregate welfare. Furthermore, the predominance of Western ethical frameworks may neglect diverse cultural perspectives, potentially leading to a form of ethical imperialism that inadequately represents global moral diversity.

While traditional ethical frameworks serve as crucial foundations for moral reasoning, their relevance to contemporary technological contexts remains debated. Classical theories such as deontology and utilitarianism, despite their philosophical rigor, often encounter challenges in addressing the distributed agency and opacity inherent in AI systems (Mittelstadt et al., 2016). Furthermore, the predominance of Western ethical traditions in mainstream discourse risks marginalizing pluralistic perspectives that may more effectively capture the relational and contextual dimensions of responsibility, as highlighted in African communitarian ethics or Eastern notions of harmony and interdependence (Ess, 2020). This situation raises significant questions regarding the values that underpin emerging AI ethics frameworks and how inclusivity can be achieved in their global application. Consequently, while historical ethical traditions remain intellectually significant, the complexity of technological ethics necessitates an adaptive, intercultural, and practice-oriented approach that integrates normative theory with applied governance.

In conclusion, the evolution of ethical thought reflects an ongoing interaction between philosophical theory and practical moral judgement. Although significant advances have been made in articulating ethical principles, persistent difficulties remain in their application to contemporary challenges, particularly those arising from rapidly advancing technologies such as AI. Addressing these complexities requires a more integrated and context sensitive approach to ethics that is capable of responding to dynamic socio technical conditions.

The field of digital ethics, particularly as it relates to AI, is an evolving area of study addressing the ethical concerns surrounding the development and deployment of digital technologies. Central to digital ethics is the concept of establishing frameworks to guide the responsible utilization of AI technologies. As Floridi (2021) asserts, digital ethics has emerged as a crucial component in addressing the complex moral and societal challenges precipitated by the widespread implementation of digital platforms and AI. The necessity for a robust ethical framework is increasingly apparent, especially as AI systems begin to significantly influence decision-making processes across various domains.

One of the critical debates within digital ethics is the establishment of universal principles to govern AI's development and the complexities involved in creating a global ethical framework. Principles such as transparency, fairness, and accountability, have become central tenets in AI ethics. However, there is also a recognition of the potential risks of ‘ethics-washing,’ where organizations superficially adopt ethical principles without meaningful implementation (Floridi, 2019).

The intersection of AI and ethics presents novel challenges regarding the moral agency of machines. Hanna and Kazim (2021) discuss the concept of a ‘dignitarian approach’ to AI ethics, focusing on human dignity as a guiding principle. They posit that

while AI can enhance efficiency, the systems themselves lack moral agency. Consequently, ethical considerations should focus on how AI technologies impact human users, ensuring that dignity and human rights remain central to their design and implementation.

Stahl (2022) extends this argument, suggesting that digital ethics should transcend individual AI systems to consider the broader ecosystem within which these technologies operate. This ecosystemic perspective emphasizes the interconnected nature of digital technologies, where ethical issues in AI cannot be isolated from broader societal and environmental contexts. Stahl (2021) underscores the importance of ethical foresight in addressing long-term consequences, advocating for a sustainable and inclusive approach to AI governance.

Operationalizing AI ethics remains a significant challenge, as it necessitates the translation of abstract ethical principles into concrete practices. Morley et al. (2023) identify several barriers to operationalizing AI ethics, such as the absence of standardized ethical guidelines and the varying interpretations of what constitutes ethical AI. They advocate for a more collaborative approach, wherein AI developers and users cooperate to ensure that ethical considerations are embedded throughout the AI lifecycle.

While the literature on digital ethics emphasizes the significance of principles such as transparency, fairness, and accountability, critics contend that principle-based approaches often lack the specificity necessary for practical implementation (Bietti, 2020). The proliferation of ethical guidelines and frameworks has resulted in conceptual redundancy, with limited mechanisms for evaluation or enforcement (Jobin, Ienca, and Vayena, 2019). Moreover, framing digital ethics primarily around universal principles risks neglecting the socio-technical realities of AI systems, which are influenced by contextual, cultural, and economic factors (Greene, Hoffmann, and Stark, 2019). The

dignitarian approach proposed by Hanna and Kazim (2021) effectively re-centres human welfare, yet questions persist regarding how such human-centred values can be operationalized in global, data-driven ecosystems governed by corporate interests. Thus, while digital ethics provides a crucial moral compass for AI governance, its efficacy ultimately depends on the integration of enforceable standards, interdisciplinary oversight, and ongoing reflexivity concerning power, accountability, and cultural inclusivity.

In conclusion, digital ethics, particularly in the context of AI, aims to provide a moral framework to guide the responsible development and utilization of technology. The primary challenges involve the operationalization of ethical principles, ensuring that AI systems are designed and deployed in ways that safeguard human dignity, rights, and societal well-being. As AI continues to evolve, digital ethics will play a crucial role in navigating the complex moral landscapes created by emerging technologies.

The ethical governance of AI has emerged as an increasingly critical area of focus, given the rapid integration of AI systems into social, commercial, and governmental sectors. An expanding body of literature critiques the development, deployment, and societal implications of AI technologies, with particular emphasis on fairness, accountability, and transparency. The field of ethics plays a pivotal role in the adoption, implementation, and utilization of AI systems, as organizations address the implications of incorporating these technologies into their operations.

One of the primary ethical concerns in AI is the potential for adverse consequences resulting from inadequately designed systems. Favaretto et al. argue that researchers must prioritize ethical conduct in big data behavioural studies to prevent harm to individuals and communities (Favaretto et al., 2020). This perspective underscores the

necessity of embedding ethical considerations into the design and implementation of AI systems to mitigate risks associated with data misuse and algorithmic bias.

Transparency in AI systems constitutes a significant ethical consideration. Vakkuri et al. underscore the necessity of transparency concerning algorithms and data to cultivate trust in AI technologies (Vakkuri et al., 2019). In the absence of a comprehensive understanding of AI system operations, stakeholders may exhibit reluctance in adopting these technologies, thereby hindering their potential advantages. Jobin and Ienca further corroborate this perspective, asserting that transparency is a fundamental prerequisite for the ethical governance of AI (Jobin & Ienca, 2019). They contend that while establishing public trust in AI is crucial, it must not compromise scrutiny and accountability.

Furthermore, the development of ethical frameworks for AI is essential for guiding organizations in their implementation of these technologies. Morley et al. emphasize the importance of translating ethical principles into practical tools and methods that organizations can utilize (Morley et al., 2019). This translation process is crucial for ensuring that ethical considerations are not merely theoretical but are actively integrated into the operational practices of organizations.

The introduction of AI ethics guidelines and frameworks has also been a significant development in the field, however Schiff et al (2020) caution that the homogeneity of these documents may pose challenges in addressing the diverse ethical concerns associated with AI. This observation underscores one of the primary challenges with AI, namely that it is a vast and rapidly evolving field and highlights the need for ongoing dialogue and collaboration among stakeholders to ensure that ethical frameworks are comprehensive and adaptable to the evolving landscape of AI technologies.

Floridi conceptualizes AI as a unique form of agency, distinguishing it from traditional notions of intelligence. He argues that the ethical considerations surrounding AI emerge from the separation between agency and intelligence, highlighting risks, the balance between ethical principles and legal norms, and the concept of 'soft ethics', ethics beyond mere compliance (Floridi, 2023). Floridi (2023) suggests that the evolution of AI will be driven by synthetic data generation, reducing reliance on historical datasets and addressing concerns about data privacy and sensitivity. This shift allows AI to maintain high-quality training while navigating privacy challenges.

Blackman (2022a) contends that ethical considerations are integral and must be systematically integrated throughout the AI development lifecycle. He identifies three primary risks associated with AI: scale, speed, and moral complexity, and posits that organizations should address ethical risk with the same diligence as financial or legal risk. According to Blackman, a mere commitment to overarching ethical principles such as fairness or transparency is inadequate unless these principles are translated into actionable protocols, staff training, and evaluative mechanisms (Blackman, 2020).

While Blackman's approach usefully advances the operationalisation of AI ethics within organisations, it risks conflating ethical responsibility with corporate self-interest, particularly in the absence of independent oversight (Bietti, 2020). His framework therefore benefits from being situated alongside broader models that emphasise professional virtue, reflexivity, and external accountability as integral to sustaining genuinely ethical AI governance (Dignum, 2019; Vallor, 2016).

Blackman and Euchner (2024) emphasize that ethical responsibility in AI should not be relegated solely to compliance teams. Instead, it should align with broader trends in responsible innovation and be integrated into organizational culture and decision-

making processes at all levels. They propose that ethics should be a cross-functional concern, avoiding the traditional siloed approach to ethical considerations.

In addressing the ethical challenges posed by AI, Floridi (2023) proposes a comprehensive framework of ethical principles. He acknowledges the proliferation of AI ethics guidelines since 2017, highlighting the resultant complexity due to overlapping or conflicting principles. Through a comparative analysis of prominent guidelines, including the Asilomar Principles, the Montreal Declaration, and IEEE's Ethically Aligned Design, he identifies a convergence around five core ethical principles, with the aim of establishing a cohesive ethical framework for AI governance. These five principles are as follows: Beneficence, which mandates that AI should enhance human well-being, preserve dignity, and support planetary sustainability; Nonmaleficence, which requires AI systems to avoid harm, protect privacy, and ensure security; Autonomy, which emphasizes the respect for human autonomy, with AI designed to support rather than undermine human decision-making; Justice, which advocates for fairness, the elimination of bias, and equitable access to AI's benefits; and Explicability, a principle unique to AI, encompassing transparency and accountability, necessitating that AI's operations and impacts are comprehensible and that its developers are held accountable.

Floridi highlights that explicability is essential because AI operates as an autonomous form of agency that is often opaque. This principle enables other ethical standards by making AI systems more transparent and accountable. He concludes that this unified framework of ethical principles could serve as a foundation for regulations, technical standards, and best practices, facilitating a shift from principles to practices that ensure AI serves society's best interests (Floridi, 2023).

While the literature presents valuable frameworks for embedding ethics into AI design and governance, much of it assumes that ethical principles can be universally

applied across diverse social and cultural contexts. This presumption risks overlooking the situated nature of moral reasoning and the socio-technical contingencies that shape AI's real-world effects (Birhane, 2021). Furthermore, the proliferation of overlapping ethics guidelines has led to conceptual saturation without clear mechanisms for enforcement or evaluation (Jobin, Ienca and Vayena, 2019). Critics contend that without institutional accountability, ethics frameworks may serve as legitimising tools rather than transformative instruments for responsible innovation (Greene, Hoffmann and Stark, 2019). Therefore, while principle-based models such as Floridi's (2023) provide valuable normative coherence, their effectiveness ultimately depends on the establishment of enforceable standards, cross-sectoral governance structures, and ongoing critical reflection on power, inclusivity, and societal impact.

In conclusion, ethical considerations are fundamentally crucial to the successful adoption, implementation, and utilization of AI systems. Prioritizing ethical considerations, promoting transparency, and developing robust frameworks enable organizations to navigate the complexities of AI while fostering responsible innovation. Continuous dialogue among stakeholders is imperative to address the ethical challenges associated with AI, ensuring these technologies serve the broader interests of society.

As organizations enhance their comprehension of the ethical and societal implications of AI, subsequent endeavours should concentrate on translating ethical principles into implementable practices, examining the diverse global impacts of AI, and considering long-term societal transformations. Facilitating cross-sector collaboration and incorporating global perspectives will be essential to ensure that the development of AI remains inclusive and sustainable.

The preceding analysis of ethics underscores that moral integrity in the AI era must extend beyond mere philosophical commitment, necessitating its assimilation into

corporate strategy and organizational structure. The convergence of significant technological risks (scale, speed, and moral complexity), stringent regulatory requirements (such as the EU AI Act), and the need for robust corporate governance frameworks positions ethical fluency at the core of effective leadership in commercial organizations.

Ethical fluency constitutes a fundamental leadership capability, ensuring that organizations perceive ethical conduct not merely as a compliance obligation but as a strategic asset that enhances corporate reputation and stakeholder trust. Leaders must advocate for values-based leadership, aligning their behaviours with core organisational and personal values to cultivate the trust and accountability essential for ethical conduct. The strategic role of leadership encompasses managing the significant organizational and systemic challenges inherent in the adoption of AI.

2.8 Leadership, Organisational Theory and Systemic Change

A prominent issue identified in the literature is the ‘accountability gap’ that arises when AI systems function without sufficient human oversight. Additionally, assertions that ethical AI fosters trust frequently lack empirical support, and organizations risk engaging in 'ethics-washing', superficially adopting principles without implementing meaningful, substantive changes.

Ethical fluency enables leaders to translate core ethical principles, including beneficence, non-maleficence, autonomy, justice, and explicability, into concrete organisational practice. As Blackman (2020) argues, ethical principles remain insufficient unless they are operationalised through actionable protocols, workforce training, and evaluative mechanisms embedded across the AI development lifecycle. Ethical responsibility must therefore be treated as a cross functional leadership concern,

extending beyond siloed compliance functions and informing decision making at all organisational levels.

The field of Organization Development (OD) has undergone significant evolution since its inception in the mid-20th century. It emerged from human relations and behavioural science fields, with the objective of enhancing organizational effectiveness and individual well-being through planned change (Burnes & Cooke, 2012). Kurt Lewin's seminal work on action research and change theory in the 1940s established the foundation for OD, emphasizing the significance of participative methodologies and empirical inquiry in organizational change (Reason & McArdle, 2006). Over subsequent decades, OD has expanded to encompass a diverse array of approaches and has experienced shifts in its theoretical and practical foci.

2.8.1 Historical Foundations

Kurt Lewin's action research, particularly his concepts of group dynamics and field theory, were instrumental in establishing the early foundations of Organizational Development (OD). Lewin's proposition that change should be a participatory process, wherein individuals are actively engaged in diagnosing and resolving problems, became a fundamental principle of OD practices (Burnes & Cooke, 2012). This participative approach aimed to address both organizational performance and employee satisfaction. In the 1960s, OD gained further momentum through the work of consultants such as Richard Beckhard and Edgar Schein, who integrated systems thinking with organizational change processes (Hinckley, 2014).

Seminal approaches in OD, such as sensitivity training, socio-technical systems, and team building, focused on enhancing interpersonal relations within organizations. These interventions primarily targeted leadership development and the improvement of group performance. The fundamental objective of OD during this period was to establish

adaptive organizations that balanced efficiency with humane working environments (Grieves, 2000). Nevertheless, this initial emphasis on interpersonal relations faced criticism for its insufficient attention to broader structural and systemic factors influencing organizational performance.

2.8.2 Theoretical Frameworks and Developments

OD frameworks have evolved to address the limitations of early humanistic models. Systems theory and contingency theory became influential, as they emphasized the complex, interdependent nature of organizational components. These theories posited that OD should consider both internal dynamics and external environmental factors when designing interventions (Rothwell & Stavros, 2015). Moreover, Burke and Litwin's (1992) model of organizational change introduced the concept of transformational leadership, proposing that effective OD must align with profound changes in organizational culture and values.

In more recent developments, constructionist and postmodern approaches have entered the field, reflecting a paradigm shift from the mechanistic perspectives of the past. Marshak and Grant (2008) posit that contemporary OD practices are increasingly informed by discursive and narrative approaches, which emphasize the role of language and social constructs in shaping organizational realities. This shift underscores the field's growing recognition of power dynamics, identity, and cultural influences in organizational life.

2.8.3 Critique and Challenges

Despite its advancements, Organizational Development (OD) has faced numerous critiques over the years. A significant challenge is its limited focus on structural inequalities within organizations. Early OD interventions were predominantly confined to improving interpersonal relations without adequately addressing systemic issues such as

power imbalances, economic constraints, and technological change (Cooke, 1998). Furthermore, OD's reliance on subjective methodologies, including action research and participatory techniques, has raised concerns regarding the consistency and replicability of its outcomes.

Another critique pertains to the field's capacity to adapt to rapidly evolving technological and economic changes. While OD was initially conceptualized as a human-centred approach to change, the increasing complexity of contemporary organizations and the impact of globalization have necessitated more robust, scalable interventions capable of managing large-scale transformations (Burke, 2018). Consequently, the future of OD may require a more substantial integration of technology-driven solutions and data analytics to maintain relevance in modern organizational contexts.

2.8.4 Gaps and Future Directions

The existing literature on OD highlights significant gaps in addressing global and cross-cultural contexts. Many OD theories and models have been formulated within Western frameworks, which restricts their applicability to non-Western organizational settings (Hinckley, 2014). Additionally, further empirical research is required to assess the long-term effectiveness of OD interventions across various industries and cultural environments. The integration of insights from disciplines such as organizational psychology, technology management, and international business could potentially address these identified gaps.

Although OD has evolved to incorporate greater theoretical sophistication, its humanistic and participatory foundations continue to attract critique for insufficient engagement with structural power, digital transformation, and the ethical challenges posed by emerging technologies. As Cummings and Worley (2019) observe, OD’s traditional emphasis on interpersonal processes and cultural alignment has not always

kept pace with the complexities of data driven and AI enabled organisations operating within global and highly dynamic environments. Similarly, postmodern and dialogic approaches, while valuable in foregrounding discourse, identity, and meaning, risk diluting accountability by privileging narrative over material conditions and measurable outcomes (Marshak and Grant, 2011). These critiques point to the need for a renewed integration of OD principles with contemporary governance and ethical frameworks, enabling organisational change interventions that remain both participatory and responsive to the technological, societal, and moral demands of AI informed transformation.

2.8.5 Dialogic OD

Dialogic Organisational Development represents a significant shift from traditional diagnostic and prescriptive change models that rely on top-down intervention and linear problem solving (Marshak and Bushe, 2021). Grounded in complexity science and interpretivist social theory, Dialogic OD views organisations as socially constructed systems in which meaning, identity, and action emerge through language, interaction, and collective sense making (Bushe, 2021). Rather than assuming change can be designed and imposed, this approach recognises that order and innovation often arise from uncertainty, dialogue, and self-organising processes.

Dialogic OD prioritises conversational processes over predetermined solutions, emphasising the transformation of narratives and interpretive frames rather than the diagnosis and correction of discrete problems. Change is understood as emergent and non-linear, shaped through participation, reflection, and the co creation of meaning. The approach encompasses both structured interventions, such as Appreciative Inquiry and Open Space, and less formal processes that enable autonomous and self-initiated change. In doing so, Dialogic OD maintains the humanistic and democratic foundations of OD

while offering greater adaptability in dynamic organisational contexts (Cheung Judge and Holbeche, 2015).

At its core, Dialogic OD positions dialogue as a mechanism for engagement and transformation, extending beyond facilitated discussion to engage with the underlying narratives that shape organisational behaviour (Bushe and Marshak, 2014). Recent scholarship has further emphasised the importance of psychological safety, leadership vulnerability, and inclusive participation in enabling dialogic processes to contribute to more equitable and just organisational outcomes (Kwon and Kwon, 2023). By involving diverse stakeholders and encouraging multiple perspectives, Dialogic OD supports more inclusive sense making and decision making, which is particularly relevant in ethically sensitive and technologically complex domains.

Bushe's generative change model highlights how dialogic processes can enable rapid and substantive transformation by activating the self-organising capacity of human systems and mobilising stakeholder engagement (Bushe, 2021). As a result, Dialogic OD is especially well suited to complex and uncertain environments in which traditional planned change approaches are insufficient. Frameworks such as the Cynefin model further support its application in contexts characterised by complexity and chaos, where emergent and adaptive responses are required. This positions Dialogic OD as a relevant approach for organisations grappling with AI enabled transformation, where ethical uncertainty, socio technical complexity, and evolving risk demand participatory and generative forms of change.

2.8.6 Critique of Dialogic OD

Despite its strengths, Dialogic OD is not appropriate in all organisational contexts. In highly regulated sectors or organisations characterised by strong hierarchical cultures, its emergent and participatory orientation may conflict with expectations for

clear authority, procedural consistency, and directive leadership (Institute for Organisation Development, 2024). Such environments often require structured, linear change processes that sit uneasily alongside dialogic approaches.

The emphasis on self-organising and emergent change can also introduce unpredictability, which may be problematic for organisations that prioritise stability, control, and predefined outcomes (Zeno, 2013). Furthermore, because Dialogic OD focuses on shifts in narratives, mindsets, and cultural meaning, its outcomes can be difficult to quantify using conventional performance metrics. This presents challenges for organisations that require demonstrable return on investment or clearly defined key performance indicators to justify change interventions (Bushe and Marshak, 2014).

Concerns have also been raised regarding accountability. By favouring collective sense making and distributed agency, Dialogic OD may obscure responsibility for specific decisions and outcomes, creating ambiguity in implementation and follow through (Zeno, 2013). Scholars have additionally questioned its scalability and institutional reliability, noting that an emphasis on discursive transformation can marginalise material conditions and measurable organisational impact (Marshak and Grant, 2011). Its applicability within risk averse or compliance driven environments therefore remains contested, as open-ended dialogue may conflict with governance and regulatory requirements (Burnes, Hughes and By, 2018).

Nevertheless, Dialogic OD offers valuable mechanisms for fostering trust, inclusivity, creativity, and ethical reflexivity, particularly in complex and uncertain contexts such as AI enabled transformation. The central challenge lies in integrating its adaptive and participatory strengths with the structural clarity and accountability of diagnostic approaches. A blended model may therefore be required to achieve both legitimacy and responsiveness in contemporary organisational change.

2.8.7 The Cynefin Model

The Cynefin framework is a sense-making model designed to assist organizations and individuals in navigating complexity and uncertainty within decision-making processes (Snowden & Boone, 2007). It serves as a practical tool for leaders to discern the nature of their situations and select appropriate strategies (Snowden, 2003). The framework comprises five domains, clear, complicated, complex, chaotic, and confused (disorder), which reflect distinct types of systems and decision environments, with specific methodologies proposed for each (Kurtz & Snowden, 2003).

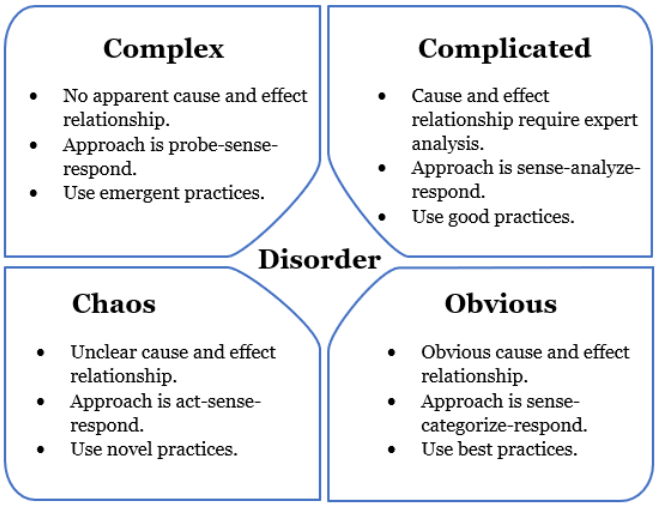


Figure 2.1: The Cynefin Framework

The Cynefin model draws upon several theoretical frameworks, most notably complexity theory, systems thinking, and cognitive science. Complexity theory is central, particularly in recognizing the unpredictable and non-linear relationships present in complex systems (Snowden, 2003). The model distinguishes between systems that are predictable, where best practices or expert knowledge can be applied (clear and complicated), and those that necessitate experimentation and adaptation (complex and chaotic) (Kurtz & Snowden, 2003). By integrating these perspectives, Cynefin provides a

nuanced approach to decision-making that contrasts with more prescriptive models, such as traditional contingency or rationalist theories.

Several empirical studies have substantiated the efficacy of the Cynefin framework across diverse contexts. Snowden and Boone (2007) illustrated its application in business leadership, emphasizing its utility in comprehending dynamic and emergent phenomena. In public policy, the model has been utilized to address governance and crisis management, particularly in situations where traditional predictive models prove inadequate (Pell, 2016). More recently, researchers have employed the Cynefin framework to examine organizational agility and innovation, specifically in adapting to technological disruptions and global uncertainties (Baran & Woznyj, 2020). Its application in complex and chaotic environments renders it particularly suitable for sectors such as healthcare, software development, and emergency management (Paley, 2010).

2.8.8 Strengths and Weaknesses

One of the principal strengths of the Cynefin framework lies in its ability to support sense making in complex and chaotic contexts where linear decision-making models are inadequate. Its distinction between complicated and complex problems challenges reductionist approaches and encourages experimentation, learning, and adaptive responses (Snowden and Boone, 2007). The framework's applicability across multiple sectors further enhances its practical relevance.

Despite these strengths, Cynefin is not without limitations. Critics note that the boundaries between domains are not always clearly delineated, increasing the risk of misclassification, particularly as situations evolve dynamically across domains (French, 2013). The categorisation of organisational realities into discrete domains may also oversimplify the fluid and overlapping nature of real-world contexts (Paley, 2010).

Moreover, while Cynefin promotes adaptive thinking, its application often relies on subjective judgement, introducing potential bias and inconsistency in interpretation (Kirsch, 2021). Although this heuristic flexibility is a core strength, it also constrains empirical validation and limits the framework's utility as a predictive or diagnostic tool (Kurtz and Snowden, 2003).

Nonetheless, Cynefin's emphasis on sense making rather than control remains particularly relevant in domains characterised by uncertainty, non-linearity, and ethical ambiguity, such as AI governance. Integrating the framework with complementary approaches, including systems thinking, responsible innovation, and dialogic OD, may enhance its capacity to inform leadership practice and organisational adaptation. While Cynefin has made a significant contribution to decision making theory in complex environments, further research is required to strengthen its empirical grounding and theoretical integration.

2.8.9 Leadership Development

Leadership development has emerged as a critical area of focus within organizational studies, reflecting the necessity for effective leadership to navigate complex environments. The field is characterised by diverse theoretical frameworks, methodologies, and practical applications.

Research indicates that leadership development programs significantly enhance organizational performance by fostering essential leadership competencies. Longe's qualitative study highlights that organizations with robust leadership development initiatives are more likely to experience improved performance outcomes, as these programs cultivate transformational leadership qualities and support succession planning (Longe, 2023).

Moreover, the literature suggests that traditional approaches, which often emphasize individual traits, are inadequate for understanding leadership as a dynamic and adaptive process (Turner & Baker, 2017). This perspective aligns with findings from Purdy, who asserts that leadership development programs can motivate high-potential individuals and necessitate ongoing support mechanisms such as mentoring and coaching to ensure the application of learned competencies (Purdy, 2016).

Transformational leadership, as proposed by Bass and Avolio (1994), emphasizes the significance of leaders inspiring and motivating followers toward achieving exceptional outcomes. This approach is recognized for fostering innovation and enhancing organizational performance. Similarly, authentic leadership, emphasizing self-awareness and ethical behaviour, has gained prominence for its potential to build trust and encourage employee engagement (Avolio and Gardner, 2005).

Shared leadership models, which distribute leadership responsibilities across teams, have also emerged as significant, particularly in dynamic and complex organizational environments. This approach, highlighted by Pearce and Conger (2003), promotes collaborative decision-making, adaptability, and collective efficacy.

The transformational leadership framework is frequently cited as a foundational theory underpinning effective leadership development. For instance, Wijewantha emphasizes the role of transformational leaders in developing future leaders, thereby reinforcing the importance of follower development within leadership training (Wijewantha, 2019). Additionally, the integration of emotional intelligence within these programs has been demonstrated to enhance their effectiveness (Ullrich et al., 2020).

The integration of values-based leadership has brought ethical considerations to the forefront. Copeland (2014) asserts that aligning leadership behaviours with core organisational and personal values is essential for fostering trust and accountability.

Additionally, the intersection of leadership with diversity and inclusion, as explored by Ospina and Foldy (2009), highlights the necessity of context-sensitive approaches to leadership that address systemic inequities.

Despite the recognized benefits, challenges persist in measuring the return on investment (ROI) of leadership development initiatives, with numerous organizations failing to assess their impact adequately (Turner & Baker, 2017). This gap underscores the necessity for further empirical research to establish clear connections between leadership development interventions and organizational performance outcomes (Day & Dragoni, 2015). The literature advocates for a comprehensive approach to leadership development that incorporates both theoretical foundations and practical applications to foster effective leadership in diverse organizational settings.

Although leadership development is widely regarded as essential to organisational performance, much of the supporting literature remains normative, with limited longitudinal evidence linking interventions to measurable outcomes (Day et al., 2021). Additionally, prevailing models such as transformational and authentic leadership risk overemphasising individual traits at the expense of collective, ethical, and systems-based competencies needed to navigate complex, technologically mediated environments (Alvesson and Einola, 2019).

In conclusion, leadership development integrates transformational and authentic leadership theories, shared models, and values-based approaches to adapt to evolving organizational and societal demands. These frameworks collectively offer robust strategies for cultivating effective and ethical leadership.

An important capability pertaining to leadership is the organizational context by acknowledging that the responsible adoption of AI necessitates a comprehensive, system-wide transformation, aligning strategic clarity with operational governance. This requires

systemic change literacy, which extends beyond the interpersonal focus often critiqued in early organizational development approaches, to effectively manage the integration of AI into organizational culture and values. Leaders must advocate for values-based frameworks, ensuring that the organization perceives ethical conduct as a strategic advantage rather than merely a compliance obligation. Ultimately, systemic change literacy ensures that governance and ethical frameworks are consistently applied across all organizational levels and decision-making processes, thereby translating responsible innovation from a theoretical imperative into sustainable practice.

2.9 Synthesis, Research Gaps and Conceptual Model

Several enduring challenges in the governance of AI are identified in the literature. Firstly, there exists a significant disparity between high-level ethical principles and their practical implementation, as tools for translating these values into specific design requirements, metrics, and control mechanisms remain insufficiently developed (Rahwan et al., 2025; Zajko, 2025; Zhou, 2025). Secondly, organizations must navigate overlapping internal policies alongside fragmented, cross-jurisdictional regulatory frameworks, often resulting in inconsistent governance arrangements that hinder the development of coherent and integrated governance structures (Zhou, 2025; Garrigues, 2024). Thirdly, the literature indicates that commercial AI governance frequently fails to adequately represent the perspectives of affected stakeholders, particularly marginalized communities, while prioritizing legal compliance and reputational risk over broader societal impact (Zajko, 2025; Simply Sustainable, 2025). In response, there is requirement for increased empirical investigation across sectors, enhanced stakeholder participation, and systematic evaluation of how governance mechanisms influence AI behaviour and organizational outcomes, including their implications for performance, accountability, and trust (Rahwan et al., 2025; Xia, 2026; Toward AI Governance, 2022).

The literature review indicates that effective leadership in the era of AI cannot depend solely on individual competencies. Instead, it requires a comprehensive, multidisciplinary capability model that incorporates ethical, regulatory, and organizational imperatives. The essential leadership capabilities emerge at the intersection of five core themes, AI risks, regulation, governance, ethics, and organizational change, thereby defining a model centred on adaptive accountability within complex digital environments.

This synthesis establishes four interdependent traits for Responsible AI Leadership:

2.9.1 Ethical Fluency and Operationalisation

Responsible AI leadership necessitates that leaders possess ethical fluency, which enables them to convert abstract moral principles into tangible and actionable organizational protocols. This capability extends beyond mere normative compliance and addresses the existing gap between aspirational rhetoric and meaningful implementation. It is rooted in the convergence of five fundamental ethical principles: beneficence, nonmaleficence, autonomy, justice, and explicability. In particular, leaders must prioritize Explicability, a distinctive AI principle that encompasses transparency and accountability, which is essential for mitigating risks associated with opaque algorithmic agency. Ethical fluency ensures that bias mitigation, privacy protection, and risk assessment are integrated throughout the AI lifecycle, rather than being confined solely to compliance teams.

2.9.2 Adaptive Governance and Sense-Making

Leaders are required to establish governance structures that are both rigorous and adaptive, balancing accountability with organisational agility. This demands advanced sense making capabilities in contexts characterised by uncertainty and non-linearity. The

Cynefin model provides a suitable framework for addressing complex and chaotic AI related challenges, including the implementation of safe and responsible AI practices, by encouraging experimentation and emergent responses rather than reliance on prescriptive linear models. Adaptive governance therefore entails aligning mandatory legislative instruments, such as the EU AI Act, with voluntary and structured risk management standards. This approach is reflected in unification models such as that proposed by Dotan (2024), which offer organisations clearer pathways for coherent and responsive AI governance.

2.9.3 Dialogic Accountability and Inclusivity

The complexity inherent in AI systems necessitates a transition from traditional, individualistic leadership models to those that promote collective action and distributed responsibility. Dialogic Accountability utilises the methodologies of Dialogic OD, which are inherently suited to addressing complex adaptive challenges by emphasizing conversation, participation, and shared meaning-making. This approach mitigates the risk of ambiguity in implementation by prioritizing transparent communication, psychological safety, and leadership vulnerability, thereby fostering collective self-reflection and ownership. Moreover, by engaging diverse stakeholders through dialogue, leaders can address systemic inequities and ensure that AI design and deployment are inclusive and culturally relevant, thus mitigating the risk of ethical imperialism identified in Western-centric ethical traditions.

2.10 Discussion and Conclusion

The literature reviewed highlights the dualistic nature of AI adoption, emphasizing both its transformative potential and the significant risks it poses to organizations and society. On one hand, AI facilitates innovation, operational efficiency, and economic growth, with projections estimating its contribution to global GDP as

substantial by 2030. However, this optimistic perspective necessitates a critical examination of the ethical, practical, and regulatory challenges that accompany AI integration.

The commercial opportunities presented by AI, particularly through automation and enhanced data analytics, are unprecedented. The potential to optimize operations, personalize customer interactions, and facilitate innovation is widely recognized. However, the implementation of AI is not without its complexities. A recurring pattern across the literature highlights concerns such as algorithmic bias, privacy infringements, and the potential diminution of human agency. These risks are exacerbated by a lack of standardized frameworks for evaluating AI performance and ensuring transparency, thereby rendering organizations susceptible to ethical breaches and operational inefficiencies.

Regulation emerges as a fundamental element of the discourse, with frameworks such as the EU AI Act offering a structured, albeit controversial, approach to governance. While proponents assert that regulation can promote ethical compliance and market stability, critics caution against the potential inflexibility of these measures, particularly in light of AI's rapid evolution. This tension between innovation and oversight reflects broader concerns regarding the balance between progress and societal well-being. The literature reveals an urgent need for adaptive governance models that can accommodate technological advancements without impeding innovation.

Ethical considerations are central to the discourse surrounding AI. The principles of beneficence, non-maleficence, autonomy, justice, and explicability provide a robust foundation for ethical AI deployment. However, their translation into actionable practices remains inconsistent across industries and regions. The prevalent occurrence of ‘ethics-washing’ in corporate contexts undermines genuine efforts to establish trust and

accountability. The Cambridge Analytica scandal serves as a salient example of the consequences of neglecting ethical imperatives, highlighting the interplay between data governance, regulatory compliance, and public trust.

Critically, the literature emphasizes the significance of interdisciplinary and cross-sectoral collaboration in addressing the challenges posed by AI. Stakeholders, including policymakers, technologists, and business leaders, must engage in ongoing dialogue to refine regulatory frameworks and ethical guidelines. Such collaboration is essential to ensuring that AI's benefits are equitably distributed and its risks effectively mitigated.

In conclusion, the adoption of AI represents a complex interplay of opportunities and risks. While its potential to revolutionize industries is undeniable, this transformation necessitates a prudent approach to governance and ethics, accompanied by appropriate regulatory measures.

As AI technologies continue to evolve, future research must focus on bridging the gap between theoretical principles and practical implementation. By fostering transparency, accountability, and adaptability, organizations can navigate the challenges of AI integration, ensuring that innovation aligns with societal values and public trust. This equilibrium will be crucial in shaping an AI-enabled future that is both equitable and sustainable.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Introduction and Research Objectives

This chapter describes the research methodology formulated to fulfil the study's aims and objectives. It commences with an exposition of the philosophical foundations underpinning the research and their relevance to the selected phenomenon. Subsequently, the philosophical stance is conveyed, accompanied by a justification of the chosen methods. The chapter further elaborates on the data collection instruments and analysis procedures, emphasizing their strengths and limitations. Ethical considerations and methodological constraints are examined in the concluding section.

This study was guided by the following objectives:

1. To identify and analyse external regulatory influences that impacted the ethical implementation of AI in the commercial sector.
2. To assess ethical concerns and potential risks associated with data privacy, algorithmic bias, and transparency in the deployment of AI systems. This encompassed both the theoretical comprehension of ethics in data science and its practical application within routine business operations. The objective was to identify the discrepancies between ethical principles and their practical implementation, as well as to comprehend the challenges organizations encounter in adhering to ethical standards.
3. To evaluate internal organisational practices and preparedness for ethical AI deployment, including governance, employee education, and leadership commitment. This objective also examined the extent of organisational understanding of AI-related risks and the potential legal and reputational consequences of ethical and regulatory breaches. Measures such as internal

training programs and ethical guidelines were assessed for their role in promoting responsible AI usage.

4. To develop a comprehensive framework that promoted transparency, accountability, and fairness in AI applications. The framework aimed to synthesise themes from the research findings to support best practices in ethical data management and responsible AI deployment. It was intended to reflect the real concerns and priorities of stakeholders and to enhance understanding of AI risk perceptions and ethical considerations among businesses, policymakers, and researchers.

3.2 Research Philosophy and Positioning

The development of a research methodology begins with ontology, which examines the nature of reality and what can be known about the human world (Anfara & Mertz, 2014). This study adopted an interpretivist philosophical stance, positing that reality is socially constructed and is best comprehended through the meanings individuals ascribe to their experiences (Bryman & Bell, 2015). One of the primary strengths of interpretivism lies in its capacity to explain the complexities of human experience.

Interpretivism is a prevalent research paradigm, particularly within the domain of information systems (Walsham, 1995). Epistemology pertains to what is considered acceptable knowledge within a field of study (Saunders et al., 2019), and interpretivism, from an epistemological perspective, is acknowledged for its ability to explain causal relationships by concentrating on how individuals comprehend and interpret their actions and interpretations within a research context (Matta, 2015).

Acknowledging the subjectively lived experiences of individuals is deemed crucial in rapidly evolving fields such as AI, which may be susceptible to 'groupthink' or unduly influenced by excessively positive or negative media portrayals. This domain is

particularly vulnerable to narratives propagated by prominent influencers and entrepreneurs, which reinforce their worldviews and are not easily contested due to the swiftly changing nature of this disruptive technology. Essentially, opinions can only be confirmed or refuted retrospectively, with the benefit of hindsight at a future point in time.

An interpretive study was utilized as a research methodology to fulfil the research objectives and attain a comprehensive understanding of how ethical practices and risks associated with AI adoption were perceived by individuals and organizations. This approach is grounded in interpretivism, a research philosophy that seeks to comprehend and interpret the subjective experiences, meanings, and perspectives of individuals and groups.

Interpretivism is particularly well-suited for investigating complex and multifaceted topics, such as perceptions of ethics and risks, as it required the researcher to adopt an empathetic approach (Saunders et al., 2019). This approach provided valuable insights into the subjective experiences and contextual factors that shaped the perception of AI risks and the role of ethics in mitigating those risks.

There were several reasons why an interpretive approach was advantageous for this study. First, it was important to acknowledge that ethical behaviour and risk perception were inherently subjective and significantly influenced by the values, beliefs, and contextual factors of both individuals and organisations. Engaging in an interpretive study facilitated a more profound examination of these subjective experiences and illuminated the intricate perceptions of these critical aspects concerning the implementation of AI. The researcher's professional background in Information Technology contributed significantly to this endeavour.

Secondly, the importance of context is a fundamental principle of interpretive research. This approach allows for the examination of how organizational culture, industry, and external factors influence the perception of ethics and risks. A comprehensive understanding of the context is crucial for the development of effective risk management strategies, the cultivation of essential competencies, the development of AI literacy and the fostering of a culture that upholds ethical practices.

Despite its strengths, interpretivism is not without limitations. Critics frequently highlight issues of subjectivity, and the challenges associated with generalizability. Given that interpretivist research relies heavily on the researcher's interpretation of data, there exists a potential for bias to influence the findings (Bonache, 2020; Alborough & Hansen, 2012). For instance, Kamau emphasizes the necessity for pragmatism in research methodologies, suggesting that while interpretivism provides valuable insights, it should be complemented by alternative approaches to enhance the robustness of findings (Kamau, 2022).

In alignment with the study's interpretive philosophical stance, which asserts that reality is socially constructed, the analysis of the findings must acknowledge the researcher's role in the interpretive process. This research adopts a practitioner-scholar approach typical of doctoral work in Business Administration (DBA), where the analysis of complex organizational phenomena is informed by professional experience.

The interpretation of the data, particularly the significant gap between aspirational rhetoric and institutional readiness identified in the findings, is informed by the researcher’s background in Information Technology and organizational leadership. This professional context provided essential Systemic Change Literacy and deep knowledge of AI systems, ethics, governance, and organizational transformation, which was critical in

selecting, framing, and interpreting the survey items that assess Responsible AI competencies.

While this expertise enhanced the ability to design an instrument relevant to industry challenges, the practice of reflexivity was actively maintained to mitigate the potential for researcher bias. For example, when interpreting the high self-reported scores on the Responsible AI Literacy Index (mean 4.02, median 4.17) alongside the low scores for knowledge of AI governance models (mean 3.07) and training adequacy (mean 3.11), the researcher was conscious of not simply accepting the high self-assessment as fact. Instead, the analysis focused on the discrepancy between perceived literacy and practical competency, leveraging the interpretive lens to suggest that self-assessed comprehension may not correspond with actual, practical knowledge, which requires further objective assessment.

This reflexive approach ensures that the interpretation of the quantitative patterns, such as the widespread demand for training and the low scores for governance clarity, is grounded in the objective data while simultaneously offering contextually rich organizational insight, thereby enhancing the credibility of the findings for organizational leaders.

While interpretivism is traditionally aligned with qualitative methods, its application within this research is justified through a mixed methods approach. The use of quantitative data is not intended to generalise findings universally but to surface patterns of perception that are then interpreted within their organisational and cultural contexts.

The survey instrument, though comprising primarily Likert-scale items, was designed to capture subjective viewpoints on risk awareness, responsible AI, governance, and leadership readiness. These perspectives, while quantified for the purposes of pattern

recognition, are understood to be socially and situationally constructed. Thus, the role of interpretivism in this study is to provide a contextual lens through which numerical trends are examined, not as objective truths, but as expressions of shared meaning within a professional community.

Furthermore, the interpretivist approach facilitates the analysis of statistical trends, particularly concerning leadership behaviours, ethical comprehension, and organizational culture. This method enables a more profound understanding of the underlying reasons behind observed phenomena, aligning with the broader objective of the study to inform leadership frameworks for the responsible implementation of AI in commercial organizations.

In this context, the application of interpretivism enhances the quantitative approach by promoting reflexivity and contextual awareness. It recognizes that the importance of numerical responses is not solely in their frequency but in how they represent the evolving narratives and values inherent in contemporary leadership practices.

3.3 Research Design and Approach

This study utilized a sequential explanatory mixed-methods approach. Initially, a quantitative survey was administered to gather insights from a diverse range of organizations across various sectors and countries concerning their perceptions of AI risks and ethical practices. This was followed by a qualitative case study phase, which provided context-specific information on AI implementation within organizations. This dual approach explicated the complex ways in which AI risks were perceived and managed, alongside the ethical considerations involved and the degree of preparedness for developing, deploying, and using AI models responsibly. A second survey was employed, informed by the preceding two stages, to assess the readiness of organizations,

the implementation of responsible AI practices, and the extent to which executive leaders are equipped to navigate the evolving AI landscape.

The survey questions were meticulously designed to elicit critical perspectives, events, and perceptions that aligned with the research objectives. Each question was crafted based on existing literature and case studies, with each corresponding to one or more specific research objectives.

The qualitative phase employed a desk-based case study methodology to investigate detailed and context-specific information regarding organizations implementing AI. This real-world examination provided valuable insights into the complexities associated with AI deployment, highlighting both collective and individual attitudes towards ethics and risk within these organizations. The knowledge obtained served as a robust foundation for understanding current practices, identifying potential shortcomings, and informing the development of a model for ethical AI deployment in the future.

The application of an interpretive approach was instrumental in capturing a multitude of perspectives, thereby illuminating both similarities and differences. Moreover, it facilitated the discovery of latent and often overlooked hazards associated with AI deployment, particularly those related to trust, ethics, and social implications.

Remaining current in a rapidly evolving field presents a persistent challenge, as new research, data, and scholarly articles emerge almost daily. Previously accepted knowledge is frequently supplanted by fresh insights, novel theoretical perspectives, and newly uncovered evidence. This dynamic landscape necessitates continuous engagement with the latest developments to ensure relevance and rigour. Furthermore, the proliferation of speculation within mainstream media complicates this task. Much of the coverage is characterized by bias and sensationalism, underscoring the need for critical

scrutiny and careful evaluation of sources. Distinguishing credible information from conjecture is essential to maintaining scholarly integrity and analytical depth.

A critical area of investigation was regulation and governance, which required substantial effort and extensive desk-based research. Much of the literature in this domain was characterized by esoteric terminology, dense legal language, and region-specific nuances, posing challenges to interpretation and comparative analysis.

As the research advanced, the focus underwent an evolution. Despite growing attention to AI regulation, governance, and ethics, the literature reveals a relative paucity of analysis on how organisations can adapt in practice to realise AI's benefits while responsibly managing its associated risks.

This gap underscores the necessity for a more pragmatic, organizational-level perspective within the broader regulatory and ethical discourse. It also highlighted the critical significance of leadership competence, behaviour, and AI literacy in the development, deployment, and utilization of AI models.

It was imperative to develop a meticulously designed survey instrument featuring clear and unbiased questions that aligned with the research objectives. The questionnaire was required to be 'unambiguous, reliable, and valid for the purpose for which it was to be used' (Remenyi et al., 2010). Conducting a pre-test with a small sample size facilitated the identification and rectification of potential issues prior to full deployment. Implementing a robust sampling strategy was crucial to ensure that the sample accurately represented the target population, employing techniques such as stratified sampling or random sampling, which were appropriate for the research design. Given the international scope of the survey, ethical considerations were addressed, including obtaining informed consent, ensuring confidentiality, and adhering to data protection regulations across different jurisdictions.

The research methodology utilized a sequential explanatory mixed-methods approach, offering a comprehensive understanding of the risks, ethical implications and organizational readiness concerning the development, deployment, and utilization of AI models in the context of AI implementation within organizations.

This approach integrated qualitative and quantitative data to enhance theoretical understanding and offer practical guidance for promoting responsible AI practices across various business contexts.

3.4 Qualitative Phase: Contextualisation and Analysis

The transition from the initial exploratory survey (Survey #1), which identified key themes related to ethical attitudes, practices, and broad perceptions of AI, to the targeted instrument used in Survey #2 was facilitated by a focused desk-based case study analysis (Phase 2). This phase was crucial for contextualizing the practical challenges organizations face when attempting to operationalize the strong ethical consensus observed in the initial data.

The qualitative phase utilized a desk-based methodology to gather context-specific information on organizational implementation, focusing on entities navigating complex AI governance and ethical environments. The case selection criteria were focused on organizations that had made public commitments to Responsible AI, experienced notable AI-related incidents, or operated in highly regulated sectors (e.g., Financial Services and Information Technology, the sectors highly represented in the sample).

Data sources for this phase included:

1. Publicly Available Organizational Documents: Corporate strategy statements regarding AI implementation, ethics codes, annual reports concerning technological risk, and published governance frameworks.

2. Regulatory and Compliance Documents: Analysis of organizational responses to evolving legislative pressures, such as the EU AI Act, to understand how abstract rules were being integrated into internal protocols.
3. Incident Reports and Public Analysis: Desk review of high-profile cases involving algorithmic bias, data misuse, or sanctions, which provided evidence of the 'accountability gap' in real-world application.

The knowledge obtained from these sources was systematically analysed to identify the specific barriers to implementation and the required leadership capabilities, thereby informing (alongside results from the first survey) the instrument for Survey #2. Content and thematic analysis was applied to the collected documents to extract structured insights relating to four key areas:

1. Clarity of Responsible AI Strategy: Identifying tangible evidence of clear organizational strategies and guidance for legal compliance, an area later rated critically low in Survey #2.
2. Operational Governance Deficits: Documenting instances where knowledge of governance models (e.g. NIST AI Risk Management Framework) appeared insufficient for organizational deployment, confirming the low self-assessed knowledge scores later found in Survey #2.
3. Training and Education Gaps: Extracting qualitative context regarding the specific content and format of existing leadership education and training programs to validate the strong consensus observed in Survey #1 regarding the need for ethical training.
4. Leadership Engagement: Assessing how organizational rhetoric translated into specific roles and responsibilities related to AI integration and deployment, particularly in international organizations.

By systematically translating the qualitative insights regarding implementation barriers into quantifiable measures, the case study phase ensured that the subsequent large-scale quantitative assessment (Survey #2) directly targeted the precise leadership competencies and organizational deficits required for effective Responsible AI implementation.

3.5 Quantitative Phase: Surveys and Data Collection

3.5.1 Quantitative Surveys

Quantitative surveys offer several methodological advantages in academic research, including efficient data collection from large populations, systematic and objective analysis, and the capacity to test hypotheses through the examination of relationships between variables. A key strength lies in the generalisability of findings, as large sample sizes within the positivist paradigm enable inferences to be extended beyond the immediate study population (Park et al., 2020). This capacity supports the development of broader theoretical insights and practical applications, reinforcing the value of quantitative surveys within the academic research landscape.

Quantitative surveys are widely employed in academic research due to their capacity to enhance the reliability and validity of findings. A key advantage lies in their efficiency and cost effectiveness in data collection, enabling systematic examination of defined populations through structured instruments that support statistical analysis (Tanti et al., 2021). Such surveys allow researchers to investigate beliefs, attitudes, and behaviours at scale, producing data that can be analysed objectively using established statistical methods (Simarmata et al., 2022). Their prevalence within information systems research further reflects their suitability for capturing data from large and diverse populations within constrained timeframes and resources (Sawah et al., 2008). In addition, the replicability of quantitative survey designs strengthens methodological

rigour and enhances the credibility of research outcomes (Creswell and Plano Clark, 2018; Almeida, 2017).

The generalisability of findings derived from quantitative surveys represents a key methodological strength. Within the positivist paradigm, large sample sizes enable the generation of inferences that extend beyond the immediate study population, thereby supporting broader theoretical development and practical application (Park et al., 2020).

Quantitative surveys also offer a systematic and objective approach to data collection and analysis. By transforming numerical data into structured insights, quantitative methods facilitate hypothesis testing and reduce the influence of subjective interpretation (Almeida, 2017). This objectivity enhances the reliability and consistency of findings, as quantitative approaches are designed to describe phenomena in a standardised and replicable manner (Suprapmanto, 2024).

Quantitative surveys facilitate the exploration of relationships between variables, which is essential for hypothesis testing. As noted by Peddle et al., the quantitative phase of research allows for the examination of correlations among variables, providing insights that can inform future studies (Peddle et al., 2020).

The approach was to secure a substantial, highly relevant, sample size to enhance the statistical power and yield more reliable and generalizable results. By incorporating participants from a diverse array of organizations, sectors, and countries, the study captured a wide range of perspectives and experiences, thereby improving the quality of the data. The inclusion of multiple sectors and countries facilitated comparative analysis, enabling the identification of patterns and variations across different contexts. This approach significantly augmented the research value by allowing for the exploration of how variables interact differently in various environments. Surveying a diverse and extensive sample increased the likelihood that the findings would be applicable to a

broader population, thus enhancing the study's external validity. This aspect is particularly crucial in business research, which often seeks to provide guidance across different settings. A large and varied dataset permitted the application of advanced statistical techniques, yielding deeper insights into the complex relationships between variables.

Both surveys were developed utilizing Google Forms, which offered a familiar, user-friendly and widely recognized platform for participants to access via web browsers or mobile devices. The responses were automatically collected in real-time, resulting in a single comprehensive file of all responses, which could be downloaded in comma-separated values (CSV) format.

To foster consistency in responses, the survey instrument included an introductory section defining key terms and concepts relevant to the study, including ethics, AI, governance, regulation, and responsible AI. This approach was intended to establish a shared interpretive framework and minimise variation arising from differing understandings among participants.

The quantitative study employed a non-probability sampling strategy, combining convenience sampling, whereby participants were selected based on accessibility and willingness to participate, with purposive sampling to ensure inclusion of individuals possessing characteristics relevant to the research objectives.

To encourage candid and accurate responses, the principles of confidentiality and anonymity were clearly communicated to all participants. Respondents were informed that no identifying or demographic data would be collected or stored, thereby reducing potential response bias and supporting open and honest participation.

3.5.2 Qualitative Data Sources and Analysis Strategy

The study utilized a sequential explanatory mixed-methods approach, incorporating a desk-based qualitative case study methodology to furnish context-specific insights into AI implementation within organisations. This qualitative phase functioned as a crucial methodological bridge, adopting an interpretivist perspective to explore the meaning-making processes and practical challenges encountered by organizations, thereby complementing the pattern recognition obtained from the quantitative surveys.

3.5.3 Data Sources and Scope

The desk-based case study focused on organizations highly engaged in AI development, deployment, or usage within the commercial sector, particularly those operating in regions or industries with evolving regulatory pressures. This methodology was instrumental in illuminating the complexities associated with AI deployment and highlighting collective attitudes towards ethics and risk.

The data sources reviewed were publicly available organizational and regulatory documents, providing real-world evidence of implementation attempts and governance deficits. These included:

1. **Organizational Policy Documents:** Published statements, white papers, and corporate reports detailing internal AI strategy, ethical guidelines, and formalized governance frameworks. These sources provided evidence of organizational commitment and the strategies employed to manage risks related to data privacy, algorithmic bias, and transparency.
2. **Regulatory and Compliance Documents:** Analysis of how organizations responded to external regulatory influences, such as the General Data Protection Regulation (European Commission, 2018), and frameworks related to AI

governance. This scrutiny addressed the research objective of identifying and analysing external regulatory influences.

3. **High-Profile Case Studies and Incident Analysis:** Desk review of publicly reported organizational events involving AI failures, data breaches, or sanctions. These cases provided critical contextual evidence of the discrepancies between ethical principles and their practical implementation, offering insights into latent hazards associated with AI deployment, particularly those related to trust and social implications.

3.5.4 Qualitative Analysis Strategy

The analysis applied to these desk-based sources utilised content and thematic analysis, which aligns with the interpretivist need to comprehend the subjective experiences and contextual factors shaping the perception of AI risks and ethics. The primary objective was not hypothesis testing but rather the extraction of systematic, context-specific insights regarding organizational preparedness and leadership requirements.

The analysis focused on four critical areas derived from the research objectives:

1. **Clarity of Responsible AI Strategy:** The documents were analysed to confirm the extent of formalisation and communication of ethical guidelines and risk management strategies, assessing the organization's understanding of AI-related risks and potential consequences.
2. **Operational Governance Deficits:** Content analysis was used to identify explicit references to specific governance models (e.g. formal risk frameworks) and their integration into routine business operations, thereby informing the assessment of organizational preparedness.

3. **Training and Education Gaps:** The review examined publicly reported evidence regarding internal training programs and leadership commitment to employee education on ethical guidelines, assessing their role in promoting responsible AI usage.
4. **Leadership and Competency Requirements:** The language and tone of executive-level documents were scrutinized to identify priorities and demonstrated commitment to developing the AI literacy and essential competencies required to uphold ethical practices.

By systematically translating these qualitative insights into specific operational themes and practical barriers, the desk-based case study phase ensured that the subsequent design of the second quantitative survey directly targeted the practical leadership competencies and governance deficits crucial for effective Responsible AI implementation. This triangulation of findings enhanced the credibility and trustworthiness of the overall study.

The first survey comprised a series of questions pertaining to ethical attitudes and practices in business and data usage. It included four demographic questions and 48 survey questions utilising the Likert scale to measure various aspects of the research focus. All questions were mandatory.

The primary aim of this questionnaire was to capture a diverse array of perspectives from a broad demographic base, ensuring that the findings reflect a wide spectrum of opinions and experiences. To achieve this, participant recruitment was conducted through three main channels: the SurveyCircle portal, LinkedIn, and direct email invitations. Individuals deemed appropriate within the researchers' network were provided with a flyer containing a QR code and URL to facilitate access to the survey.

People were approached directly via email and LinkedIn which together with the responses from SurveyCircle and the flyer resulted in 118 responses. This level of response was deemed sufficient as it was capable of generating meaningful independent insights, which could subsequently inform the development of questions for the second survey.

LinkedIn is a professional social networking platform that facilitates career development, recruitment, and knowledge dissemination. It allows individuals to connect with peers, expand their professional networks, pursue employment opportunities, and engage with industry-related content. The platform was selected as a primary source for recruiting participants due to the presence of numerous employed professionals in the commercial sector, which constitutes a highly relevant, key target demographic for the research. No other platform demonstrated such a high degree of relevance for the target demographic, thereby positively impacting the quality of the output. The platform also facilitated the inclusion of participants based on their sector, country, role, and tenure within their respective fields.

The platform facilitates direct communication with '1st connections,' defined as individuals who have previously accepted an invitation to connect. Consequently, the strategy involved engaging potential participants from my existing network and extending invitations to relevant individuals to become 1st connections. Upon their acceptance, they were subsequently invited to participate directly in the survey.

The second survey examined leadership and organisational readiness, the adoption of responsible AI, and leadership development in the context of advancing AI technologies. Specifically, the survey comprised a series of questions pertaining to responsible AI and AI literacy in leadership, as detailed in the appendices. It included

four demographic questions and 38 items based on the Likert scale. All questions were mandatory.

LinkedIn first connections were utilized to ensure specific and focused alignment with the target demographic, comprising executives, leaders, consultants, policymakers, and professionals engaged in AI strategy, development, or usage within private sector organizations. This included individuals responsible for overseeing or implementing AI initiatives, shaping regulatory and policy frameworks, and specializing in governance, compliance, and ethics to facilitate responsible AI adoption.

Over 1000 individuals were approached directly via email and LinkedIn. Combined with 50 responses gathered through SurveyCircle, this resulted in 197 completed surveys. All respondents answered all the questions, resulting in 7486 data items. This response level was considered sufficient to yield meaningful, independent insights and served to complement the extensive desk-based research undertaken.

3.6 Data Analysis Strategy

In quantitative academic research, analytical methods refer to the statistical techniques employed to process and interpret numerical data. These methods are applied following data collection to test hypotheses, identify patterns, explore relationships, and draw valid conclusions. The primary objective is to transform raw data into meaningful insights. Analytical methods in quantitative research encompass statistical tools used to explore, test, and interpret numerical data. They include descriptive statistics for data summarization and inferential statistics for making predictions or testing hypotheses, thereby enabling researchers to draw robust, generalizable conclusions. The selection of a method is contingent upon the research question, data type, and distribution.

This approach encompassed descriptive statistical analysis, including measures of central tendency (mean, median, and mode) and measures of dispersion (range, variance,

and standard deviation), to summarise and interpret the distribution of survey responses (Field, 2018), frequency distributions (counts and percentages of responses), and data visualization techniques (tables, histograms, pie charts).

The data files from both surveys were systematically analysed using a combination of tools, including Excel's Data Analysis functionality and the Python programming language. In certain instances, it was necessary to engage in iterative data inquiry as new insights emerged. These analytical methods facilitated the extraction of meaningful insights and trends from the dataset, which are explored in the Findings section of this document.

3.7 Ethical Considerations, Reliability and Validity

In the realm of business research, which frequently examines organizations, employees, consumers, and market behaviour, researchers are required to navigate intricate ethical landscapes characterized by commercial sensitivity, power imbalances, and potential conflicts of interest (Saunders, Lewis, and Thornhill, 2019).

Ethical considerations were integral to the design and implementation of this research, reflecting the complexities inherent in business research involving organizations, employees, and sensitive commercial data (Saunders, Lewis & Thornhill, 2019). All potential participants were made aware of the study's aims, objectives, and research procedures (Saunders & Lewis, 2014). This ensured that participants were fully informed prior to their engagement, in accordance with the principle of informed consent (Bryman & Bell, 2015). Participants were informed of their right to withdraw from the study at any time without consequence.

Anonymity was maintained throughout the data collection process, and no identifying information was collected. This approach mitigated the risk of potential data

breaches and fostered participant trust and willingness to engage with the research (Singh, 2010).

This study was meticulously designed and conducted in accordance with established ethical standards for business research. The questions were crafted to be non-intrusive, ensuring that participants were not required to disclose sensitive company information. Neutral and non-leading language was employed to facilitate open and unbiased responses and participants were informed about the intended use of the findings and were given the option to receive a summary of the results. Furthermore, participants were assured that their responses would be used exclusively for academic purposes and reported in aggregate form (Singh, 2010). Additionally, participants were informed of how the findings would be used and offered the option to receive a summary of results. These measures ensured transparency, safeguarded participant autonomy, and upheld the integrity of the research.

Given the proprietary nature of business information, confidentiality was stringently upheld, with all data managed in accordance with the General Data Protection Regulation (GDPR) (European Commission, 2018). Survey responses were securely stored within encrypted, password-protected digital environments, thereby ensuring both data security and participant privacy. Crucially, the research was conducted independently of any commercial or organizational affiliations that might have compromised objectivity or introduced conflicts of interest (Zikmund et al., 2013). By maintaining transparency, safeguarding data, and upholding participant autonomy, the study adhered to the ethical standards typically expected in business research. Furthermore, both surveys were designed to be easily replicable, thereby enhancing the transparency and integrity of the research process.

Ensuring the reliability and validity of this study was a critical consideration throughout the research process, particularly given the mixed methods design. Mixed methods research presents specific challenges, as it must meet the quality standards of both qualitative and quantitative paradigms while achieving meaningful integration (Tashakkori & Teddlie, 2010). To support construct validity, survey instruments were developed using established practices, including consistent use of Likert scales and clear instructions to enhance respondent comprehension (Saunders, Lewis & Thornhill, 2019). Each survey underwent a pilot test, which proved invaluable in refining question clarity, eliminating redundancy, and improving overall coherence. All survey questions were made mandatory, ensuring complete datasets for analysis and reducing the risk of missing values that could compromise internal consistency and inferential validity.

Google Forms was chosen as the survey platform due to its standardized format and secure data handling processes. The system ensured consistency in survey administration and automatically logged each response into a flat CSV file. This format facilitated structured data management while preserving participant confidentiality. The data remained encrypted and password-protected within the Google ecosystem until the survey concluded, after which it was securely transferred to Apple iCloud for continued encrypted storage.

Methodological triangulation underpinned the research design, integrating qualitative insights with quantitative measures to enhance inference quality and contextual understanding (Creswell & Plano Clark, 2018; Bryman & Bell, 2015). By synthesizing findings across methods, the study aimed to ensure coherence and depth in its interpretations, minimizing the risk of superficial convergence and supporting the trustworthiness and practical relevance of the results (Easterby-Smith, Thorpe & Jackson, 2021).

This study adopted an interpretivist philosophical stance, implemented through a sequential explanatory mixed-methods design. Consistent with this paradigm, the research moved beyond objective replication and statistical generalizability to focus on understanding the complex meanings individuals assign to social phenomena within specific contexts. Consequently, traditional metrics of reliability and validity were reconceptualised as trustworthiness, credibility, authenticity, and interpretive coherence (Lincoln & Guba, 1985).

To ensure credibility, the study maintained a strict alignment between research objectives and survey instruments, covering core constructs such as RAI leadership, governance awareness, ethical risk, and organisational preparedness. Credibility was further reinforced through methodological triangulation, where qualitative desk-based analysis was used to contextualise quantitative survey findings, ensuring depth and coherence.

Trustworthiness was bolstered by transparently documenting the research process and integrating data from diverse organisational and regulatory sources. Furthermore, the researcher engaged in reflexive awareness regarding their professional background and influence on data interpretation. This approach ensured that the analysis remained grounded in empirical patterns rather than normative assumptions. Collectively, these strategies captured the rich, subjective dimensions of organisational experience, providing robust, contextually grounded findings suitable for developing the TRAIL framework.

3.8 Limitations and Chapter Summary

Methodological limitations are intrinsic to all forms of academic research, representing constraints that impact the scope, reliability, generalizability, or interpretability of findings. In the context of business research, which frequently aims to

investigate dynamic, context-specific, and multifactorial phenomena, these limitations are particularly pronounced and necessitate careful consideration to ensure transparency and rigor (Saunders, Lewis, and Thornhill, 2019).

Methodological choices inherently present such limitations. Quantitative approaches are often critiqued for their lack of contextual depth, whereas qualitative methods may be perceived as less replicable or systematic (Collis and Hussey, 2014). Mixed methods designs endeavour to address these limitations but introduce complexities in integration and analysis that may compromise coherence if not executed proficiently (Tashakkori and Teddlie, 2010). Additionally, temporal and contextual constraints are particularly pertinent in business research. Numerous studies are cross-sectional, providing merely a temporal snapshot, which restricts the ability to infer causality or capture evolving organizational dynamics (Robson and McCartan, 2016). Rapid technological, economic, or regulatory changes can swiftly render findings obsolete, especially in fields such as AI, innovation, or global supply chains.

As with all academic research, this study was subject to several methodological limitations that may influence the scope, generalisability, and interpretability of its findings. Recognising and transparently addressing these limitations is essential for upholding the rigour, credibility, and reproducibility of the research (Saunders, Lewis and Thornhill, 2019).

3.8.1 Sampling, Access, and Contextual Constraints

One common limitation is related to sampling strategy and sample size. In survey research, response bias and low participation rates can distort results and reduce confidence in their applicability to the broader business context (Hair et al., 2020).

This study employed purposive sampling to target participants with relevant experience and insights into AI ethics and organisational decision-making. While this

approach allowed for depth and relevance, it may have restricted the extent to which findings can be generalised to the broader business population (Hair et al., 2020). Furthermore, the relatively restricted sample size, constrained by access and time limitations, may have impacted the statistical representativeness of the quantitative component.

Future research employing larger, randomised samples across a wider range of industries and geographical contexts could enhance external validity and further assess the generalisability of the study's conclusions. Additional value would be gained from deeper examination of cultural attitudes towards governance, regulation, and ethics, particularly in relation to responsible AI adoption.

Access to participants and organisational data posed another constraint. As noted by Bryman and Bell (2015), business research often involves commercial sensitivities and gatekeeping, which can restrict access to candid data or limit the range of participating firms. While efforts were made to gather diverse perspectives, the scope of access ultimately shaped the selection of cases and the depth of qualitative insights.

3.8.2 Methodological and Design Trade-Offs

The research was conducted within a cross-sectional timeframe, offering a snapshot of participants' views and practices at a specific point in time. While this design enabled efficient data collection, it limited the ability to capture evolving organisational dynamics or infer causality (Robson and McCartan, 2016). Longitudinal research would be better positioned to track changes in attitudes and practices related to responsible AI deployment over time, particularly given the rapidly changing technological and regulatory environment.

Given the interpretivist stance adopted in this study, the role of the researcher was acknowledged as an active instrument in the research process. While this enabled a

deeper exploration of meaning and context, it also introduced the potential for bias in data interpretation. As Collis and Hussey (2014) highlight, qualitative data is particularly vulnerable to subjectivity stemming from the researcher's worldview and prior assumptions. To mitigate this, reflexivity was actively practised throughout the study, and the use of triangulation between qualitative and quantitative findings helped enhance credibility and minimise interpretive bias.

Although the survey instruments were carefully designed and piloted to ensure clarity and flow, the measurement of abstract constructs, such as ethical awareness, trust, or AI risk perception, remained a challenge. As these concepts are inherently complex and context-dependent, there may have been limitations in construct validity (Saunders and Lewis, 2014). While the use of Likert scales and clear instructions supported consistency, further validation of instruments using established scales or psychometric testing could enhance measurement reliability in future research.

The study employed a mixed methods design to explore both the prevalence of attitudes (quantitative) and the reasoning behind them (qualitative). While this approach offered valuable complementarity, it also involved trade-offs. For instance, some aspects of organisational culture and ethics may have required deeper ethnographic immersion than was feasible within the research design. As Bryman and Bell (2015) note, no single method can capture the full complexity of organisational phenomena; thus, methodological decisions must balance depth and breadth.

Tashakkori and Teddlie (2010) highlight that coherent synthesis in mixed methods research requires careful alignment of epistemological assumptions and analytical procedures. Although this study used sequential triangulation and convergence to strengthen inference quality, the integration may not have fully captured all nuances

across methods. Future research may benefit from more elaborate mixed methods frameworks, including parallel data collection or joint displays for enhanced integration.

Organisational responses to responsible AI can be shaped by local regulatory frameworks, cultural values, and sector-specific norms. While the research context was clearly defined, readers should exercise caution in generalising conclusions to other domains. Comparative studies across sectors or international settings could provide more comprehensive insights.

3.8.3 Temporal, Ethical, and Resource Limitations

Ethical considerations and commercial sensitivities may influence the decision not to collect certain data items. The sensitivity surrounding areas such as internal practices, intellectual property, and failures was duly respected. However, as Zikmund et al. (2013) note, such limitations can affect the richness and transparency of business research, particularly when reputational concerns lead to cautious disclosure. Nonetheless, ethical considerations were prioritised to ensure participant trust and compliance with GDPR (European Commission, 2018).

The research was ultimately constrained by the practical limitations inherent in doctoral studies, specifically, the restricted availability of time, financial resources, and researcher capacity. These constraints influenced the scale of data collection and the breadth of analytical techniques employed. While sufficient data were gathered to address the research objectives, future studies could expand upon this foundation using larger datasets, extended timeframes, or more granular analytical tools.

This chapter offered a comprehensive exposition of the research methodology employed to achieve the study's aims and objectives, beginning with its philosophical foundations and advancing through the research design, data collection, and analytical procedures. Philosophically, the study adopted an interpretivist stance, positing that

reality is socially constructed and best comprehended through the meanings individuals attribute to their experiences. This approach is particularly apt for investigating complex, multifaceted topics such as perceptions of risks, ethics and readiness in the rapidly evolving field of AI, as it acknowledges subjective experiences and contextual factors.

The study adopted a sequential explanatory mixed methods design. The quantitative phases comprised two online surveys examining AI related risks and ethical practices across organisations operating in multiple sectors and geographical contexts. This was complemented by a desk based qualitative case study, which provided contextual insight into organisational approaches to AI implementation, governance and regulation.

The survey instruments were meticulously designed using Google Forms, featuring clear, unbiased, and mandatory Likert-scale questions, along with demographic questions, to capture a wide array of perspectives. Key terms were defined at the outset to ensure consistent understanding among participants.

The sampling method for the quantitative research was non-probability sampling, combining convenience and purposive (judgmental) sampling to target participants with relevant characteristics. Participants were recruited through the SurveyCircle portal, LinkedIn, and direct email invitations, aiming for a diverse and extensive sample to enhance statistical power and data quality. The first survey yielded 118 responses, while the second, targeting executives, leaders, and professionals involved in AI, garnered 197 completed surveys.

3.8.4 Sample Size Justification and Contextual Rigour (Survey #2)

A total of 197 professionals completed Survey #2 in its entirety, thereby providing the primary dataset for the quantitative analysis. Given the study's interpretive philosophical stance and its emphasis on generating context-specific insights for

organizational leadership (DBA objective), a non-probability sampling strategy was employed. This approach, which included purposive and convenience sampling, prioritized the relevance and seniority of participants over universal statistical generalizability.

The resulting sample size of 197 was deemed sufficient for two primary reasons:

1. **Contextual Rigour over External Validity:** The aim was to capture the informed perspective of professionals highly engaged in AI strategy, governance, and organizational change. The sample largely consisted of seasoned industry participants, with the majority possessing substantial work experience, and significant representation coming from pivotal, risk-aware sectors, notably Information Technology, Financial Services, and Consulting. This targeted approach ensures that the findings reflect seasoned industry perspectives on highly complex, technical subjects, even if the external validity is constrained.
2. **Sufficiency for Inferential Analysis:** The achieved sample size supports the planned analytical depth, including the use of enhanced statistical modelling techniques such as the regularised linear model and cluster analysis. These techniques require a sufficient number of observations to identify stable and meaningful signals, analyse the variation in preparedness, and prevent the risks of overfitting inherent in survey-based organizational research. The sample size was therefore adequate to move beyond descriptive patterns to identify the statistical drivers underpinning Responsible AI capability.

Accordingly, interpretation of the findings prioritised contextual trustworthiness and analytical rigour in relation to the target population of international organisational leaders, rather than broad statistical generalisation to the wider population. Descriptive statistical techniques were applied to both datasets to summarise and interpret their core

characteristics. Ethical considerations were embedded throughout the research design and implementation, and survey instruments were developed in line with established methodological standards to support reliability, validity, and credibility. The chapter also provides a transparent account of methodological limitations and identifies avenues for extending and deepening future research.

CHAPTER IV

FINDINGS

4.1 Introduction

This chapter presents the principal findings of the research as described in Chapter 3 Research Methodology, and is divided into two sections, each examining the outcomes of the two quantitative questionnaires administered during the study.

4.2 Survey 1: Ethical Attitudes and Practices

The initial quantitative survey garnered 118 responses regarding ethical attitudes, practices, and perceptions of AI, resulting in a dataset comprising 118 records. The survey data reveals several key themes and insights, highlighting areas warranting further investigation in the qualitative, desk-based phase. Additionally, it informed the development of questions for the subsequent survey.

The most common age group among respondents was 25-34. Respondents covered a range of six age groups, showing diverse participation. The gender distribution was relatively evenly split.

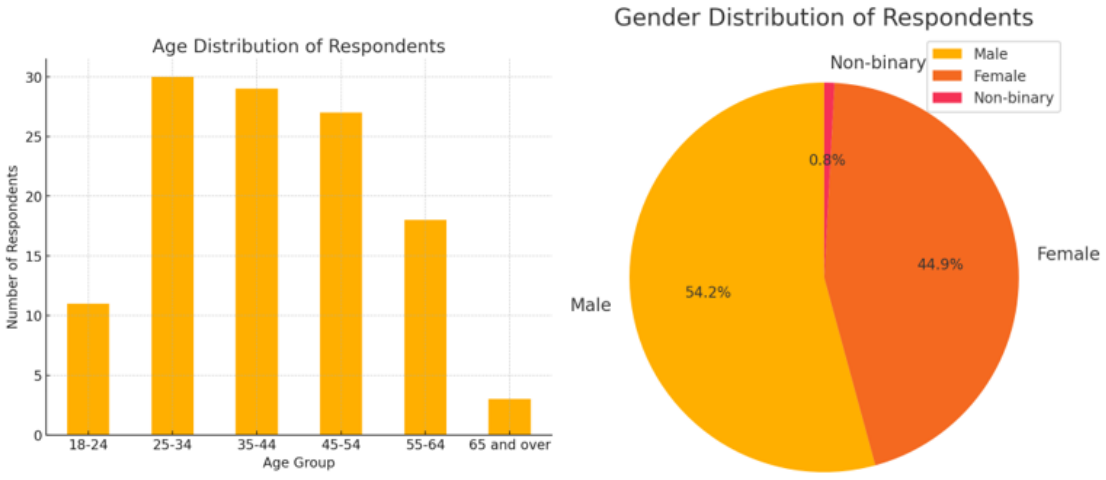


Figure 4.1: Age and Gender Distribution

Figure 4.2 shows most respondents have over 20 years of work experience, reflecting seasoned professionals. Seven distinct ranges of work experience are

represented, from less than a year to over two decades. The UK has the highest number of respondents, followed by other regions like Europe (excluding the UK), Asia, and others.

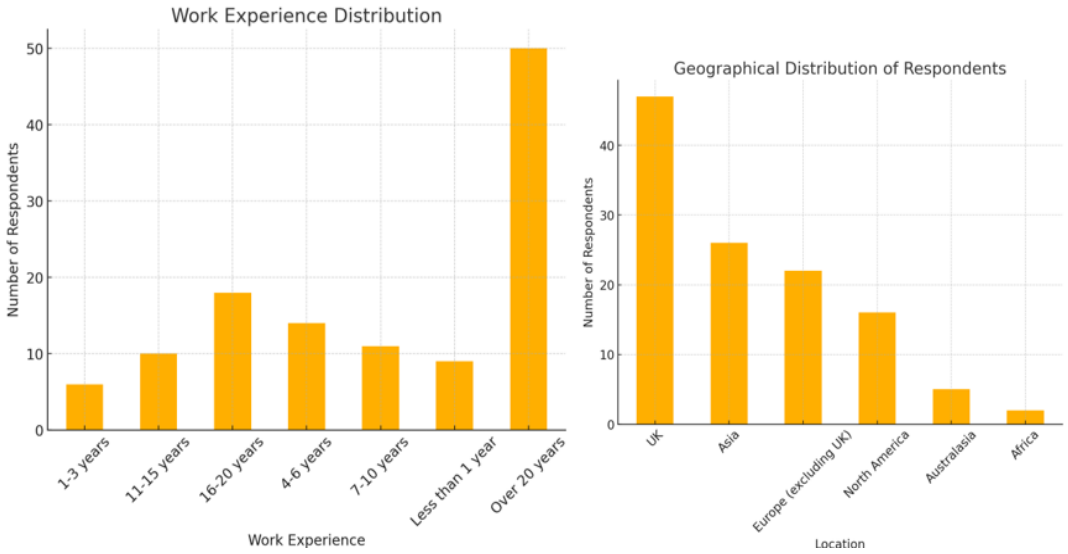


Figure 4.2: Work Experience and Geographical Distribution

The majority of participants in Figure 4.2 possess substantial work experience, indicating that the insights may be reflective of seasoned industry perspectives. A significant proportion of respondents from the UK suggests a potential emphasis on region-specific viewpoints concerning ethics and AI.

The survey results indicate a significant and widespread concern among respondents regarding the ethical management, privacy, and governance of customer data within the realm of AI. There was unanimous consensus (100%) that companies should establish stringent policies to prevent the misuse of customer data, highlighting a strong expectation for organizational accountability.

All respondents either 'Strongly agree' or 'Agree' with the statement: 'Companies should have strict policies in place to prevent misuse of customer data.'

Furthermore, 97% of respondents either 'Strongly agree' or 'Agree' with the following statements: 'It is important to ensure the security and privacy of customer data

at all times,' 'Data breaches should be promptly disclosed to affected customers,' 'Users should have the right to know how their data is being used in AI systems,' and 'Privacy protection should be a primary consideration in AI system design.'

Additionally, as Figure 4.3 shows, 81% of respondents either 'Strongly agree' or 'Agree' with the statement: 'I have concerns regarding the usage and processing of my personal data in AI systems.'

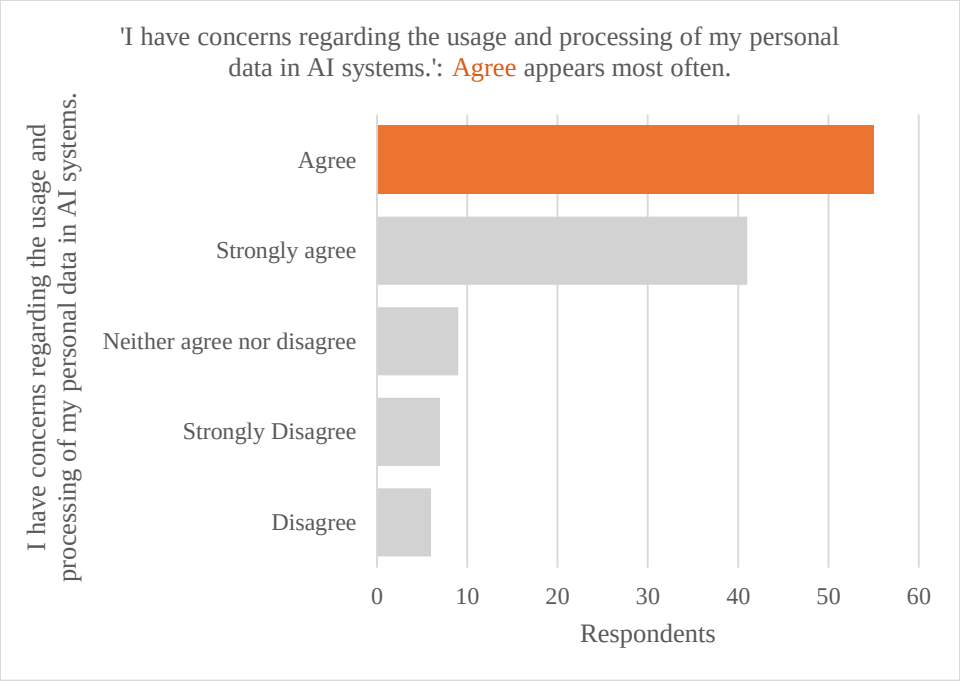


Figure 4.3: Concerns Regarding Personal Data

Transparency emerged as a near universally endorsed requirement, with 99% of respondents indicating agreement or strong agreement with the statement ‘companies should be transparent about how they collect and use customer data’. This finding reflects strong support for clear disclosure of data collection and usage practices.

A significant majority (97%) agreed or strongly agreed with statements emphasising the need for continuous data security, prompt disclosure of breaches, and

user rights to understand data usage in AI systems. Notably, the same proportion supported embedding privacy protections in AI design, indicating strong alignment with principles of privacy by design.

Moreover, 81% of respondents expressed personal concern regarding the processing of their data within AI systems, indicating a widespread sense of unease among users. This sentiment is further corroborated by the finding that 95% of participants deemed data sharing without explicit consent to be unethical, underscoring a profound expectation for informed consent and ethical management, shown in Figure 4.4 95% responded either ‘Strongly agree’ or ‘Agree’ to the question ‘Sharing customer data with third parties without explicit consent is unethical.’

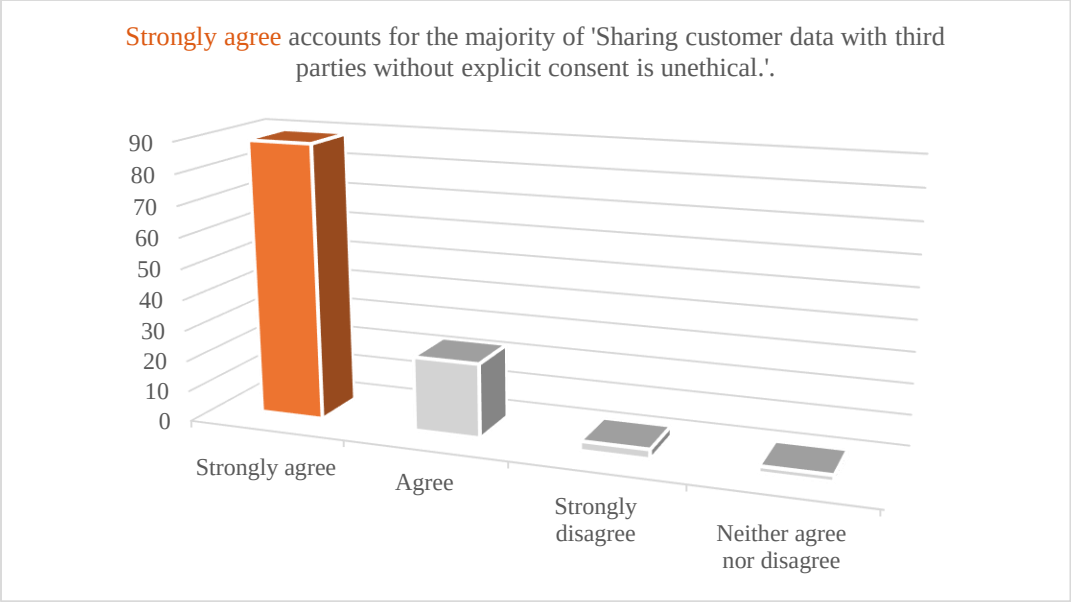


Figure 4.4: Sharing Data

Collectively, these findings underscore a strong consensus on the paramount importance of data ethics, privacy, and transparency in the development and deployment of AI systems.

There is a strong consensus on the importance of ethical conduct in both personal and professional contexts. Respondents exhibited a high level of agreement with statements such as, ‘I always endeavour to make decisions and take actions ethically.’

There exist divergent perspectives regarding the nature of ethics, specifically whether it is subjective or universal. A significant proportion of respondents, 78% indicated 'Strongly agree' or 'Agree' as shown in Figure 4.5, contend that ethical behaviour is a subjective decision influenced by individual knowledge and experience, while 22% remain uncertain.

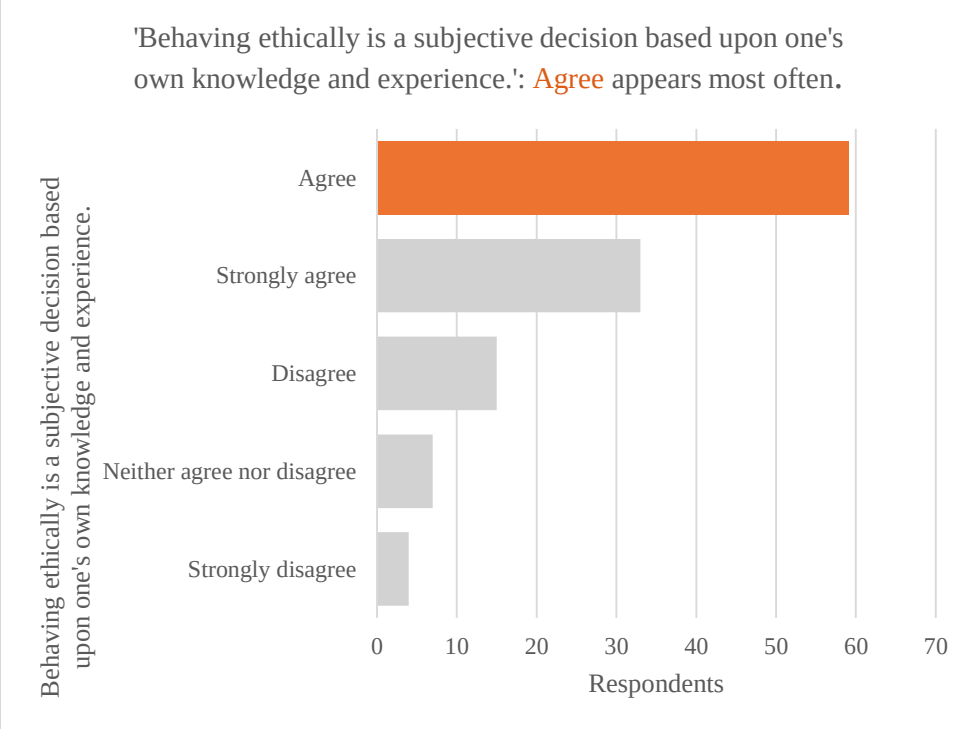


Figure 4.5: Behaving Ethically

The relationship between ethical behaviour and perceptions of AI oversight was explored. This entailed examining the correlation between individuals' ethical conduct and their views on the necessity of human oversight in sensitive AI applications. This

yielded valuable insights into how ethical principles are applied in contexts requiring rigorous AI governance.

The principles of transparency, accountability, and fairness are consistently emphasized as fundamental. A significant majority, 94% of respondents, either 'Strongly agree' or 'Agree' with the assertion that these principles should constitute the core of any ethical framework.

Respondents predominantly oppose unethical practices in the workplace, with a consensus that such practices are unacceptable and should be reported. Notably, 93% of respondents either 'Strongly agree' or 'Agree' with the statement that 'Ethical lapses by leaders should be addressed promptly and transparently.' There are nuances surrounding the reporting of policy violations that warrant further investigation.

Significant concerns exist regarding the potential for biased or inaccurate outputs from AI systems. There is substantial support for transparency, user rights pertaining to data, and privacy protections, with 94% of respondents, as shown in Figure 4.6, either 'Strongly agree' or 'Agree' with the statement that ‘Transparency about AI decision-making processes is crucial for building trust.’

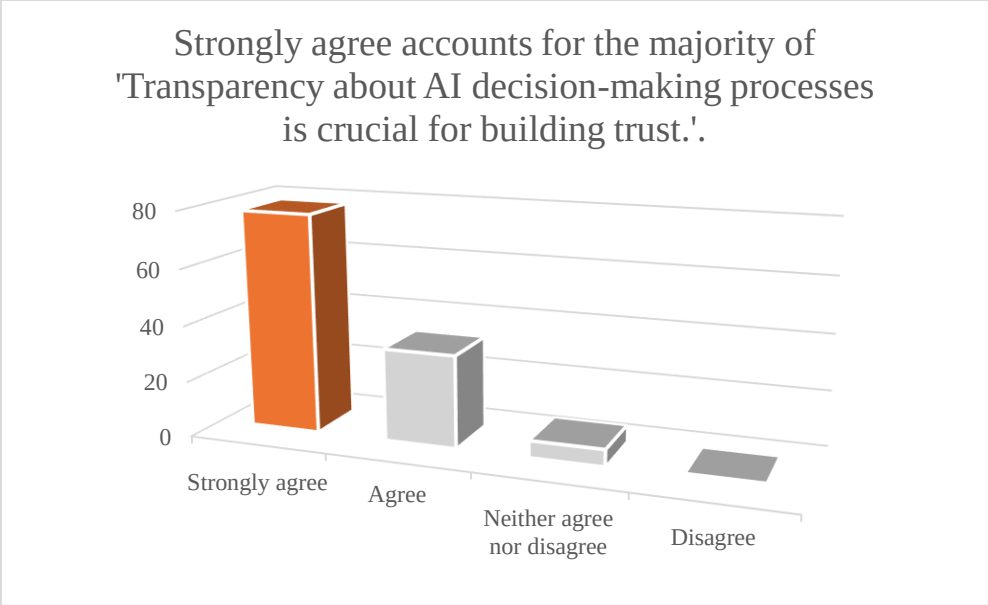


Figure 4.6: Transparency

The significance of explainability features in AI systems is also emphasized. A primary concern is the potential misuse of AI-generated content, with a high mean response of 23.4.

Responses concerning the role of ethics training were examined to assess its contribution to fostering an ethical AI culture. The significance of formalized training programs was emphasized as a foundational element for promoting ethical awareness and adherence within organizations. By exploring these dimensions, the analyses contribute to a more comprehensive understanding of how demographic factors, oversight perceptions, and training initiatives intersect to shape ethical practices in the era of AI.

There is a broad consensus on the necessity of human oversight in sensitive domains, such as politics and health, with 92% of respondents either 'Strongly agree' or 'Agree' with the assertion that 'Human oversight is necessary for AI-generated content in sensitive areas (e.g., politics, health)'. Furthermore, there is substantial support for the implementation of regular training programs on AI ethics to cultivate a more ethically oriented organizational culture.

The survey data indicates a significant awareness of ethical considerations in both conventional business practices and AI implementation. There exists a pronounced demand for transparency, accountability, and human oversight, particularly in relation to AI systems. The incorporation of these ethical considerations necessitates a structured approach encompassing training, explicit guidelines, and stakeholder engagement, with ongoing evaluation based on predetermined metrics.

A high degree of awareness and prioritization of ethical principles is evident among respondents, both in traditional business practices and in AI systems. Respondents emphasize transparency and accountability as crucial elements for fostering trust in both AI and business operations. The necessity for human oversight and appropriate labelling of AI-generated content underscores scepticism toward autonomous decision-making in critical domains. Older respondents may place greater emphasis on ethical conduct compared to younger cohorts, and individuals with more extensive work experience may exhibit increased scepticism regarding AI's potential benefits. Respondents from Europe may hold more stringent views on privacy and transparency, potentially influenced by GDPR and regional cultural norms. Those who prioritize ethical behaviour might also demonstrate stronger support for transparency, accountability, and oversight in AI systems.

4.3 Quantitative Contextualisation

The analysis of the initial quantitative survey revealed a significant and widespread awareness of ethical considerations in the implementation of AI, establishing a strong consensus on the critical importance of data ethics, transparency, and accountability. Given that this exploratory survey confirmed a high-level ethical imperative among respondents, the research necessitated an intermediate phase to examine how these principles are operationalized in practice.

Therefore, as part of the sequential explanatory mixed-methods design, a desk-based qualitative case study was conducted, utilizing publicly available organizational and regulatory documents to provide context-specific information. The primary purpose of this qualitative bridge was to examine real-world challenges, such as the accountability gap and governance deficits, to determine the practical barriers organizations face when attempting to implement Responsible AI.

The insights gathered from this analysis directly informed the instrument used in Survey #2. The focus shifted from assessing general ethical consensus (the strength identified in Survey #1) to specifically quantifying institutional readiness and operational governance. This critical methodological step ensured that the subsequent quantitative investigation targeted key organizational deficits, such as the clarity of Responsible AI strategy (later rated low at a mean of 3.30) and the provision of guidance on legal compliance (mean of 3.24), which were highlighted as practical pain points during the case study analysis. The detailed findings from this focused investigation (Survey #2) are presented in the following sections.

4.4 Survey 2: Responsible AI and Leadership Literacy

This section presents the findings from the second quantitative survey conducted as part of this doctoral study. The survey explored leadership readiness, organisational preparedness, and the extent of responsible artificial intelligence (RAI) implementation. A total of 197 participants across multiple sectors and regions completed the full survey, comprising 38 core items and 4 demographic questions. All items were mandatory, ensuring complete responses from all participants.

The survey was manually distributed via LinkedIn to 821 individuals, yielding a conversion rate of 15.47%. Given the passive nature of social media engagement, the actual visibility of the invitations is uncertain.

4.4.1 Demographic Overview

Participants primarily represented the Information Technology and Software Development, Financial Services, and Consulting sectors. Geographically, most respondents were from the United Kingdom (59), mainland Europe (39), Asia (38), and North America (25), with limited representation from Africa (11). Organisational size depicted in Figure 4.7 varied, with the largest cohort (79 respondents) working in enterprises with over 1,000 employees. This was followed by organisations with fewer than 50 employees (65 respondents), 50-249 employees (28 respondents), and 250-1,000 employees (25 respondents).

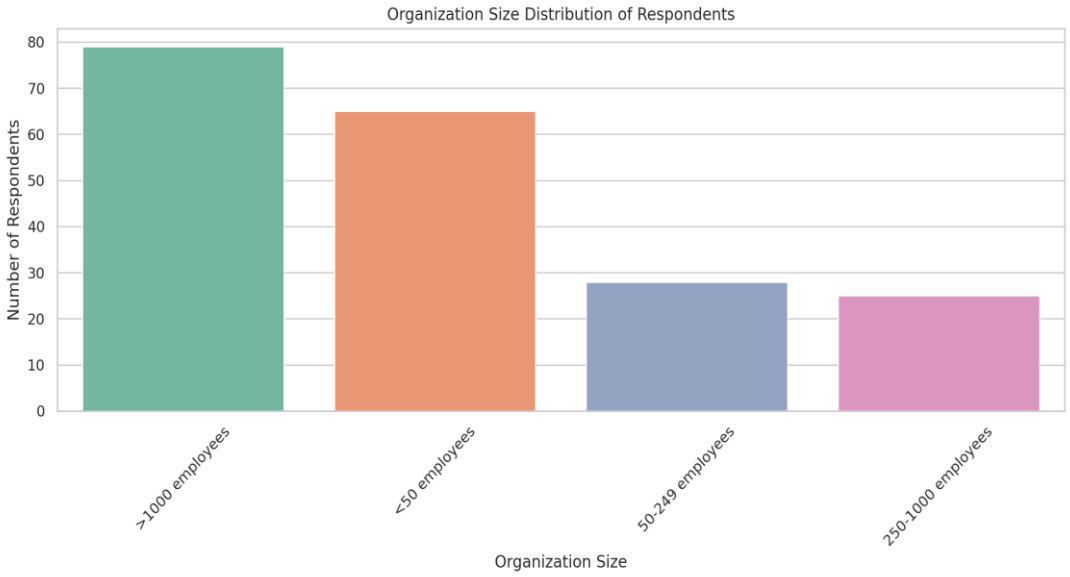


Figure 4.7: Organisation Size

For clarity, the findings that follow have been categorized into five distinct groups.

4.4.2 Key Composite Indices

To facilitate a robust and aggregated analysis of leadership readiness and capability, several items within Survey #2 were combined to form composite indices. These indices move beyond single-item responses to provide broader, more stable

measures of core constructs, such as literacy, preparedness, training efficacy, and leader confidence. The composite measures allowed for the use of advanced statistical modelling, including the regularised linear model, to identify the key drivers of organizational preparedness.

By aggregating related Likert scale items (1=Strongly Disagree to 5=Strongly Agree), these indices provide the average score for each construct. For example, the Responsible AI Literacy Index recorded a mean score of 4.02 and a median of 4.17, indicating that, collectively, respondents perceived themselves as possessing moderate to high literacy in the domain of Responsible AI. Given the identified discrepancy between self-assessed literacy and practical knowledge of governance models, the components of these key indices are defined below in Table 4.1.

Table 4.1: Components of Key Composite Indices

Composite Index	**Mean Score (Survey #2)**	**Illustrative Components (Based on Foundational Themes)**
Responsible AI Literacy Index	4.02	Items related to *understanding* the strategic potential of AI, the ethical implications of AI use, and the general belief that RAI is a critical leadership competency.
Organizational Preparedness Index	3.54	Items related to the perceived readiness of the organization to address AI challenges, the clarity of the Responsible AI strategy, and the provision of legal compliance guidance.
Training Sufficiency Index	(Varies, generally low)	Items related to having received adequate training in AI-related topics, training on ethical implications, and training on assessing AI systems for fairness and bias.
Operational Governance Knowledge Index	(Varies, generally low)	Items related to the sufficiency of knowledge of AI governance models (e.g., NIST RMF), and the need for support in understanding regulatory frameworks.

4.4.3 Understanding of AI Strategic Potential and Preparedness for AI Challenges

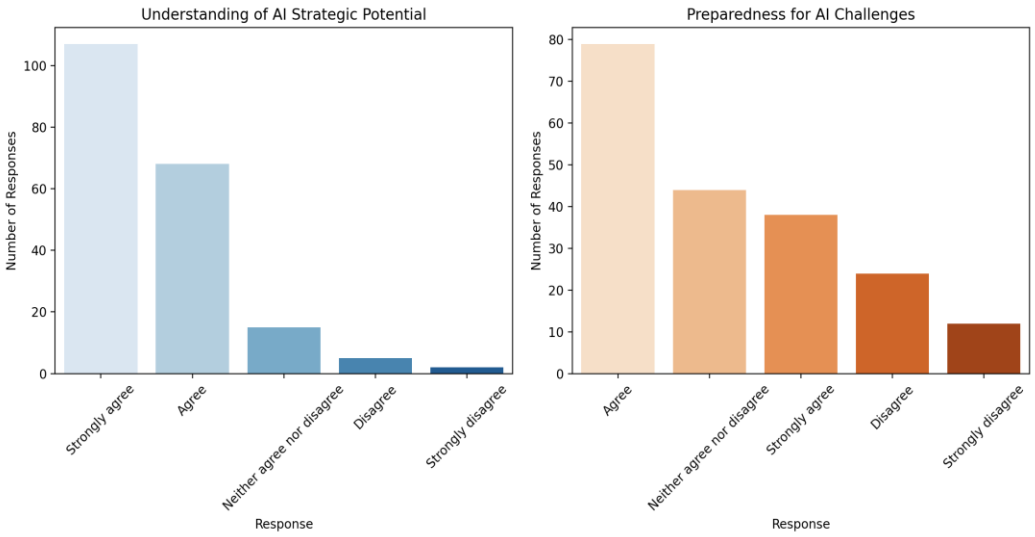


Figure 4.8: Strategic Potential and Preparedness

The left side of Figure 4.8 illustrates the distribution of responses regarding the comprehension of the strategic potential of AI. It reveals the extent to which leaders' express confidence in their understanding of AI's role within their respective industries. Conversely, the right side depicts the preparedness of organizations to address the challenges associated with AI, offering insights into their readiness to implement AI in a responsible manner.

Despite a strongly espoused commitment to Responsible AI, the survey identified substantial deficiencies in training and education. Nearly 87% of respondents reported that current leadership education does not sufficiently address the competencies necessary for effective AI implementation. There was a consistent demand for training in several critical areas, including the evaluation of AI systems for fairness and bias, leadership in AI-driven organizational change, and the societal, legal, and ethical implications of AI. Respondents also indicated that they frequently received inadequate or insufficient training on AI-related topics pertinent to their leadership roles. A clear

preference emerged for educational formats that are scenario-based or immersive, particularly those that engage with real-world case studies involving AI bias or misuse.

There exists a notable discrepancy between the recognition of AI's strategic potential and the level of preparedness, both at the individual and organizational levels, to actualize that potential. While the assertion regarding AI's strategic potential achieved a high mean score of 4.39, the perceived readiness of organizations to address the challenges posed by AI was significantly lower, with an average score of only 3.54. Respondents expressed low confidence in their capacity to manage AI systems ethically and legally, a sentiment further reflected in the limited organizational support structures available. Specifically, the clarity of organizational strategies for implementing Responsible AI was rated at a mean of 3.30, and the provision of guidance on legal compliance in AI development and deployment was rated even lower, at 3.24. These figures indicate a substantial gap between aspirational rhetoric and institutional readiness.

The Responsible AI Literacy Index, an aggregate measure derived from multiple items, recorded a mean score of 4.02 and a median of 4.17, suggesting that most leaders perceive themselves as possessing moderate to high literacy in the domain of Responsible AI. Nevertheless, the index also revealed that approximately 17% of respondents (33 out of 197) scored below 3.0, indicating significant knowledge gaps and underscoring a distinct need for further development within this subgroup.

4.4.4 Ethical Imperative, Fairness and Bias

The findings pertaining to the statement, ‘I believe promoting fairness and human dignity is essential in AI-related business decisions,’ indicate a remarkably strong consensus among respondents, as shown in Figure 4.9: 57.9% strongly agree (114 respondents), and 31.5% agree (62 respondents). This indicates that nearly 90% of the 197 respondents affirm the ethical imperative of fairness and human dignity in AI

decision-making. Such overwhelming agreement suggests that these principles are not only widely endorsed but are likely perceived as foundational to the responsible adoption of AI within organizational contexts.

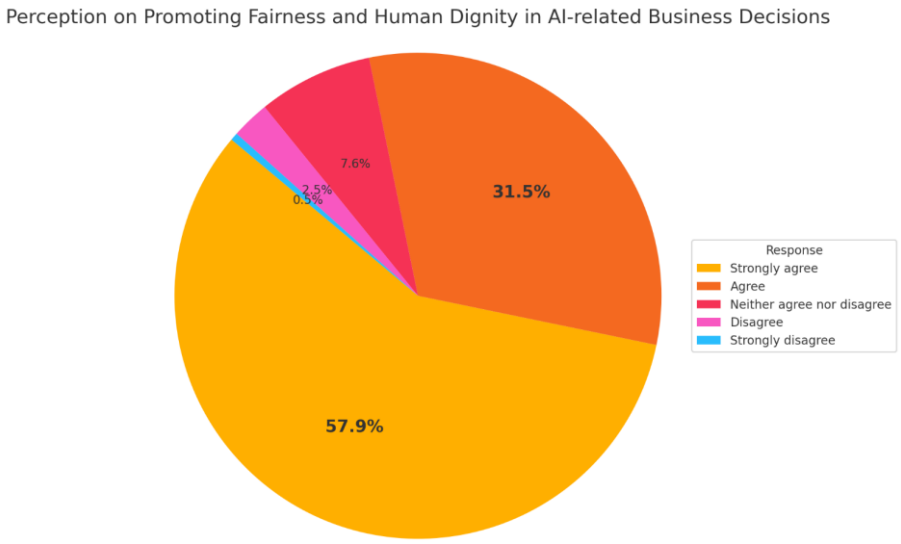


Figure 4.9: Fairness and Human Dignity

The question, 'I understand the ethical implications of using AI in business decision-making,' predominantly drew a confident response from participants. A total of 82.2% (162 respondents) expressed agreement, with 104 agreeing (52.8%) and 58 strongly agreeing (29.4%). Only 11.7% remained neutral, and a mere 6.1% expressed disagreement.

Figure 4.10: Training Understanding Limitations and Risks of AI

The question: 'I need more training in understanding the limitations and risks of AI systems' (Figure 4.10) underscores a significant demand for improved AI risk literacy among organizational leaders and professionals. A substantial majority, 73.1% of respondents, acknowledge the necessity for additional training on AI limitations and risks. This finding corroborates a recurring trend observed throughout the survey: as responsibilities for AI integration expand, so too does the recognition of existing knowledge gaps.

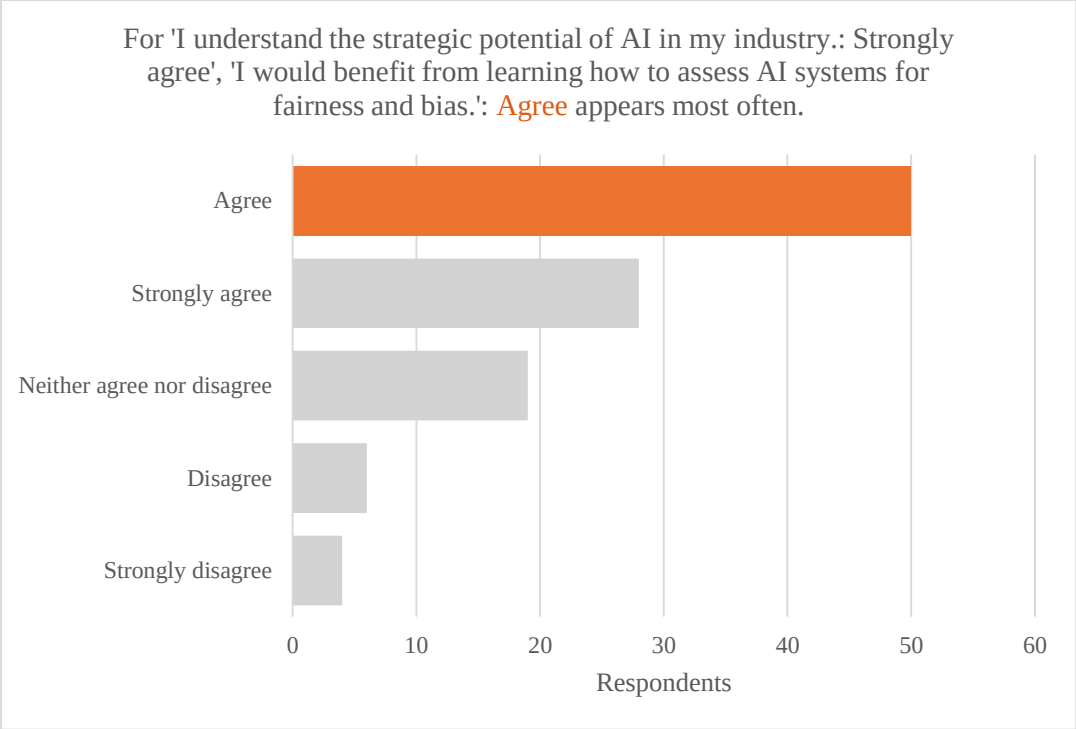


Figure 4.11: Strategic Potential and Fairness

The statement, ‘I would benefit from learning how to assess AI systems for fairness and bias,’ from respondents who already strongly agree that they understand the strategic potential of AI in their industry (Figure 4.11), reveals a critical insight: strategic awareness does not equate to ethical or technical preparedness. This highlights a significant skills gap in AI ethics, even among those confident in AI's strategic value. Nearly three-quarters of these leaders (72.9%) acknowledge that they would benefit from learning how to assess fairness and bias, a foundational aspect of responsible AI.

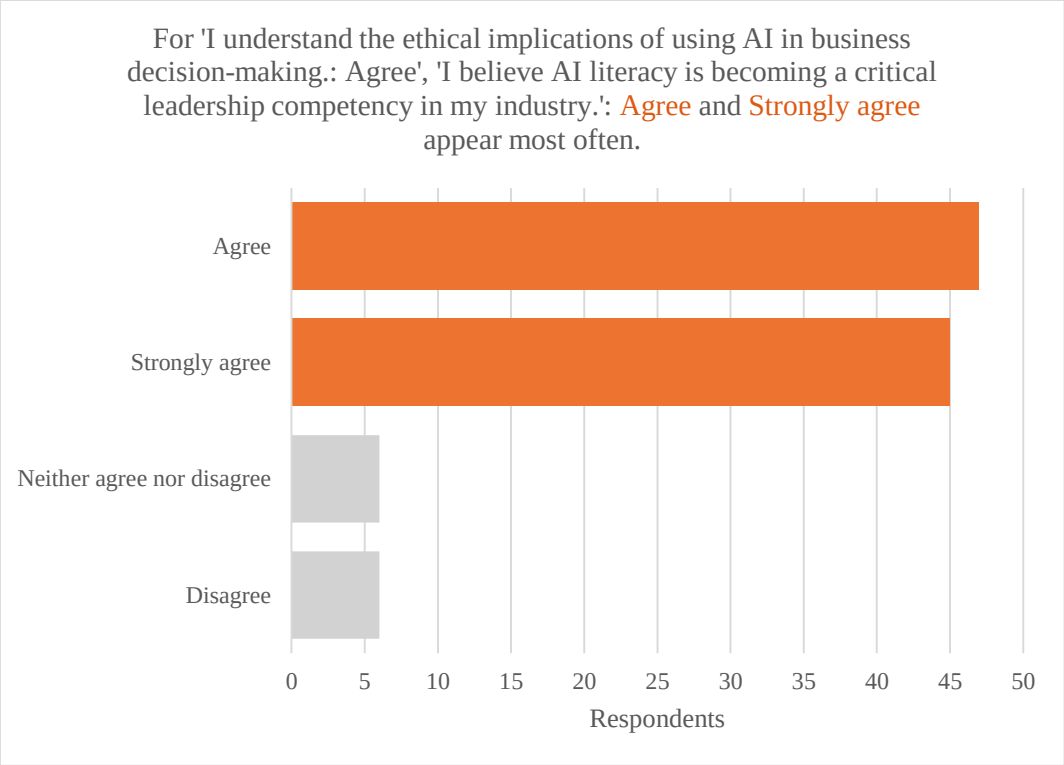


Figure 4.12: Ethical Implications and AI Literacy

The filtered dataset represents the perspectives of respondents who concur that they comprehend the ethical implications of employing AI in business decision-making. It evaluates their belief in the significance of AI literacy as an essential leadership competency. Notably, as shown in Figure 4.12, nearly nine out of ten respondents who understand AI's ethical implications also regard AI literacy as a critical leadership competency (92). This finding underscores a strong alignment between ethical awareness and the expectations of AI capability at the leadership level.

4.4.5 Leadership Roles and Evolving Responsibilities

Notably, 74.1% of respondents from organisations that operate internationally acknowledged that their roles have expanded to include responsibilities related to AI integration and deployment, as shown in Figure 4.13. This evolution was more prominent

in internationally operating organisations, suggesting a global trend toward embedding AI into strategic and operational functions.

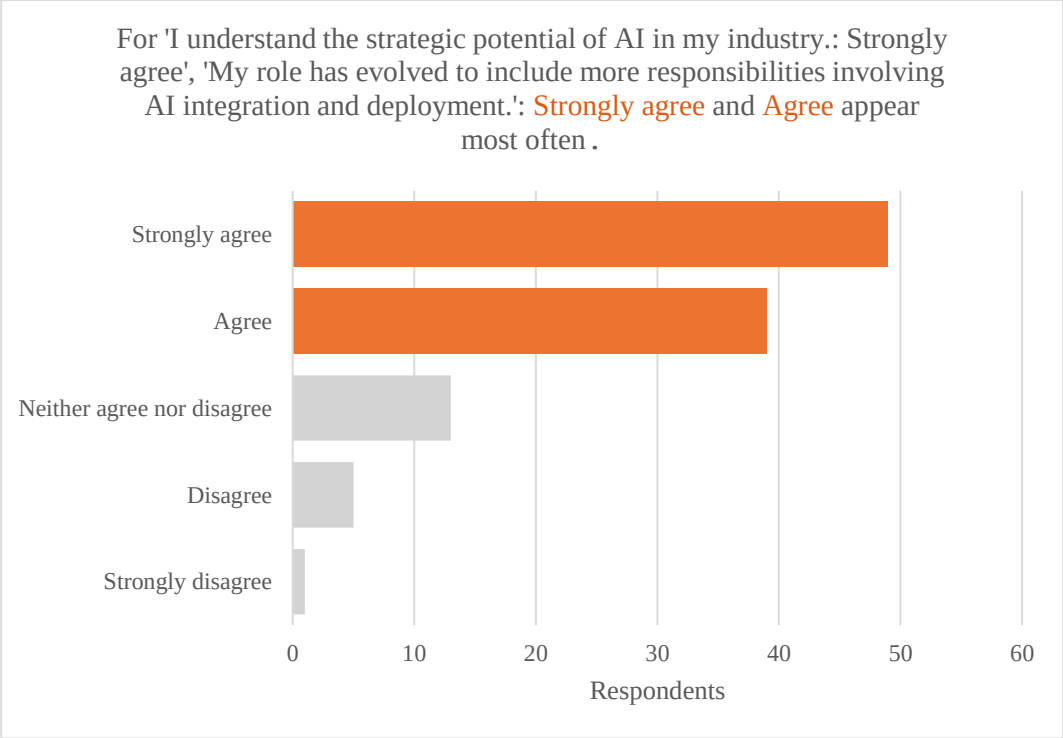


Figure 4.13: Strategic Potential and Responsibilities

A clear relationship was observed between strategic understanding of AI and the evolution of professional roles. Among respondents who reported a strong grasp of AI's strategic potential within their industry, over 80 per cent indicated that their roles had expanded to include responsibilities related to AI integration and deployment. This finding suggests that strategic AI literacy functions as an active enabler of direct involvement in organisational AI transformation rather than a passive form of knowledge.

A dominant theme throughout the data is the strong normative belief in the importance of responsible AI as a leadership competency. Over 90% of respondents agreed or strongly agreed with the statement, 'I believe that understanding Responsible

AI is a critical leadership competency in today's business environment,' yielding a mean score of 4.45 out of 5.00.

The perception of AI literacy as a critical leadership skill was affirmed by over 90% of respondents (126) in international roles. However, a gap remains between perceived strategic understanding and operational readiness. For example, while over 80% (88 respondents) who self-identified as strategically literate in AI reported expanding roles, nearly three-quarters of these leaders (72.9%) also admitted to lacking knowledge of AI fairness and bias evaluation methods, indicating a significant skills gap in AI ethics.

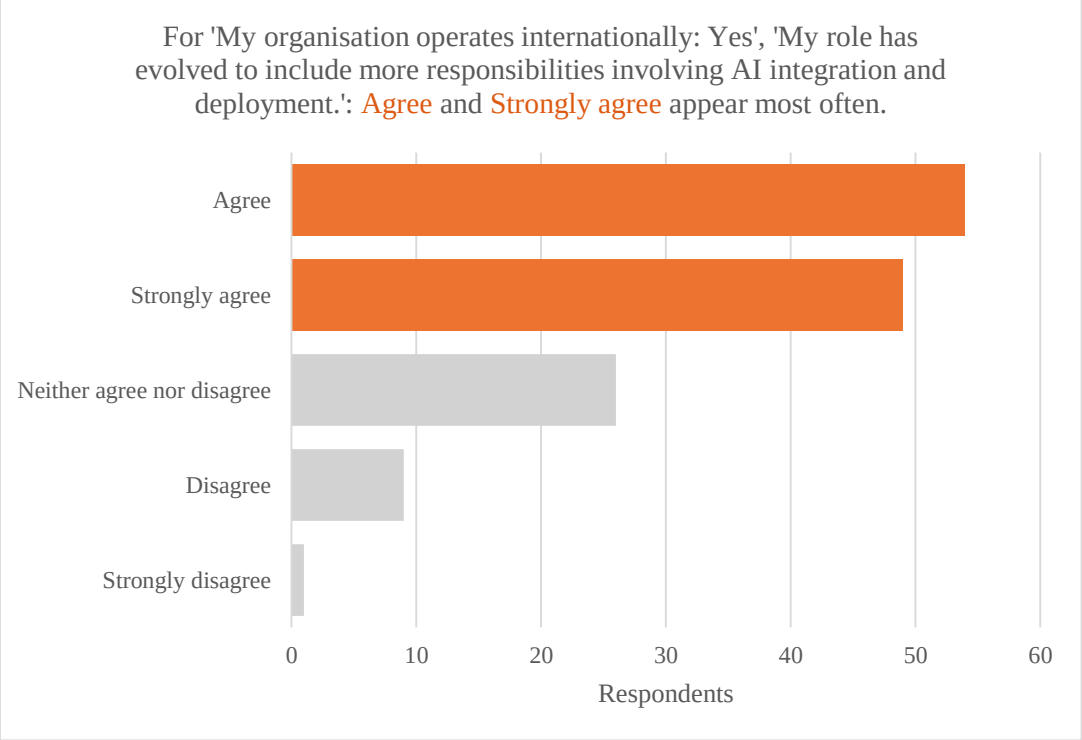


Figure 4.14: International Operation and Responsibilities

The survey item stating that respondents' roles had evolved to include responsibilities for AI integration and deployment provides insight into the changing

nature of leadership and professional practice. As illustrated in Figure 4.14, 74.1 per cent of respondents from internationally operating organisations reported an expansion of their roles in this direction. This finding indicates a substantive organisational shift, particularly within global enterprises, towards embedding AI within routine strategic and operational activities.

The statement that 'AI literacy is becoming a critical leadership competency in my industry' represents one of the most compelling endorsements within the dataset, indicating a strong consensus on the significance of AI literacy. Over 90% of respondents (126) from organizations with international operations concur that AI literacy has emerged as a fundamental leadership competency, particularly in globally operating entities. This reflects a shift in what defines leadership effectiveness in the digital era.

4.4.6 Strategic Literacy vs Organisational Readiness

Despite high strategic awareness, indicated by a mean score of 4.39 for understanding the strategic potential of AI in their industry, organisational preparedness lagged, with a markedly lower average score of 3.54 for perceived readiness to meet AI challenges. Specifically, ratings for clarity of Responsible AI strategy (mean of 3.30) and provision of legal compliance guidance in AI development and deployment (mean of 3.24) were notably low. This reflects a persistent gap between rhetorical commitment and implementation capability.

The Responsible AI Literacy Index, a composite measure, had a mean of 4.02 and a median of 4.17, indicating moderate to high self-reported literacy among most leaders. However, 17% of respondents (33 out of 197) scored below 3.0, exposing a critical subset in need of targeted development and indicating significant knowledge gaps.

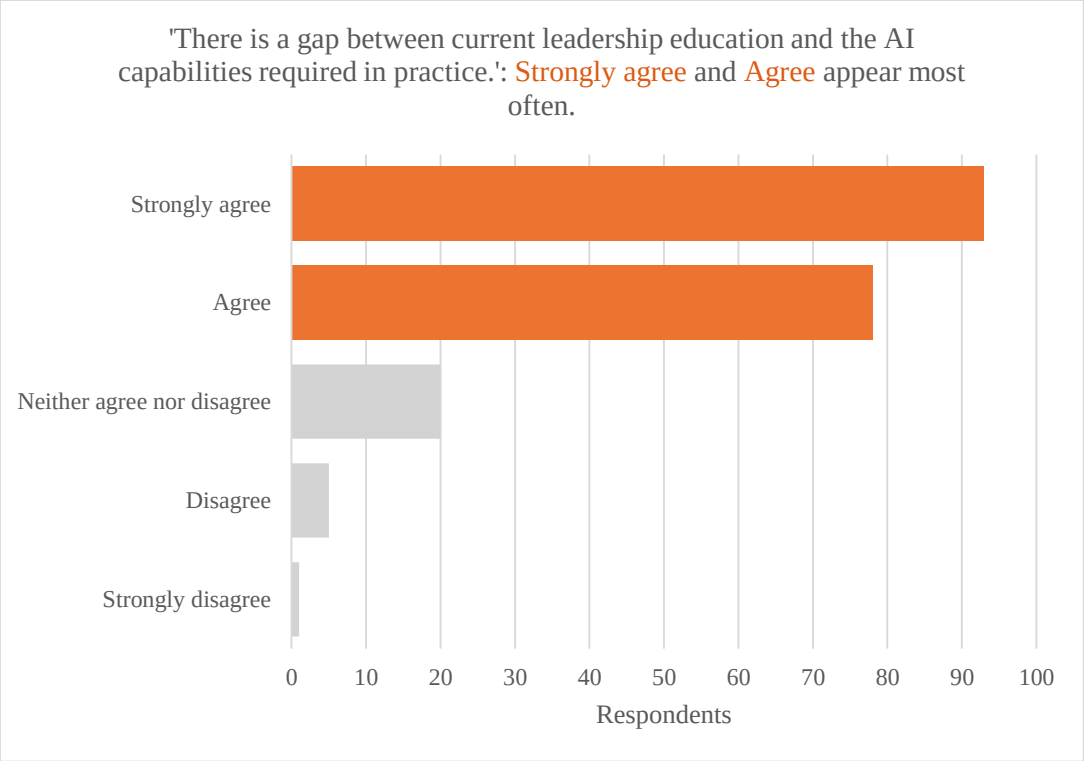


Figure 4.15: Leadership Education and AI Capabilities

A significant proportion of respondents (approximately 90%) agreed that current leadership education is insufficient for addressing AI's practical challenges, as shown in Figure 4.15. The item 'There is a gap between current leadership education and the AI capabilities required in practice' received widespread agreement, with 93 respondents strongly agreeing and 78 agreeing. Respondents consistently reported inadequate training in areas such as AI governance, ethical implications, and risk awareness.

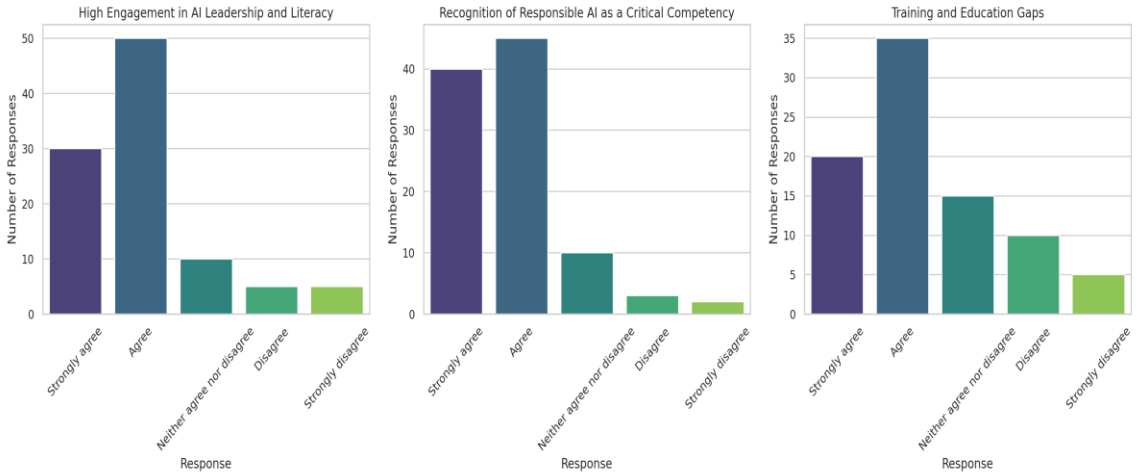


Figure 4.16: Levels of Confidence

Figure 4.16 illustrates respondents' levels of confidence in their AI literacy and engagement in AI leadership, revealing a high degree of self-reported involvement. The findings also indicate broad recognition of responsible AI as a critical leadership competency, with the majority of respondents reporting that their organisations have strategies in place for its implementation. However, the distribution of responses highlights persistent gaps in leadership development, particularly in relation to applied and scenario-based training. While many leaders perceive their training as adequate, a substantial proportion identify the need for more immersive approaches to strengthen practical capability.

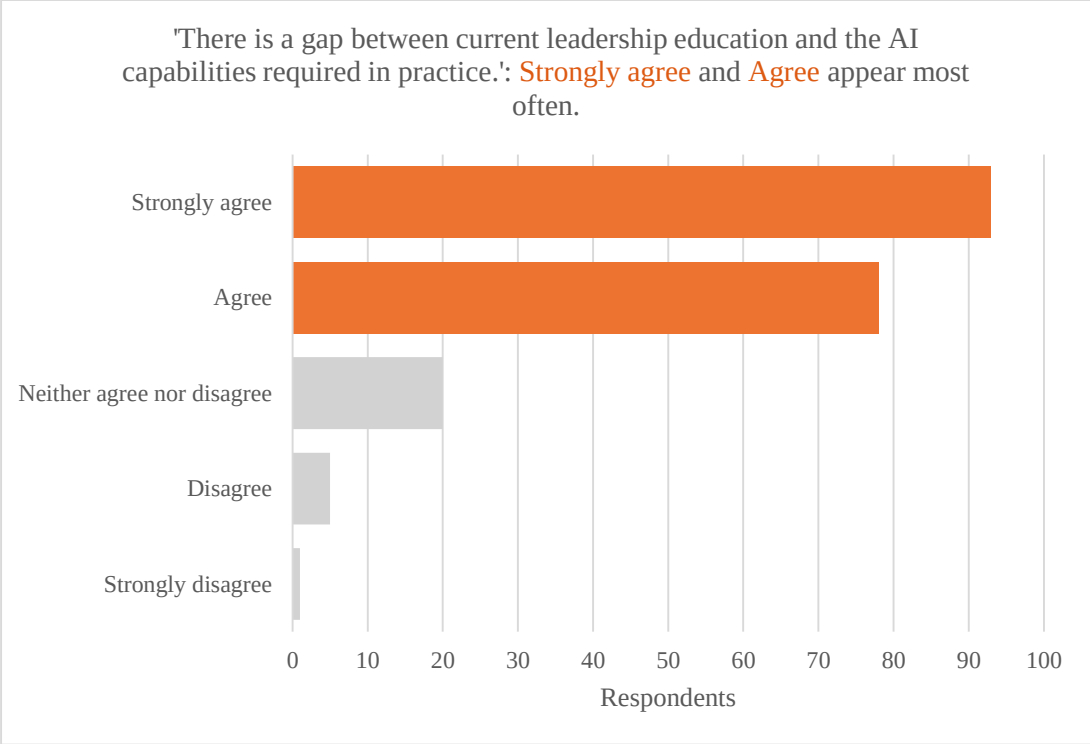


Figure 4.17: Leadership Education and AI Capabilities

Figure 4.17 depicts a notable gap existing between the current state of leadership education and the practical AI competencies required in the field. Key observations reveal that a substantial majority, nearly 90% of respondents, align with this view. Specifically, 93 respondents strongly agree, while 78 agree. Conversely, only a minimal number of respondents expressed disagreement or strong disagreement, with a modest group of 20 remaining neutral.

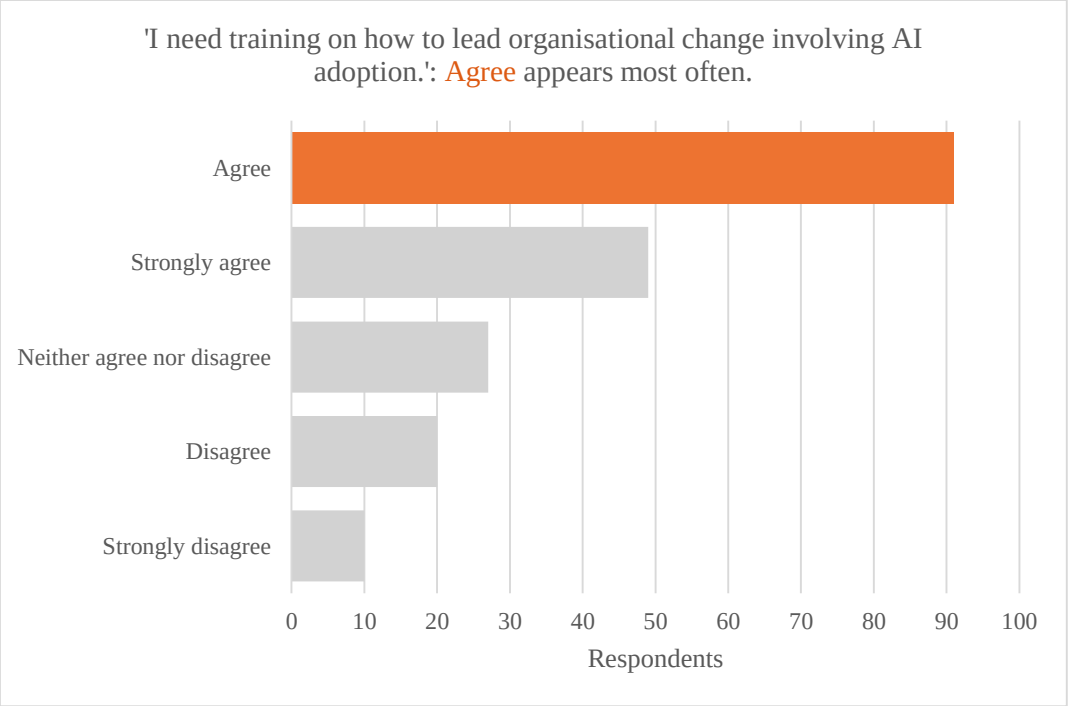


Figure 4.18: Organisational Change

The results pertaining to the statement, ‘I need training on how to lead organisational change involving AI adoption,’ in Figure 4.18 underscore a significant demand for leadership development in the context of AI integration: The survey results indicate that 91 respondents (46.2%) agreed, while 49 respondents (24.9%) strongly agreed, culminating in a total agreement rate of 71.1%.

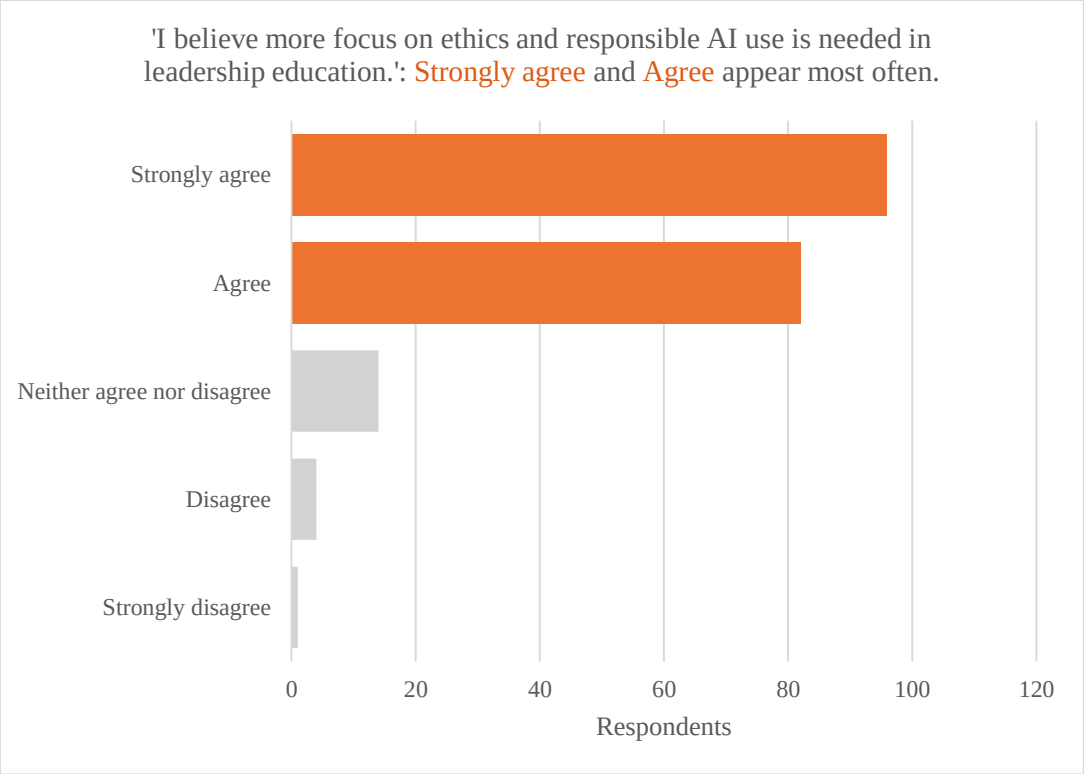


Figure 4.19: Ethics & Responsible AI

Responses to the statement ‘I believe more focus on ethics and responsible AI use is needed in leadership education,’ as presented in Figure 4.19, indicate a strong level of agreement among respondents.

4.4.7 Discrepancies in AI Governance & Regulatory Understanding

Conflicting evidence emerged regarding familiarity with AI governance models. While some sources indicated confidence in understanding frameworks such as the NIST RMF, the survey item on governance knowledge, specifically ‘My knowledge of AI governance models is sufficient for my role,’ yielded a low mean score of 3.07. This suggests that while leaders recognise the general importance of governance, their practical comprehension of specific frameworks remains insufficient.

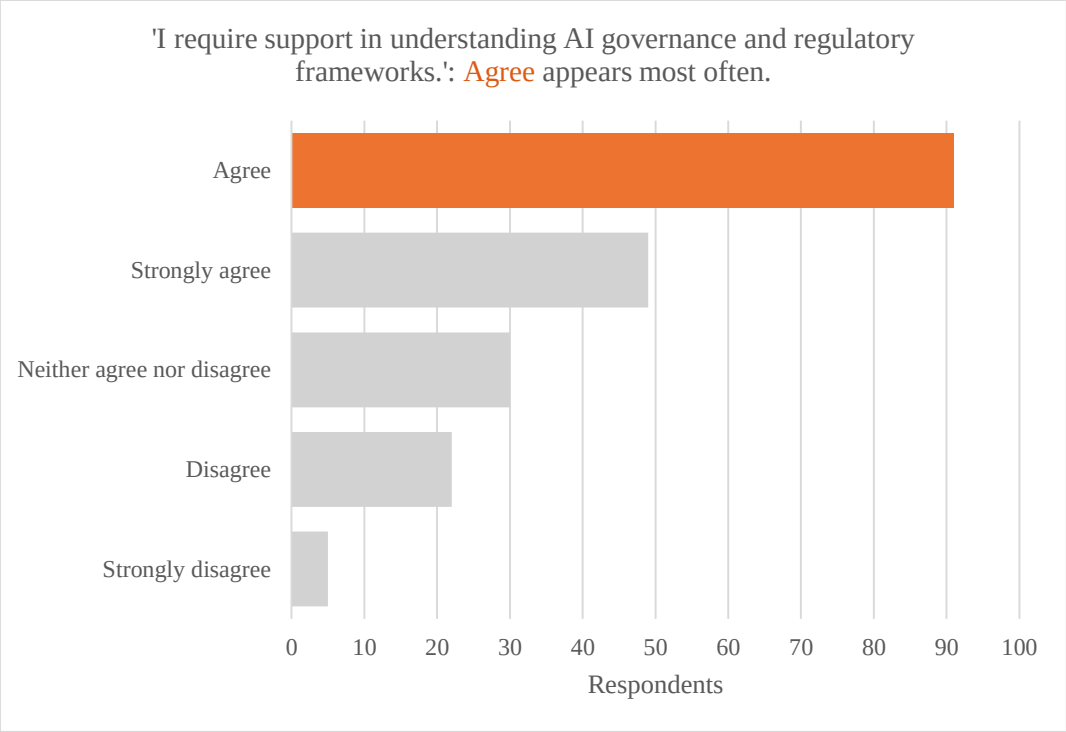


Figure 4.20: AI Governance & Regulation

The statement, 'I require support in understanding AI governance and regulatory frameworks,' (Figure 4.20) underscores a substantial demand for skill enhancement and guidance in this pivotal domain. Among the respondents, 91 individuals (46.2%) agreed, while 49 individuals (24.9%) strongly agreed, resulting in a combined agreement rate of 71.1% (140 out of 197 respondents). Additionally, 30 respondents (15.2%) neither agreed nor disagreed, 22 respondents (11.2%) disagreed, and 5 respondents (2.5%) strongly disagreed.

While 82.2% (162 respondents) claimed to understand the ethical implications of using AI in business decision-making, 73.1% expressed a need for further training in AI risks and limitations. Additionally, 71.1% (140 respondents) reported needing support in understanding governance and regulatory frameworks, highlighting a key knowledge gap in legal and ethical oversight of AI systems.

Analysis of the filtered dataset, comprising respondents who indicated a need for support in understanding AI governance and regulatory frameworks, reveals a strong perceived interdependence between governance, regulation, ethics, and legal oversight. Within this group, 91.2% (83) of respondents also reported requiring additional support in the ethical and legal supervision of AI systems. This finding suggests that many professionals conceptualise AI oversight as an integrated set of responsibilities rather than as discrete domains.

A further filtered analysis examined the perspectives of respondents who strongly agree that a disparity exists between current leadership education and the AI capabilities required in practice. It evaluated their opinions on the necessity for an increased emphasis on ethics and responsible AI use within leadership education. Nearly all respondents, exceeding 96% (90), advocate for the explicit integration of ethics and responsible AI use in leadership education. This indicates a pronounced collective recognition that technical or strategic AI capabilities alone are insufficient for leadership that is prepared for future challenges.

4.4.8 Cluster and Sentiment Analysis

Figure 4.21: Cluster Analysis

To further analyse the data, a clustering analysis was conducted, categorizing respondents into three distinct groups based on their reported AI literacy and perceived training needs, shown in Figure 4.21. The first cluster consisted of high-confidence leaders, typically affiliated with large or technologically mature organizations. These individuals reported a strong sense of preparedness and indicated relatively low support needs. The second cluster included moderate-confidence leaders who acknowledged the importance of Responsible AI (RAI) but expressed a need for additional support and training to operationalize it effectively. The third cluster was characterized by low-confidence leaders, many of whom came from smaller enterprises or less technologically advanced sectors. This group reported both limited understanding and significant training requirements.

4.4.9 Categorical Analysis of Sentiment

Average sentiment scores across categories provide additional insight into the findings. In the category of Responsible AI importance and competency, the mean score of 3.99 indicates that leaders attach high importance to responsible AI, while reporting only moderate confidence in their practical capabilities relating to implementation, risk assessment, and legal and ethical understanding.

The category of AI understanding and engagement produced a mean score of 3.90, indicating generally positive perceptions of leaders' comprehension of AI's strategic potential and their engagement with emerging AI trends. The category of training needs and gaps recorded a mean score of 3.79, reflecting a strong perceived requirement for enhanced AI education, particularly in relation to ethics, societal impact, and governance, where existing provision is viewed as insufficient. Organisational preparedness and strategy yielded the lowest mean score at 3.30, suggesting that many organisations are perceived as inadequately prepared for AI related challenges, with limited responsible AI strategies and insufficient guidance on governance and legal compliance.

4.4.10 Areas of Strength and Weakness

This section presents the highest and lowest mean scores derived from the survey responses, highlighting areas where leaders express the strongest confidence and where perceived capability gaps are most pronounced. The results reveal a clear distinction between normative awareness and strategic recognition on the one hand, and practical preparedness and organisational enablement on the other. While respondents demonstrate strong consensus regarding the importance of Responsible AI, fairness, and AI literacy as critical leadership competencies, substantially lower scores are observed in relation to governance knowledge, formal training, and organisational support mechanisms. This

divergence suggests that although responsible AI is widely valued at a conceptual and ethical level, it is not yet being matched by sufficient capability development, structured guidance, or strategic implementation within organisations.

Top 5 Areas of Perceived Strength (Highest Mean Scores):

1. I believe that understanding Responsible AI is a critical leadership competency in today's business environment. (4.45)
2. I believe promoting fairness and human dignity is essential in AI-related business decisions. (4.44)
3. I understand the strategic potential of AI in my industry. (4.39)
4. I believe more focus on ethics and responsible AI use is needed in leadership education. (4.36)
5. I believe AI literacy is becoming a critical leadership competency in my industry. (4.31)

Top 5 Areas of Perceived Weakness/Need for Support (Lowest Mean Scores):

1. My knowledge of AI governance models (e.g. NIST AI Risk Management Framework) is sufficient for my role. (3.07)
2. I have received adequate training in AI-related topics relevant to my leadership role. (3.11)
3. I have received training that includes the ethical implications of AI use. (3.13)
4. My organisation provides guidance on legal compliance in AI development, deployment and use. (3.24)
5. My organisation has a clear strategy for implementing Responsible AI. (3.30)

4.4.11 Specific Factors Influencing Organisational Readiness for Responsible AI

To deepen understanding of the findings linked to responsible AI, the analysis was enhanced by inferential statistical methods upon the dataset. These techniques move

beyond descriptive patterns and correlations to identify which organisational practices or leadership characteristics most strongly predict outcomes such as preparedness, confidence, and access to training. This shift from description to explanation provides a clearer view of the drivers that underpin responsible AI capability.

4.5 Statistical Modelling and Advanced Analysis

A regularised linear model was used to examine how leadership beliefs and governance practices shape organisational readiness. This approach reduces the influence of less informative predictors and helps prevent overfitting, which can occur when the number of predictors is high relative to the sample size. Cross-validation was used to test the model across different subsets of the data, reducing the risk that results reflect random variation. Two metrics were central: the cross-validated R^2, indicating how much variation the model explains, and the mean absolute error, indicating the scale of prediction error. Model outputs were then summarised in accessible terms to highlight the factors most associated with readiness as shown in Figure 4.22.

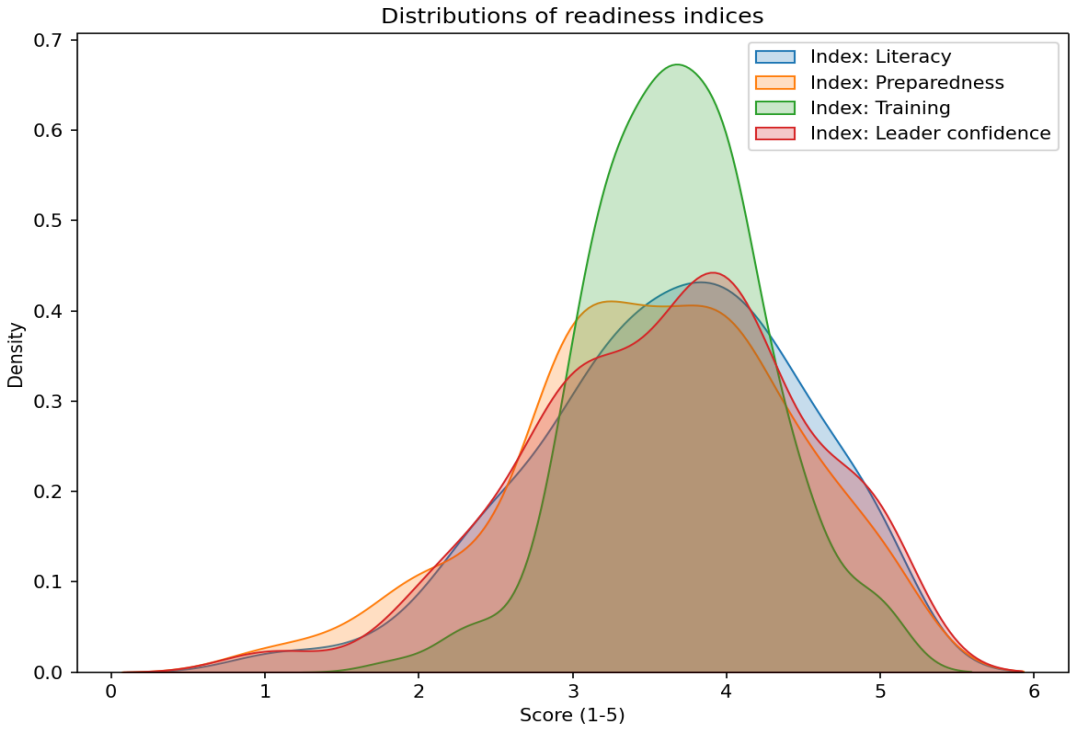


Figure 4.22: Distribution of readiness indices

The analysis (Figure 4.22) indicates that organisational preparedness for responsible AI constitutes an emergent capability shaped by strategic orientation, leadership literacy, governance practice, and ethical awareness. The predictive model demonstrated acceptable performance for survey based organisational research, with evaluation metrics suggesting a stable and meaningful signal. Leadership perceptions and governance practices accounted for a substantive proportion of the variance in preparedness, notwithstanding the subjectivity associated with self-reported data.

While the primary focus of the enhanced statistical modelling was to identify the core organizational and leadership drivers of preparedness (i.e., strategy, literacy, and governance), the analysis was extended to assess the influence of demographic and sectoral variables on the core readiness indices. This step was crucial for ensuring that the proposed capability model is appropriately contextualized to the diverse operational landscapes represented by the participants.

The initial descriptive analysis noted the sample's concentration in sectors such as Information Technology, Financial Services, and Consulting, and observed differences, such as the finding that older respondents might place greater emphasis on ethical conduct. To test the statistical significance of these observations, demographic factors (age, experience, region) and organizational factors (sector, organization size) were included as predictors in the regularised linear model alongside the core leadership and governance constructs.

4.6 Demographic and Sectoral Influences

The analysis confirmed that the strongest predictors of organisational preparedness were internal organisational factors, namely strategic clarity, governance practice, and leadership literacy. However, contextual factors were also found to exert a meaningful secondary influence.

Sectoral Influence: Statistically significant differences were observed across sectors. Respondents within financial services and information technology reported higher confidence in technical discourse and greater awareness of AI related risks, reflecting the more mature risk management and compliance cultures characteristic of these industries. This suggests that sectoral governance maturity shapes organisational starting points for the development of responsible AI capability.

Regional Influence: Regional variation was present but comparatively weaker than sectoral effects, indicating that sector specific norms and regulatory exposure, such as familiarity with GDPR among European respondents, exert greater influence on preparedness than geographic location alone.

Organisational Size: Cluster analysis revealed that respondents reporting higher levels of preparedness were more commonly associated with large or technologically mature organisations with over 1,000 employees, whereas lower confidence was more

prevalent among respondents from smaller enterprises. This pattern suggests that resource availability and established organisational infrastructure act as indirect enablers of responsible AI preparedness.

These findings underscore that while the implementation of the capability model is universally necessary, the starting level of preparedness and the specific training needs must be tailored according to the organization's sectoral and size profile.

Table 4.2 clarifies the results of the inferential analysis by categorizing the types of predictors, thereby aligning the summary of the enhanced statistical modelling with the necessity to explicitly address demographic and sectoral influences.

Table 4.2: Statistical Correlates of Organisational Readiness

Variable Category	**Predictor Variable**	**Direction of Correlation (Predicting Preparedness)**	**Finding/Source Support**
Core Organizational Driver	Strategic Clarity (clear RAI strategy)	Strong Positive	Identified as a core theme for preparedness.
Core Leadership Driver	Leadership AI Literacy	Strong Positive	Strong association with readiness; links strategic intent to operational delivery.
Core Governance Driver	Knowledge of Governance Frameworks (NIST, EU AI Act)	Positive	Correlates with stronger readiness and assures compliance.
Sectoral Context	Financial Services/IT Sector Affiliation	Moderate Positive	Respondents in these sectors report higher confidence and risk awareness.
Organizational Context	Organization Size (>1000 employees)	Moderate Positive (Implied)	High-confidence cluster typically affiliated with large organizations.
Demographic Context	Age/Work Experience	Minor/Contextual	Older respondents may emphasize ethical conduct more heavily.

Table 4.2 summarises the key variables associated with perceived organisational preparedness for Responsible AI adoption, revealing that readiness is primarily shaped by strategic clarity, leadership capability, and governance maturity, rather than by demographic characteristics alone. Leaders who operate within organisations with a clearly articulated Responsible AI strategy report higher levels of preparedness, suggesting that strategy plays a codifying role by translating ethical principles into actionable expectations and reducing ambiguity in decision making. Leadership AI literacy emerges as a critical enabling factor, with the ability to engage meaningfully with technical concepts such as explainability functioning as the connective tissue between strategic intent and operational execution. Governance knowledge further strengthens preparedness, as familiarity with recognised frameworks and regulatory requirements, including established risk management models and the EU AI Act, provides the structural mechanisms for accountability, risk identification, and compliance assurance. Sectoral and organisational context moderates these relationships, with larger organisations and those operating in regulated or technology-intensive sectors reporting greater confidence. By contrast, demographic factors such as age and work experience play a secondary, contextual role, shaping ethical emphasis rather than materially determining preparedness.

Across the dataset, literacy scores are generally higher than preparedness and training. This indicates a gap between awareness and the capacity to execute. Preparedness varies more widely across sectors and regions, with some reporting consistently lower or more uneven levels. Training is the most variable and frequently the lowest of the indices, which echoes the modelling challenges encountered. Leader confidence aligns more closely with literacy than with preparedness, suggesting that

leaders may feel comfortable discussing AI even where organisational structures remain underdeveloped.

The education gap is also notable. Respondents widely acknowledge the importance of AI literacy and ethical competence but report unmet needs in areas such as governance, societal impact, and applied practice. The preference for case-based learning suggests that experiential methods may accelerate capability building more effectively than theory alone.

Overall, the findings support a capability model in which preparedness is built through four interdependent pillars: strategic clarity, leadership literacy, operational governance, and ethical fluency. Sectoral variation reflects different starting points in governance maturity rather than different fundamental requirements. The practical implication is clear. Investment in leadership education that integrates regulatory frameworks, explainability, and scenario-based ethics, combined with a clear responsible AI strategy and internal guidance, can improve preparedness. The results also indicate that leadership-level variables are suitable indicators of organisational readiness, while highlighting the value of longitudinal data to track how training and policy interventions shape preparedness over time.

4.7 Cross-Survey Synthesis and Emerging Capability Patterns

The data reveals a high-level recognition of the ethical, legal, and strategic imperatives surrounding Responsible AI. However, this is undermined by systemic shortcomings in education, training, organisational strategy, and governance clarity. These deficiencies are particularly pronounced in the areas of governance, regulatory compliance, and ethical oversight. The findings call for leadership education programmes that incorporate immersive, scenario-based learning, focusing on real-world challenges such as algorithmic bias and regulatory compliance.

To realise the full potential of AI, it is imperative that organisations invest in leadership development and institutional infrastructure that effectively bridge the gap between ethical aspirations and operational implementation. This is discussed in the next section.

4.8 Methodology Reflection and Critique

There are several areas in the study that could provide a deeper and more nuanced understanding of Responsible AI and leadership AI literacy.

The study predominantly evaluates AI literacy and preparedness through self-reported confidence levels and agreement with specific statements. Although leaders generally perceive themselves as ‘moderately to highly prepared’ on the RAI Literacy Index (mean 4.02, median 4.17), there exists a significant discrepancy concerning their knowledge of AI governance frameworks. This observation implies that leaders' self-assessed comprehension may not correspond with their actual, practical knowledge. This discrepancy highlights an opportunity to further investigate this aspect, considering its significance in the study, to more objectively assess AI literacy.

The study could be further enhanced through greater granularity concerning leadership roles and seniority, as it does not consistently disaggregate findings by specific leadership tiers or roles. This would address a gap in understanding whether AI literacy requirements, perceived deficiencies, or training needs vary significantly among different leadership levels, such as C-suite executives, departmental heads, or project managers. It is likely that these distinct leadership levels engage with AI and its governance to varying extents.

The majority of respondents originate from the United Kingdom, Europe (excluding the UK), and North America, with significant representation from the Information Technology and Software Development, Financial Services, and Consulting

sectors. Although these sectors are pivotal for the adoption of AI, this concentration may result in findings that do not comprehensively reflect the distinct challenges, regulatory environments, or cultural perspectives on AI ethics from other important global regions (e.g., Asia, Africa, South America, Middle East) or from industries with lower levels of AI maturity.

The analyses effectively identify key findings, highlight areas of consensus and disagreement, and categorize respondents into clusters. However, they predominantly present descriptive statistics and correlations. The study does not explicitly engage in deeper causal analysis or employ inferential statistical methods to ascertain the factors that most strongly predict higher RAI literacy, enhanced organizational preparedness, or the likelihood of receiving adequate training. For instance, while cluster analysis indicates that 'High-confidence leaders' are 'often from large or tech-driven organizations,' the underlying drivers of this confidence or the impact of specific organizational practices beyond broad demographics could be more deeply examined.

The survey addresses the topic of AI and Responsible AI within a general business framework. Each AI application can present distinct ethical, legal, and governance challenges, and the perceived literacy and preparedness of individuals may vary significantly depending on the specific AI context. This provides the opportunity to delve into specific AI applications and use cases that leaders may encounter.

The survey employs a straightforward Likert scale for all questions; however, the absence of open-ended questions limits the ability to uncover the nuanced reasoning behind the quantitative responses. While effective for quantitative analysis, there is an opportunity to gather rich, qualitative insights into why leaders hold certain perceptions, what specific barriers they face in implementing Responsible AI, or what precise

elements they believe are missing from current leadership education beyond broad categories.

CHAPTER V

DISCUSSION AND CONCLUSION

5.1 Opportunities & Risks

5.1.1 The Role of AI in Sustaining Economic Growth

AI presents significant opportunities and is widely recognised as a foundational technology of the Fourth Industrial Revolution, succeeding the transformative impacts of steam power, electricity, and digital technologies (Schwab, 2016).

AI possesses the potential to revolutionize global economies, enhance organizational efficiency, and streamline the customer experience. Projections indicate substantial improvements in global GDP and the operational success of enterprises through the integration of AI technologies. However, these opportunities are accompanied by considerable risks, including data breaches, biases within AI systems, lack of transparency, and the spread of misinformation. If not properly managed, these risks could result in severe consequences, such as regulatory sanctions, fines, market exclusion, reputational damage, diminished market share, and declining stock prices.

Economic growth is primarily driven by technological innovation and cultural transformations that emphasize the pursuit of novel and beneficial ideas (Susskind, 2024). While institutions and ideas are essential for sustaining long-term economic expansion, innovation acts as a critical multiplier that enhances these effects. Directed technological change, in conjunction with well-formulated public policies, can influence the trajectory of growth and contribute to its sustainability.

From a business perspective, three primary concerns dominate strategic decision-making: revenue growth, cost reduction, and risk mitigation. For publicly listed companies, maintaining share price stability and protecting market share are also critical priorities. As businesses seek to navigate an increasingly complex economic landscape,

AI emerges as a pivotal enabler of sustained economic growth, particularly in the face of demographic shifts, resource constraints, and environmental challenges.

AI facilitates directed technological change by accelerating the shift from material-based to idea-based economic activities. This transformation encourages growth through intellectual contributions rather than resource-intensive production, thereby promoting sustainability and efficiency. By automating routine processes, enhancing human capabilities, and unlocking new opportunities, AI has the potential to redefine economic growth models. However, realizing these benefits requires careful management of associated risks and the equitable distribution of AI-driven gains.

5.1.2 AI as a Catalyst for Economic Transformation

AI plays a transformative role in facilitating economic growth by enhancing productivity, promoting innovation, and generating new market opportunities. Increasingly, governments and policymakers acknowledge AI as a substantial contributor to economic development, utilizing its capabilities to optimize industries and improve decision-making processes.

AI can significantly contribute to the automation of repetitive tasks, enabling human workers to reallocate effort towards higher value, creative, and strategic activities. This capability leads to substantial productivity improvements across various industries. Sectors such as manufacturing, logistics, and healthcare have already benefited from AI-driven efficiencies, including predictive maintenance, optimized supply chains, and enhanced diagnostic processes.

Beyond operational efficiencies, AI fosters the creation of novel products, services, and business models, thereby facilitating the emergence of entirely new markets. For example, generative AI expedites content creation, product design, and prototyping, thereby reducing the time-to-market for innovative solutions. Furthermore,

AI-driven analytics offer profound insights by discerning patterns and trends within extensive datasets, thus enabling more informed and timely decision-making for both businesses and policymakers.

The integration of AI into digital platforms, e-commerce, and financial technology can significantly enhance economic opportunities. AI enhances personalized customer interactions, strengthens fraud detection systems, and supports dynamic pricing strategies, thereby fostering competitive advantages within digital economies. Furthermore, AI-driven research accelerates scientific advancement, resulting in breakthroughs in fields such as drug discovery, renewable energy, and materials science.

5.1.3 Challenges and Risks Associated with AI-Driven Growth

AI has significant potential to drive economic growth, yet it also presents challenges that require careful consideration, particularly in relation to economic inequality. The benefits of AI are likely to accrue disproportionately to highly skilled individuals and technology intensive firms, while lower skilled workers may face heightened risks of displacement. Addressing these dynamics necessitates proactive policy responses, including investment in workforce reskilling and the development of robust social protection mechanisms to support a more equitable transition.

Furthermore, it is imperative to address the ethical and regulatory challenges associated with the adoption of AI. Concerns such as data privacy, algorithmic bias, and the ethical deployment of AI have the potential to erode public trust and impede its widespread implementation.

While AI generates new employment opportunities in fields such as AI development and data science, it can concurrently disrupt traditional sectors, leading to transitional unemployment. The ‘digital divide’ presents an additional challenge, as unequal access to AI technologies and digital infrastructure may exacerbate economic

disparities both between nations and within societies. Addressing these disparities requires coordinated international efforts to promote digital inclusion and equitable access to AI-driven innovations.

5.1.4 The Fragmented Global Governance Landscape

The business and geopolitical environment in which AI functions further complicates its responsible development and deployment. The three leading 'digital empires', namely China, the US, and the EU (Bradford, 2024a), have established distinct regulatory and governance frameworks, resulting in a fragmented global approach. This lack of a unified regulatory standard exacerbates complexity and operational challenges for organizations operating across different jurisdictions. Such fragmentation leads to inconsistent accountability measures and varied ethical standards, thereby hindering the global alignment necessary for promoting responsible AI.

While regulatory and governance frameworks are indispensable, they alone are insufficient to fully realize responsible AI. Although the landscape has demonstrated gradual improvement, the attainment of genuinely ethical AI development and deployment necessitates an organizational commitment to ethical conduct. This approach should strive to balance innovation with risk mitigation to ensure that AI is utilized for the benefit of all stakeholders, as well as for humanity as a whole.

5.1.5 The Centrality of Data in Responsible AI

AI has the potential to sustain and reshape economic growth through innovation, efficiency gains, and the creation of new opportunities. Ensuring that such growth is inclusive and aligned with societal values requires a balanced approach from policymakers and business leaders. Strategic investment in education, ethical AI governance, and digital infrastructure is therefore essential to maximise benefits while

mitigating associated risks, positioning AI as a driver of equitable and responsible economic development.

The evolving landscape of AI simultaneously generates opportunities and risks that are inseparable from the use of data. As AI systems increasingly depend on vast volumes of personal, organisational, and societal information, concerns relating to privacy, fairness, security, transparency, and governance have become more pronounced. Addressing these issues requires not only technical innovation but also robust governance frameworks, regulatory compliance, and ethical practices in data management. These dimensions of data form a critical foundation for responsible AI and will be examined in the following section.

5.1.6 Privacy and Data Protection

AI systems routinely process vast amounts of sensitive personal data, thereby posing significant privacy risks, particularly in domains such as healthcare, education, and surveillance (Dataguard, 2024; Hanna, 2025). Recent research underscores the susceptibility of student-generated content when such data are extracted from the internet or processed through unauthorized tools (Axios, 2025). At a systemic level, the incidence of AI-related privacy and security breaches increased by over 56 percent in 2024, indicating a marked escalation in AI-associated risk exposure (Stanford AI Index Report, 2025).

Mitigation requires embedding privacy by design throughout the AI lifecycle. This encompasses data minimization, anonymization, encryption, and stringent access controls (EDPB, 2025; Qualys, 2025). Organizations are increasingly implementing privacy impact assessments and transparency mechanisms to assure stakeholders of the protection of their data. These measures align with regulatory expectations, such as the

General Data Protection Regulation (GDPR) (European Commission, 2018), and contribute to maintaining trust in AI adoption (EDPB, 2025).

5.1.7 Algorithmic Bias and Fairness

Bias remains a significant concern as AI systems frequently inherit societal prejudices embedded within their training datasets. For example, gender bias has been documented in local government AI systems in the UK, where women's health issues were minimized in comparison to men's (The Guardian, 2025). Similarly, algorithmic hiring tools have demonstrated demographic skew, favouring certain demographic characteristics over others, despite measures intended to prevent discrimination (New York Post, 2025). Incidents such as the alleged algorithmic bias in State Farm's claims processing and wrongful arrests due to flawed facial recognition in Louisiana exemplify the tangible harms that can arise from the unchecked deployment of AI systems (Vincent, 2023; Hill, 2023).

Mitigation encompasses a combination of technical and organizational measures. Emerging techniques, such as affine concept editing, show promise in reducing demographic bias without compromising model performance (New York Post, 2025). Equally significant are mandatory fairness audits and bias testing, particularly in public service applications, where social outcomes can be materially affected (The Guardian, 2025). These strategies underscore the necessity for technical solutions to be bolstered by robust governance to ensure that fairness is consistently prioritized (Hanna, 2025).

5.1.8 Security Threats and Synthetic Fraud

AI systems are also vulnerable to deliberate manipulation through adversarial attacks, data poisoning and other forms of exploitation. Such vulnerabilities can undermine the reliability of safety-critical systems, including biometric authentication or traffic sign recognition (TTMS, 2025). A further dimension of risk arises from synthetic

fraud, encompassing deepfakes and fabricated identities, which threatens the stability of financial systems, e-commerce and digital communications (TechRadar, 2025a).

Mitigation necessitates a multi-layered security strategy. Defensive measures encompass the implementation of behavioural biometrics, explainable AI (XAI), and real-time monitoring. Furthermore, organizations are increasingly utilizing synthetic data within controlled training environments to minimize vulnerability to malicious manipulation while preserving system robustness (TechRadar, 2025a). These initiatives contribute to the concept of 'synthetic resilience,' an emerging paradigm in AI cybersecurity.

5.1.9 Transparency, Shadow AI and Oversight

The opacity inherent in AI decision-making processes continues to constrain accountability. The phenomenon of 'Shadow AI,' characterized by the unregulated utilization of generative tools by employees outside established organizational governance frameworks, amplifies the risks associated with data leakage, bias, and regulatory non-compliance (TechRadar, 2025b). This lack of transparency not only undermines stakeholder confidence but also hinders effective oversight (Washington Post, 2025).

Addressing this challenge requires both technical and institutional responses. Strategies such as 'deep ignorance,' which involve the exclusion of sensitive or hazardous data from training datasets, present promising methods to mitigate risk without substantially compromising performance (Washington Post, 2025). Equally, organisations must enforce clear governance protocols and ensure that human oversight is preserved in critical domains such as healthcare and social care (The Guardian, 2025).

5.1.10 Governance, Regulation and Institutional Gaps

Despite progress, governance and regulation of AI remain fragmented. In the UK, the Competition and Markets Authority (CMA) has established principles for foundation models and is expected to gain enhanced statutory powers under new AI legislation (TIME, 2025). In contrast, the United States has predominantly relied on the enforcement of existing consumer protection and privacy laws, with state Attorneys General addressing the regulatory gap (Reuters, 2025). Globally, the uneven pace of regulatory development creates uncertainty for organisations operating across multiple jurisdictions.

Effective mitigation necessitates the enhancement of governance frameworks and the closure of regulatory gaps. This entails the integration of standards pertaining to fairness, documentation, transparency, and human oversight, all of which should be supported by regular auditing and monitoring (OpenForum, 2025). Institutional oversight must evolve in tandem with technological advancements to ensure that societal risks are not externalized while economic benefits are privately accrued.

5.1.11 Integrative Perspective

The risks associated with data in AI, spanning privacy, bias, security, opacity and governance, are deeply interconnected. For example, failures in transparency compound risks of bias and privacy breaches, while weak governance amplifies vulnerability to synthetic fraud. Mitigation, therefore, cannot be viewed as a set of discrete interventions. Instead, it requires an integrated approach combining technical safeguards, organisational culture, and regulatory enforcement. Privacy concerns can be addressed through privacy-by-design principles, data minimisation, encryption, and rigorous auditing processes (EDPB, 2025; Qualys, 2025).

Algorithmic bias and unfairness demand fairness frameworks, systematic bias testing, and the application of advanced technical mitigation techniques (Hanna, 2025;

New York Post, 2025). Security threats, including adversarial attacks and synthetic fraud, call for the deployment of explainable AI, behavioural biometrics, real-time monitoring, and synthetic resilience approaches (TechRadar, 2025a; TTMS, 2025). Challenges linked to opacity and shadow AI highlight the importance of strong governance frameworks, oversight mechanisms, and the adoption of approaches such as filtering sensitive training data (TechRadar, 2025b; Washington Post, 2025). Finally, regulatory gaps necessitate both the strengthening of institutional oversight and the adaptation of existing legal frameworks to ensure enforceability across jurisdictions (Reuters, 2025; TIME, 2025).

Collectively, these strategies demonstrate that responsible AI cannot be achieved through isolated interventions but rather through integrated practices that combine technical defences, organisational governance, and regulatory enforcement (OpenForum, 2025).

5.1.12 Case Studies in Technology Governance Failures and Lessons for AI Regulation

Technological innovation often advances at a faster pace than governance frameworks, creating opportunities for breakthroughs but also exposing significant risks when validation, transparency, and accountability are insufficient. The following three case studies illustrate how governance gaps, coupled with organisational and ethical failings, can lead to harmful outcomes. Although each occurred in different sectors, all present cautionary lessons highly relevant to the governance of AI.

1. Cambridge Analytica and Facebook: Data Exploitation and Consent Evasion

The Cambridge Analytica scandal exposed significant deficiencies in digital data governance, particularly in relation to consent, transparency, and the ethical use of data. In 2014, the political consulting firm Cambridge Analytica, through a third-party

academic, collected data from Facebook users via a personality quiz application. Although only approximately 270,000 individuals consented to participate, Facebook's permissive application programming interface (API) facilitated the collection of data from their friends, resulting in the unauthorized acquisition of personal information from an estimated 50–87 million users (Cadwalladr and Graham-Harrison, 2018).

The data were used to create psychographic voter profiles, which were subsequently deployed in political campaigns including the Brexit referendum in the UK and the 2016 presidential election in the United States. Public disclosure in 2018 triggered global debates on data privacy, political manipulation, and the ethical limits of algorithmic profiling. For AI governance, the case underscores the need for explicit consent mechanisms, data minimisation, and strict limitations on secondary data use. Current proposals such as the EU AI Act and the UK's Data Protection and Digital Information Bill reflect these lessons by mandating transparent data provenance and compliance with privacy-by-design principles (European Commission, 2024; Department for Science, Innovation and Technology, 2023).

2. Theranos: Unvalidated Technology and Opaque Operations

Theranos's collapse offers a cautionary example of the dangers posed when technological claims outpace scientific validation. Founded in 2003, the company claimed to have developed a device capable of running hundreds of blood tests from just a few drops of blood. In practice, the technology was inaccurate, and the majority of tests were conducted using commercially available machines rather than the company's proprietary device (Carreyrou, 2018).

The concealment of these failings from investors, regulators, and patients resulted in significant financial and reputational damage, alongside the risk of harm to individuals receiving incorrect medical results. The parallels with high-risk AI applications are clear:

both involve complex, safety-critical systems whose performance is difficult for outsiders to verify. Provisions in the EU AI Act requiring pre-market conformity assessments, post-market monitoring, and independent auditing are designed to prevent the kind of opaque, unvalidated deployment seen in Theranos (European Commission, 2024). The Federal Trade Commission (FTC, 2023) in the United States has also emphasised that exaggerated or false claims about AI capabilities may constitute deceptive marketing, echoing the enforcement actions taken against Theranos.

3. Boeing 737 MAX: Safety Compromises and System Design Failures

The Boeing 737 MAX crisis exemplifies the severe repercussions that can arise from insufficient safety validation and inadequate risk communication. Two air accidents, occurring in Indonesia in 2018 and Ethiopia in 2019, resulted in the loss of 346 lives and were attributed to the Manoeuvring Characteristics Augmentation System (MCAS), a software mechanism designed to enhance aircraft stability. The MCAS depended on input from a single angle-of-attack sensor, lacking adequate redundancy, and had the potential to repeatedly force the aircraft's nose downward in response to erroneous data (Joint Authorities Technical Review, 2019).

Investigations revealed that Boeing had minimised regulatory scrutiny by presenting MCAS as a minor modification, avoiding extensive pilot training requirements. Internal communications indicated awareness of the risks but a prioritisation of cost and market deadlines over safety (US House of Representatives, 2020). In the context of AI governance, the 737 MAX case underlines the critical importance of system transparency, human oversight, and robust testing under realistic operational conditions. The focus on managing risks associated with high-risk systems and the monitoring of these systems post-deployment, as outlined in the EU AI Act, mirrors the regulatory reforms eventually adopted in the aviation sector. This parallel

underscores the necessity for rigorous and continuous safety assurance in software-driven systems, whether they pertain to aircraft or AI.

4. Cross-Case Analysis and Implications for AI Governance

Despite their sectoral differences, these three cases share key governance failures: lack of transparency, inadequate validation, insufficient independent oversight, and weak post-deployment monitoring. In each instance, the consequences were amplified by organisational incentives that prioritised speed, market advantage, or political influence over ethical responsibility and public safety.

The EU AI Act, the OECD AI Principles, and emerging frameworks in the US and the UK address these deficiencies through several measures, including: pre-market testing and independent conformity assessments for high-risk AI systems; mandatory transparency regarding data sources, system capabilities, and limitations; human oversight obligations to prevent unmitigated autonomous operation; and continuous monitoring and mandatory incident reporting to regulatory authorities.

By incorporating these safeguards, AI governance can mitigate the risk of similar failures, thereby maintain public trust and ensure that innovation does not compromise safety, fairness, or democratic integrity. The cases of Cambridge Analytica, Theranos, and the Boeing 737 MAX serve as compelling evidence for the necessity of robust AI regulation and comprehensive governance, highlighting that such measures are not merely precautionary but essential.

5.2 Mitigation

5.2.1 Artificial Intelligence Explainability: Technical and Ethical Dimensions

The rapid advancement of AI, particularly machine learning (ML), has led to its increasing deployment in critical systems and digital decision-support tools. The growing reliance on AI in such domains necessitates a clear understanding of its decision-making

processes to ensure trust, safety, and accountability. Explainability, often referred to as explainable AI (XAI), is a pivotal concept in this context, bridging the gap between technical functionality and ethical responsibility.

XAI methods aim to make ML models more interpretable and transparent, allowing stakeholders to understand how AI-driven decisions are made. These methods can be broadly categorised into feature-importance techniques, example-based approaches, and model-specific versus model-agnostic methods. By explaining the internal logic of AI systems, XAI contributes to building assurance and justified confidence in AI applications.

Stakeholders seek explanations of AI systems for multiple reasons, including regulatory compliance, confidence assessment, decision contestation, and ethical assurance. These motivations align closely with broader ethical principles such as fairness, accountability and transparency. As McDermid et al. (2021) observe, the drivers of XAI are inherently ethical, as they enable responsible governance by tracing explanations back to human choices made during design and implementation. XAI should therefore be understood within a wider accountability ecosystem that integrates both technical and ethical imperatives.

The range of stakeholders extends from regulators and service providers to expert and end users, each requiring explanations for interconnected but distinct purposes. These can be conceptualised as five cyclical objectives. Clarity enhances comprehensibility for non-expert users and underpins compliance with legal and regulatory frameworks. Compliance strengthens confidence by assuring stakeholders that systems operate within accepted norms. Confidence supports consent and control, as individuals are more willing to exercise informed oversight when trust is established. The ability to challenge AI-generated outcomes ensures accountability and fairness, which in turn reinforce both

confidence and compliance. Continuous improvement depends on feedback from these processes to refine models, thereby generating greater clarity and restarting the cycle. Internal stakeholders, such as developers and safety engineers, are primarily concerned with validation and performance assurance, while external stakeholders, including regulators and affected individuals, emphasise explainability as a foundation for accountability, fairness, and ethical decision making.

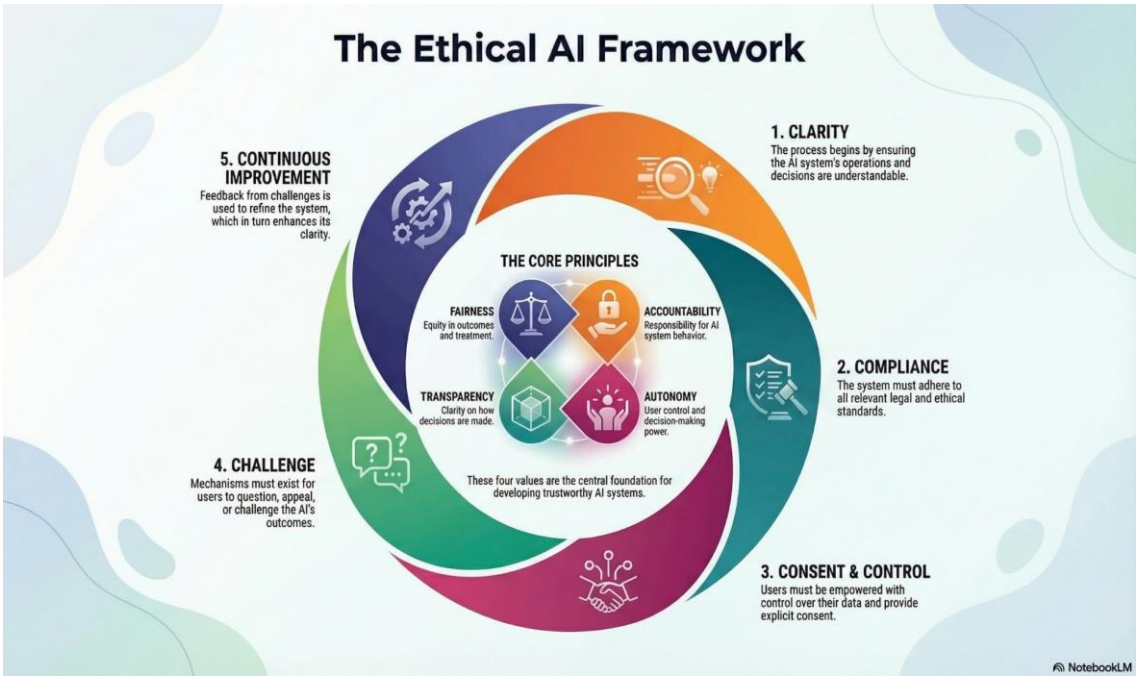


Figure 5.1: Ethical Dimensions and Cyclical Objectives of AI Explainability

Figure 5.1 illustrates the relationship between four overarching ethical dimensions, fairness, accountability, transparency, and autonomy, and five cyclical objectives of explainability: clarity, compliance, consent and control, challenge, and continuous improvement. The model highlights how ethical imperatives inform stakeholder needs for AI explanations, while cyclical objectives ensure that explainability remains dynamic, iterative, and responsive to both technical and societal expectations.

5.2.2 Ethical and Regulatory Considerations

Explainability is increasingly recognised as a key component in AI governance frameworks. Regulations such as the General Data Protection Regulation (GDPR), the EU AI Act, and sector-specific AI guidelines emphasise transparency and accountability in AI decision-making. However, while XAI methods enhance interpretability, they do not inherently provide full normative justifications for AI-driven outcomes. Ethical concerns persist regarding the potential manipulation of explanations, misleading interpretations, and an over-reliance on approximate reasoning.

Furthermore, the integration of XAI into regulatory structures remains an evolving challenge. There is limited discussion on how XAI methods align with emerging governance frameworks such as the EU AI Act or US AI policy guidelines. Further research is needed to explore the effectiveness of XAI in identifying and mitigating biases within AI models and datasets, as well as to assess the cognitive biases influencing human interpretation of AI-generated explanations.

5.2.3 The Future of AI Explainability

Explainability is essential for fostering trust in AI, particularly in domains such as healthcare, finance, and law enforcement where automated decisions have significant ethical and societal impact. Advances in explainable AI (XAI) can strengthen human–AI collaboration by promoting alignment with ethical values, mitigating discrimination, and enhancing accountability.

XAI is also critical for transparency and compliance. By clarifying how decisions are reached, it provides assurance to both regulators and stakeholders (Guidotti et al., 2018). Regulatory expectations are increasingly stringent as governments legislate against bias and discriminatory practices. For example, New York City requires independent audits of automated hiring systems, while the European Union's AI Act sets

comprehensive requirements for ‘high-risk’ applications across multiple sectors (European Commission, 2021).

Future research should prioritise empirical evaluation, user experience analysis, and legal study to refine methods and ensure their effective application. Addressing these challenges is essential to ensuring that AI systems are not only technically robust but also ethically responsible and socially legitimate. Embedding XAI within broader governance frameworks will enable policymakers, developers, and regulators to advance AI that is transparent, trustworthy, and aligned with human-centred principles.

5.2.4 Responsible AI

Responsible AI (RAI) offers a structured approach that aligns AI development and deployment with ethical, transparent, and human-centred values. Responsible AI can be defined as the practice of designing and implementing AI systems in ways that are fair, transparent, accountable, and consistent with societal values (Jobin et al., 2019). Robust data governance is essential to reduce bias and error by ensuring the quality, traceability, and representativeness of training data. Effective governance can reduce these disparities by embedding quality controls and continuous monitoring.

Stakeholder communication fosters transparency, signalling to users and buyers where AI is deployed and clarifying data provenance and model implications. Board-level engagement and cross-functional collaboration ensure that responsibility does not rest solely with technical teams but is embedded across leadership and organisational functions (Holistic AI, 2022). A commitment to human-centred AI requires designing systems that augment rather than replace human capacities, with emphasis on feedback loops and usability (Shneiderman, 2020).

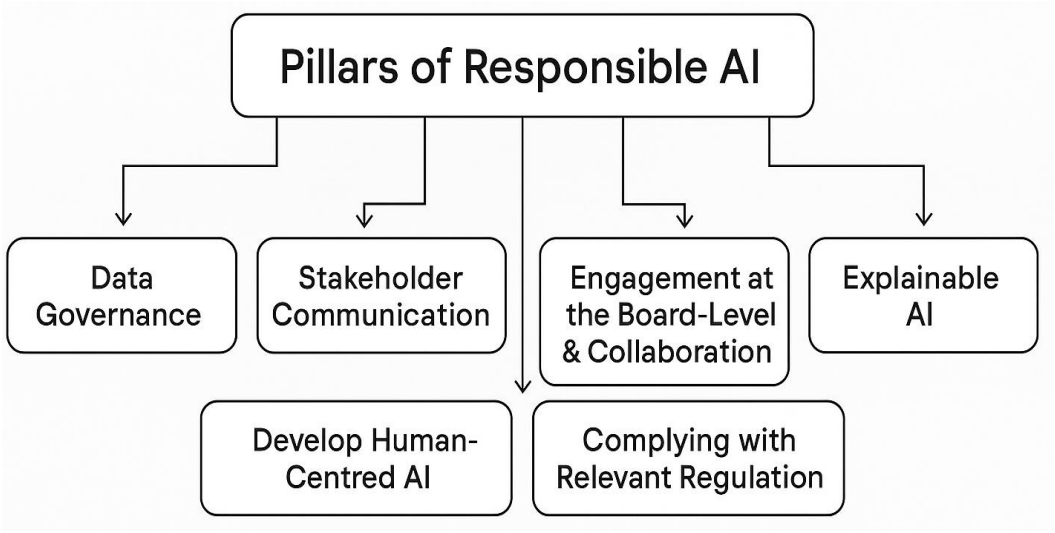


Figure 5.2: Pillars of Responsible AI

It is crucial that this process does not become a mere procedural formality, as such an approach would diminish its potential advantages. Instead, it should consistently be perceived as a contextual human endeavour, wherein individuals have the capacity to influence ethical practices by embracing care, accountability, and responsibility.

A study by the Infosys Knowledge Institute found that only 2% of companies met responsible AI (RAI) standards (Infosys Knowledge Institute, 2022). The report concludes that the failure to integrate RAI into environmental, social, and governance (ESG) strategies creates a substantial blind spot, exposing organisations to both financial and reputational risks. Notably, 77% of businesses reported financial losses as a result of poorly implemented AI. By contrast, companies that have invested in RAI leadership practices reported significant advantages, including 39% lower incident costs and reduced severity of risks. These findings suggest that well-implemented RAI standards should not be regarded merely as a compliance requirement but as a strategic driver of organisational resilience and sustainable growth.

This argument is reinforced by Holistic AI, which highlights that embedding RAI principles such as data governance, explainability, and board-level engagement helps

organisations anticipate and mitigate harms while also creating competitive advantage (Holistic AI, 2022). Similarly, the OECD (2021) stresses that trust in AI is fundamental to realising its economic and societal benefits, with responsible governance frameworks necessary to ensure safety, fairness, and accountability. Together, these studies underscore that RAI is increasingly central not only to ethical practice but also to long-term financial sustainability and stakeholder trust.

5.2.5 Extending Porter's Five Forces: Responsible AI Capability and Stewardship

Porter's Five Forces framework, introduced by Michael E. Porter in 1979, is a strategic tool used to analyse the competitive structure of an industry. It helps organisations assess the intensity of competition and thus the attractiveness or profitability of entering or remaining in a given market (Porter, 1979; 2008). The five forces are:

1. **Threat of New Entrants** – New competitors can enter the market and erode profitability unless there are substantial barriers to entry such as high capital requirements, strong brand identities, or regulatory constraints.
2. **Bargaining Power of Suppliers** – Powerful suppliers can demand higher prices or reduce the quality of goods and services, affecting profitability.
3. **Bargaining Power of Buyers** – When buyers have significant power, they can force prices down or demand higher quality, squeezing margins.
4. **Threat of Substitute Products or Services** – The presence of alternatives can limit the price companies can charge and drive innovation.
5. **Industry Rivalry** – Intense competition among existing players can lead to price wars, increased marketing costs, and the need for continuous innovation (Porter, 2008).

5.2.6 Contemporary Relevance

While sectors such as technology are predominantly influenced by network effects, ecosystem strategies, and rapid innovation cycles rather than traditional barriers to entry or supplier power (Teece, 2018; Jacobides, Cennamo, and Gawer, 2018), they nonetheless offer the opportunity to explore new dynamics enabled by enterprises implementing AI.

Porter's Five Forces framework traditionally examines external competitive pressures, including suppliers, buyers, substitutes, new entrants, and industry rivalry. Nevertheless, the utilization and implementation of AI presents an opportunity to develop a distinct internal strategic competence dimension, which can fundamentally influence shape an organisation's position in its market. This sixth force should reflect an organisation's capacity for trustworthy and responsible AI adoption, in other words, its organisational AI integrity.

Porter's Five Forces model (Porter, 2008) explains how industry structure and external pressures shape organisational competitiveness. Yet, as AI becomes embedded in business strategy, there is increasing recognition that internal competencies, particularly those linked to responsible and trustworthy AI, are equally decisive in strengthening or weakening its competitive position. To address this gap, a sixth force is proposed: *Responsible AI Capability and Stewardship* as shown in Figure 5.3.

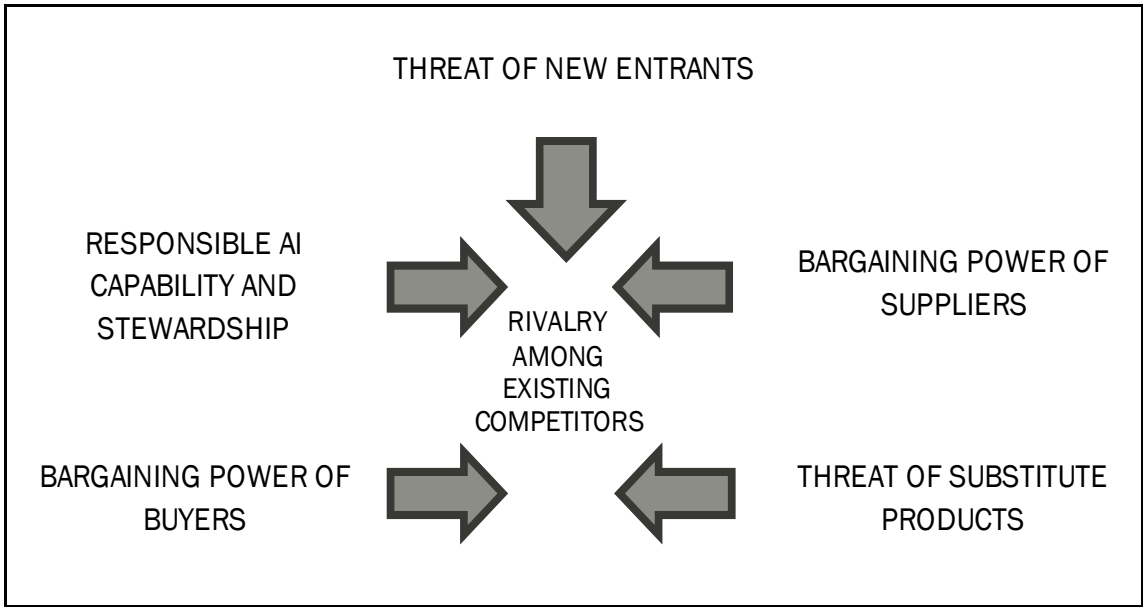


Figure 5.3: Porter's Five Forces with a 6th Force Added

Responsible AI Capability and Stewardship reflects the extent to which an organisation ensures robust data practices, compliance with evolving regulation, effective governance, ethical culture, and leadership competence in managing AI responsibly. These attributes are not only compliance enablers but also differentiators that foster resilience and competitive advantage (Floridi, 2022; Mökander and Floridi, 2023).

The necessity of proposing Responsible AI Capability and Stewardship as a distinct competitive force stems directly from the empirical finding that failure to operationalize ethical intent creates profound organizational vulnerability. While Porter's model traditionally focuses on external competitive pressures, the integration of AI systems means that internal competence, culture, and governance are now decisive factors in organizational risk and resilience.

This internal capability is critical because the integration of AI introduces hazards that traditional governance structures often fail to manage, as evidenced by significant technology governance failures. Cases like the Boeing 737 MAX crisis, which resulted from a prioritisation of cost and market deadlines over rigorous safety validation,

illustrate the catastrophic consequences when systemic governance is insufficient in software-driven, safety-critical systems. Similarly, failures in data governance, such as the Cambridge Analytica scandal, expose the severe regulatory and reputational damage associated with a lack of transparency and insufficient ethical stewardship.

Failure to embed Responsible AI principles exposes organisations to a range of risks, including regulatory sanctions, financial penalties, reputational harm, and economic loss. Empirical evidence suggests that organisations lacking mature Responsible AI practices experience higher incident related costs, whereas those investing in Responsible AI leadership report substantially lower financial impact, with reductions of up to 39% in AI related incident costs (Infosys Knowledge Institute, 2025).

Therefore, Responsible AI Capability and Stewardship is not merely an ethical commitment or a compliance overhead; it is a strategic differentiator that fundamentally determines a firm's ability to maintain legitimacy and trust within its digital ecosystem. This force reflects the reality that sustainable competitive advantage is increasingly governed by an organization’s capacity to ensure robust data practices, compliance with evolving regulation, ethical culture, and leadership competency in managing AI responsibly. By integrating this sixth force, the model fully accounts for the contemporary challenges where internal ethical integrity is essential for achieving long-term resilience and sustained competitive success.

This sixth force complements the original framework by influencing each of Porter’s existing dimensions. For suppliers and buyers, integrity strengthens bargaining power and loyalty, as organisations able to demonstrate fairness, security, and accountability inspire greater trust (Jobin, Ienca and Vayena, 2019). Regarding new entrants, robust governance and ethical leadership create reputational and compliance barriers that raise the threshold for market access, particularly under emerging regulatory

regimes such as the EU AI Act (European Commission, 2021). In relation to substitutes, responsible and transparent AI deployment reduces the likelihood of displacement by alternatives positioned as more trustworthy. Within competitive rivalry, integrity becomes a strategic differentiator: while irresponsible firms may seek short-term gains, they are more exposed to regulatory sanctions, reputational backlash, and erosion of legitimacy (Cath, 2022).

This new force therefore recognises that sustainable advantage in AI-intensive industries is increasingly shaped by an organisation's ability to manage ethical and governance challenges proactively. By embedding responsible AI into corporate strategy, firms build resilience against regulatory uncertainty and stakeholder scepticism, while positioning themselves as trustworthy partners in digital ecosystems (McKinsey, 2023). This aligns with scholarly arguments that accountability and transparency are not only ethical imperatives but also sources of long-term strategic strength (Floridi, 2022).

In short, Responsible AI Capability and Stewardship strengthens Porter's model by incorporating internal responsibility and stewardship as a decisive force in modern competition. It reflects the reality that competitive success in AI-driven markets depends as much on trust, governance, and ethical leadership as on managing the traditional external forces.

5.2.7 Corporate Culture, Leadership Competency and Responsible AI

Corporate culture and leadership competency are critical determinants of whether organisations can effectively develop and deploy responsible and ethical AI. Culture shapes the norms, values and behaviours that guide how technology is adopted and applied, while leadership competency provides the vision, skills and accountability to align AI initiatives with ethical and risk-aware practices (Schein, 2017; Bass and Avolio, 1994).

A culture that prioritises transparency, inclusivity and accountability creates the conditions for responsible AI, ensuring that ethical principles such as fairness, human oversight and privacy are embedded into decision-making (Floridi et al., 2018). In contrast, weak cultural foundations can exacerbate risks, including algorithmic bias, opaque decision-making and reputational harm. Leadership plays a pivotal role in shaping this culture by setting expectations, fostering cross-disciplinary collaboration, and embedding responsible AI standards into governance structures (Wilson and Daugherty, 2018).

Leadership competency in this context extends beyond technical knowledge to encompass ethical literacy, regulatory awareness and the ability to manage complex risks. Research shows that organisations where executives actively champion responsible AI are better positioned to mitigate harms and build stakeholder trust (OECD, 2021). This aligns with the growing consensus that AI governance is not solely a technical challenge but a strategic leadership imperative (Jobin, Ienca and Vayena, 2019).

By cultivating a responsible corporate culture and enhancing leadership competency, organisations can integrate responsible and ethical AI into their operations in ways that reduce risk and promote resilience. These factors are therefore not peripheral but central to the mitigation of AI risks and to the sustainable adoption of AI technologies.

5.3 Awareness, Training and Framework

5.3.1 Introduction

The findings of this research support and extend the existing literature on responsible AI by emphasising the pivotal role of leadership in guiding ethical and governance practices. Scholars such as Dignum (2019) and Morley et al. (2021) have argued that effective AI governance requires not only robust technical safeguards, but

also strong organisational leadership grounded in ethical principles. This study corroborates that claim, revealing that while many leaders express ethical concerns regarding AI systems, they lack the structured knowledge or frameworks to act with confidence. The concept of AI literacy, emerging in the work of Long and Magerko (2020), becomes especially relevant here, as participants highlighted gaps in understanding key AI concepts, such as bias, explainability, and algorithmic accountability, at senior decision-making levels. Furthermore, although policy instruments like the EU Artificial Intelligence Act (European Commission, 2021) aim to enforce responsible AI practices, this research suggests that compliance-driven approaches alone are insufficient without concurrent leadership development and cross-functional collaboration. This aligns with findings by Fjeld et al. (2020), who argue that ethics principles must be operationalised within organisations through culture, training, and leadership support. The present study contributes to this discourse by offering empirical insight into the disconnect between high-level awareness and everyday decision-making in AI implementation, underscoring the need for leadership-centred frameworks for responsible AI.

5.3.2 Commentary on First Survey Findings

The findings presented in the previous section and recommendations presented herein are derived from the results of the initial quantitative survey and emphasize the establishment of a systematic approach to formulating ethical guidelines, training personnel, and promoting transparency and stakeholder engagement.

The analyses provided a foundation for several key recommendations aimed at enhancing ethical practices within organizations, particularly in relation to AI usage. A primary recommendation is the development of clear and comprehensive ethical guidelines to govern the application of AI technologies. These guidelines would serve as

a framework to ensure that AI is utilized responsibly and in alignment with organizational values.

Regular training programs on AI ethics are also emphasized as a critical measure. Such initiatives seek to equip employees with the knowledge and competencies required to address ethical dilemmas arising from the use of AI systems. Furthermore, enhancing transparency in AI decision-making processes is identified as a vital step toward building trust and accountability in AI-driven environments.

The establishment of robust mechanisms for reporting and addressing unethical AI practices is another significant recommendation. These mechanisms should be accessible and effective, enabling organizations to identify and rectify potential ethical breaches promptly. Lastly, engaging stakeholders in meaningful discussions about AI ethics and transparency is highlighted as a strategy for fostering a culture of inclusivity and shared accountability. Collectively, these recommendations underscore the importance of a structured and proactive approach to ethical governance, training, and stakeholder involvement in the evolving landscape of AI.

5.3.3 Implementation Strategies

An organisation's approach to ethics can be aligned with their core values, history, industry and external factors. That is, it will be unique to each organisation and therefore cannot be predefined or subject to a 'one size fits all' approach. Effective implementation of ethical and responsible AI requires coordinated structures and processes. One approach is to establish a dedicated task force responsible for the development, monitoring, and continual refinement of ethical guidelines. Regular workshops and training sessions should be introduced to ensure that employees at all levels develop awareness and competence in AI ethics. Transparency can be promoted through structured reporting mechanisms that document AI decision-making processes,

enabling scrutiny and accountability. To address potential risks of malpractice, organisations should also provide confidential channels for reporting unethical practices without fear of reprisal. Stakeholder forums can serve as an important platform for dialogue, allowing diverse perspectives to inform policy development and ensuring that ethical frameworks remain responsive to evolving expectations.

Organisations that successfully embed and sustain an ethical culture are better positioned to secure competitive advantage in the context of AI adoption. The field of organisational development provides a means of integrating ethical education and practice as a holistic and strategic intervention, strengthening both organisational capability and cultural coherence. This contrasts with isolated or unstructured initiatives, whose effects are often short lived and insufficient to support enduring ethical change.

Organizations that lack ethical alignment risk violating regulations, alienating their customers, and damaging their brand reputation. Consequently, this may adversely affect their share price, market share, and overall ability to thrive in their respective marketplace.

5.3.4 Metrics for Evaluation

The evaluation of ethical AI practices requires clear and measurable indicators. Employee awareness and understanding of AI ethics can be assessed through regular surveys, providing insight into the effectiveness of training and communication strategies. The number and quality of ethical guidelines developed and disseminated reflect organisational commitment to responsible practice. The frequency and reach of training sessions offer a measure of institutional investment in capability building. Reporting rates of unethical practices indicate both employee confidence in available mechanisms and the extent to which accountability processes are functioning. Finally, stakeholder participation in ethical discussions demonstrates the inclusiveness of

governance structures and the responsiveness of organisations to diverse perspectives. Together, these metrics provide a basis for continuous monitoring and improvement of AI ethics initiatives.

5.3.5 Theoretical and Practical Implications

The findings of this study carry important implications for both theory and practice. Theoretically, the results extend current leadership and organisational learning literature by positioning AI literacy as a foundational leadership competency, essential for engaging with the ethical and strategic dimensions of AI adoption. This builds on the work of Argyris and Schön (1996), who advocate for double-loop learning in complex environments, by suggesting that responsible AI implementation requires leaders to question not just outcomes, but the underlying values, assumptions, and governance mechanisms guiding AI use. Practically, the study highlights an urgent need for structured leadership development programmes that address ethical reasoning, cross-disciplinary communication, and AI-related decision-making under uncertainty. Such programmes would go beyond technical training and equip leaders with the tools to engage in value-based dialogue and risk-based oversight, essential elements of responsible AI governance (Fjeld et al., 2020; Gasser and Almeida, 2017). Furthermore, organisations should integrate AI ethics and governance into corporate strategy, ensuring that responsibility is distributed across executive, legal, and technical domains. In doing so, AI adoption becomes not merely a question of innovation or compliance, but a leadership-driven process grounded in trust, accountability, and organisational integrity.

5.3.6 Critical Reflection and Limitations

While this study offers valuable insights into leadership readiness for responsible AI, several limitations must be acknowledged. First, the research was based on a purposive sample of leaders predominantly drawn from organisations in the UK and

Europe, which may limit the generalisability of findings to global contexts with differing regulatory, cultural, or technological infrastructures. Second, although participants occupied senior roles, their responses may reflect perceived organisational positions rather than personal convictions or actions, raising the possibility of social desirability bias (Fisher, 1993). Additionally, given the evolving nature of AI policy, particularly developments like the EU AI Act, findings represent a snapshot in time and may not fully capture shifting regulatory awareness or organisational adaptation. Methodologically, the quantitative components provided rich data, but due to time and resource constraints, the sample size was modest, limiting the breadth of sectoral comparison. Finally, the study focused on leadership perceptions rather than empirical measurement of AI deployment outcomes or the effectiveness of governance practices. Future studies could address this by triangulating leader perspectives with organisational policy documents, AI implementation metrics, or staff-level data to provide a more holistic view. Despite these limitations, the study offers a timely and relevant contribution to the discourse on AI ethics, leadership competence, and organisational readiness.

Cultural variations exert a substantial influence on ethical considerations, with geographical and demographic contexts shaping individuals' responses. Analysing these differences can provide valuable insights into how cultural norms and societal values affect perceptions of ethics in various settings. This perspective is essential for understanding the broader implications of ethical behaviour and transparency, particularly in the context of AI.

Practical challenges in implementing ethical frameworks and AI transparency require further investigation. Examining real-world trade-offs and the operational difficulties of maintaining ethical standards in practice can illuminate the discrepancies between theoretical principles and their application in diverse contexts. Moreover, ethical

concerns often manifest differently across industries and job functions, underscoring the need for industry-specific analyses. Such examinations can uncover nuanced ethical dilemmas unique to particular sectors and professional environments.

A persistent tension exists between ethical decision making and the pursuit of business objectives. Examining how individuals and organisations navigate this balance offers important insight into how values are prioritised within commercial contexts. The subjective nature of ethics, shaped by personal experience and demographic factors, further contributes to divergent and sometimes conflicting interpretations of acceptable conduct, as illustrated by high profile organisational failures such as Cambridge Analytica, Theranos, and the Boeing 737 MAX.

Organizational complexity further complicates ethical decision-making. The structures and cultures of organizations play a pivotal role in shaping the ethical actions of their members. Analysing these influences could provide a framework for fostering ethical behaviour at all levels. Equally significant is the evaluation of training programs aimed at promoting ethical leadership and behaviour, as their efficacy directly impacts the development of ethical competencies.

Finally, the mechanisms for reporting unethical behaviour require careful scrutiny. The accessibility and efficacy of these systems are vital for fostering accountability and transparency within organizations. Longitudinal studies tracking changes in ethical attitudes over time can offer a dynamic perspective on how perceptions and practices evolve, contributing to a more comprehensive understanding of ethics in an ever-changing global landscape.

In conclusion, the mechanisms for reporting unethical behaviour warrant thorough examination. The accessibility and effectiveness of these systems are crucial for promoting accountability and transparency within organizations. Longitudinal studies

that monitor changes in ethical attitudes over time can provide a dynamic perspective on the evolution of perceptions and practices, thereby contributing to a more comprehensive understanding of ethics in an ever-evolving global context.

5.3.6.1 The Researcher's Role and Reflexivity

As a doctoral study grounded in the interpretive tradition and conducted within the framework of a Business Administration (DBA) thesis, this research adopts a practitioner-scholar stance. This perspective dictates that the analysis and interpretation of the findings are necessarily informed by the researcher's professional background and engagement in organizational leadership and technological governance.

This reflexive approach was critical for framing the research questions, designing the constructs for the quantitative surveys, and, most importantly, interpreting the results. The finding of a significant readiness gap and the observation that responsibility for AI is frequently siloed in technical or compliance departments, were interpreted not merely as statistical patterns but as symptoms of a deeper, systemic failure in organizational structure and culture. This interpretation is informed by the researcher's experience in navigating the discrepancy between high-level ethical aspirations and the practical difficulties of implementation within commercial settings.

By actively engaging in reflexivity, the study acknowledges that the lens of professional practice influenced the focus on actionable literacy, and the design of the TRAIL framework (in section 5.13) as a strategic, cross-functional intervention, rather than a purely technical or compliance-driven solution. This transparency ensures that while the findings are grounded in empirical data, the theoretical and practical implications are presented with an awareness of the inherent subjectivity involved in interpreting complex organizational change dynamics. This practice enhances the

trustworthiness of the conclusions and the relevance of the proposed TRAIL framework for the target audience of senior organizational leaders.

5.3.7 Pathways for Future Research

This study highlights several avenues for future research in the field of responsible AI and leadership. First, there is a clear need for longitudinal studies that examine how leadership development interventions influence responsible AI adoption over time, particularly in terms of ethical governance and cultural integration. Second, future research could explore sector-specific variations in AI governance, comparing industries with mature regulatory oversight (such as finance or healthcare) to those with looser or emerging standards (such as creative industries or retail). Additionally, given the study's findings on uneven levels of AI literacy across senior roles, future work could investigate the impact of targeted executive training programmes, using both qualitative and quantitative measures of leadership competency and organisational AI maturity (Huang et al., 2020). Comparative studies across national and regional contexts would also enrich the literature, especially as differing legal frameworks such as the EU AI Act, the U.S. AI Bill of Rights, and China's AI ethics guidelines continue to shape practice and perception (Wirtz, Weyerer and Geyer, 2019). Finally, there is scope for further empirical testing of the framework proposed in this thesis, including its adaptability, usability, and relevance in real-world organisational settings. Such studies could offer robust validation and refinement, ensuring that leadership in AI remains not only informed but genuinely responsible.

5.4 Discussion of Second Survey Findings

5.4.1 Introduction

The findings from the second survey address the third research question of this study, which pertains to how leaders in commercial organizations perceive their

preparedness, responsibilities, and ethical obligations concerning AI. The analysis sought to interpret the data collected from survey respondents, who are senior professionals and decision-makers, within the broader theoretical framework of responsible AI adoption and AI literacy.

The results of the survey provides empirical support for a central proposition of this study: that AI literacy is swiftly becoming a defining characteristic of contemporary leadership, particularly for organizations operating on an international scale. Consequently, the increasing integration of AI into global business operations necessitates a comprehensive re-evaluation of leadership competencies and development frameworks.

The results help explain an evolving leadership paradigm; wherein traditional business acumen must be augmented by technological fluency, ethical foresight, and regulatory awareness. The findings also highlight a readiness gap: leaders are assuming responsibility for AI integration at a pace that surpasses their level of preparation. Furthermore, while there is widespread endorsement of fairness and human dignity in AI-related decisions, a disconnect persists between ethical intent and practical competence.

The implications of the findings extend to both theoretical and practical domains. They highlight the urgent need for comprehensive leadership development that integrates governance, ethics, and strategic foresight. Leadership roles are evolving in response to the challenges of AI, requiring education and training that equip leaders to implement technology responsibly, navigate regulatory complexity, and foster public trust.

Organisational culture should also adapt to embed ethical considerations into decision making, while strategic planning should prioritise the alignment of technological innovation with societal values. Together, these insights underscore the interdependence

of leadership education, governance structures, and cultural transformation in shaping responsible AI adoption.

The findings of this study reinforce the growing body of literature that views responsible AI not as a purely technical construct but as a socio-organisational endeavour grounded in leadership decision-making and ethical accountability. For instance, Dignum (2019) emphasises the need for AI systems to be embedded within human values and institutional practices, a point echoed by participants in this study who highlighted the absence of clear organisational frameworks for aligning AI use with corporate ethics. Furthermore, while prior research by Floridi et al. (2018) has outlined the principles of ethical AI, such as 'beneficence, non-maleficence, autonomy, and justice', this study found limited evidence that such principles have been operationalised at leadership levels. Instead, AI governance remains fragmented, often relegated to compliance functions or IT departments, suggesting a disconnect between high-level ethical awareness and actionable leadership practice.

The findings also contribute to emerging debates on AI literacy, reinforcing Senge and Scharmer's (2020) argument that the future of organisational leadership depends on the capacity to engage with complex and cross-disciplinary knowledge systems. The data reveal that although many leaders report a conceptual understanding of AI, few have undertaken targeted development to acquire the specific competencies necessary for responsible oversight. This highlights a significant gap in both academic discourse and organisational preparedness, suggesting that leadership education has yet to fully address the demands of responsible AI adoption.

These findings collectively highlight the pressing necessity for organizations to establish the necessary infrastructure for capability development, formalize AI-related leadership development, and align leadership competencies with the principles of

responsible innovation and digital transformation. This section will elaborate on these themes, offering a detailed analysis of the implications of the quantitative results for future leadership development and strategic talent planning in the AI era.

5.4.2 Contextualizing the Readiness Gap (Qualitative Interpretation)

The empirical support derived from the quantitative results regarding the readiness gap and the disconnect between ethical intent and practical competence, was significantly contextualized by the desk-based qualitative case study (Phase 2). This intermediate phase involved a focused analysis of publicly available organizational documents, policy statements, and high-profile incidents (such as those illustrating governance gaps in data exploitation, unvalidated technology, and safety compromises).

The qualitative analysis reinforced the core conclusion that, despite widespread ethical awareness and strong strategic endorsement of responsible AI principles, organizations frequently struggle to translate aspirational rhetoric into actionable protocols and consistent governance structures. Specifically, the review of corporate policies and practices revealed three critical operational weaknesses that confirmed the quantitative deficits:

1. **Siloed Responsibility:** The case studies indicated a consistent pattern where responsibility for AI governance was often confined to technical, legal, or compliance functions. This reinforced the finding that holistic, cross-functional oversight, which requires board-level engagement, was severely limited.
2. **Lack of Operationalisation:** While organizational documents often affirmed principles such as ‘fairness’ and ‘transparency,’ the qualitative data provided limited evidence of practical ethical competence, meaning, the required tools, mechanisms, or training necessary for leaders to evaluate these outcomes within

algorithmic systems or to mitigate bias effectively. This gap validates the finding that leaders lacked confidence in assessing AI-related risks.

3. **Prioritization of Speed over Validation:** Failures in technology governance, such as the Boeing 737 MAX crisis, demonstrated the severe repercussions when organizational incentives prioritize speed, cost, and market advantage over rigorous validation, transparency, and accountability. This qualitative evidence provides crucial context for interpreting the quantitative data showing that leaders perceive high strategic potential for AI but low governance preparedness, underscoring the necessity for robust governance and regulatory fluency.

The insights gleaned from this phase were essential in framing the discussion, underscoring that the challenge of responsible AI is not primarily one of awareness, but one of structural and systemic failure to integrate ethical and governance frameworks into routine operational practice. This systemic view ultimately justifies the need for the structured intervention provided by the TRAIL framework.

Ultimately, the discussion advances the argument that responsible AI leadership requires more than awareness; it demands structured, interdisciplinary development aligned with the realities of AI deployment. These findings contribute not only to academic discourse on leadership in the digital era but also to the practical design of leadership frameworks for ethical and effective AI adoption.

5.5 AI Literacy

Over 90% of respondents concur that AI literacy has become an essential leadership competency, especially for organizations operating globally. This widespread consensus signifies a shift in the criteria for leadership effectiveness in the digital age. Leaders within international organizations are increasingly required to comprehend AI technologies and their strategic implications, make informed and ethical decisions

regarding AI systems, and navigate complex global regulatory environments. This strong agreement suggests that AI literacy is now viewed as a competitive necessity rather than merely a specialized technical skill.

The findings emphasise the need to embed AI literacy within leadership development frameworks, revising executive programmes to incorporate both foundational knowledge and applied competencies. Leadership skills must be aligned with principles of responsible innovation, governance, and ethical practice to ensure effective oversight of AI. Although the importance of AI literacy is widely acknowledged, the research reveals persistent deficiencies in knowledge, training, and regulatory understanding, signalling a critical area for organisational and educational reform.

This scenario outlines a definitive course of action: while leaders recognize the necessity, it is crucial for organizations to establish the necessary learning ecosystem for capability development. The results provide substantial empirical evidence supporting a central proposition of this study: that AI literacy is rapidly becoming a defining attribute of modern leadership. The development of this competency is imperative; it is crucial for maintaining relevance, managing risks, sustaining competitiveness, and ensuring long-term success in the AI era.

5.6 Risks, Governance & Regulation

Approximately 75% of respondents indicated a need for enhanced training on the limitations and risks associated with AI. This finding underscores a recurring theme in the survey: as responsibilities for AI integration expand, so does the recognition of existing knowledge gaps, indicating a high demand for training and development in this area.

The considerable consensus among respondents indicates that leaders are increasingly aware of AI as not only a technical asset but also a potential source of risk. This is particularly evident in relation to issues such as bias and discrimination, lack of explainability, data privacy and misuse, and overreliance on automated outputs.

While previous findings indicated a strong self-reported ethical awareness, the survey responses highlight a deficiency in technical expertise or applied understanding. Although ethical intent may be present, leaders may not fully comprehend how AI risks manifest in practice or how to effectively mitigate them.

The findings indicate that AI risk training extends beyond a mere issue of competency; it functions as a strategic facilitator of responsible innovation, compliance, and trust, thereby advancing organizational maturity and practical resilience. These findings should inform the development of AI leadership training programs, which ought to prioritize the inclusion of real-world case studies that illustrate AI failures and their associated harms. Furthermore, these programs should provide practical tools for risk identification and mitigation and incorporate cross-functional perspectives, including legal, technical, and operational viewpoints.

The significant demand for additional training highlights the essential need to comprehend the risks and limitations of AI as a core leadership capability in the AI era. These findings offer a compelling justification for incorporating AI risk literacy into leadership development frameworks, particularly as part of broader initiatives to promote responsible AI governance.

A noteworthy majority of over 70% of respondents acknowledged the need for assistance in understanding AI governance and regulation, highlighting an insufficient knowledge to develop, deploy and use AI safely. This finding indicates that many leaders and professionals feel inadequately prepared to navigate or comply with the evolving

frameworks governing AI. Considering the intricate and rapidly advancing nature of AI regulations, such as the EU AI Act, NIST RMF, and OECD guidelines, this demand was not unexpected. Respondents may encounter challenges in comprehending the implications of legislation, implementing governance frameworks within their specific contexts, and aligning organizational policies with legal obligations, particularly if their skill set does not naturally align with this type of leadership competency.

The current complexities of AI governance and regulation, much in a nascent stage, presents a substantial challenge, underscoring both a critical organizational risk and a potential opportunity to enhance awareness and improve competency in this domain. Insufficient comprehension of risk may lead to non-compliance, reputational damage, or ethical violations, which could result in financial penalties and potentially cause irreparable harm to an organization's position within its market.

There exists a distinct opportunity to enhance organizational resilience and advance responsible AI deployment through strategic investment in support, training, and advisory services. Executive briefings on regulatory developments can augment awareness of evolving legal requirements, while the incorporation of AI risk, governance, and regulatory training into leadership programs can cultivate the competencies necessary for ethical oversight, in pursuit of responsible AI. Furthermore, appointing or consulting with AI compliance, legal, and ethics officers can provide specialized expertise, ensuring that organizational practices remain aligned with both regulatory expectations and societal values. Despite the growing awareness of responsible AI, the data indicates that knowledge of governance and regulation remains a significant blind spot. Addressing this issue will be crucial for ensuring that responsible AI principles are not only understood but also legally and strategically integrated into practice. This directly aligns with the research's emphasis on AI literacy and regulatory readiness among leadership.

5.7 Ethical and Responsible AI

The results of this study reveal a significant gap between leaders' awareness of responsible AI principles and their ability to apply these principles in practice. While a majority of participants acknowledged the importance of ethical AI, governance structures, and regulatory compliance, few had received formal training or guidance on operationalising these responsibilities. This disconnect suggests that ethical intent is not being effectively translated into organisational routines or leadership behaviours. For instance, leaders frequently identified 'fairness' and 'transparency' as critical objectives, yet they expressed uncertainty regarding the evaluation of these outcomes within algorithmic systems. This observation aligns with previous research on the challenges of implementing AI ethics (Morley et al., 2021; Mittelstadt et al., 2016). Furthermore, there was evidence of over-reliance on technical teams or compliance officers to manage AI risks, rather than engaging cross-functional leadership or board-level oversight, which aligns with observations by Eitel-Porter (2021) that AI responsibility is often siloed.

Participants also expressed low confidence in their ability to assess AI-related risks, particularly around explainability and accountability, highlighting a widespread need for improved AI literacy among organisational leaders. This reinforces the argument by Long and Magerko (2020) that AI literacy should be considered a core leadership competency, not merely a technical skill set. Taken together, these findings indicate that responsible AI adoption requires more than policy alignment; it demands cultural, structural, and educational interventions at the leadership level.

The findings reveal a significant skills gap in AI ethics, even among leaders who express confidence in the strategic value of AI. Nearly three-quarters of these leaders acknowledge the necessity of acquiring the ability to assess fairness and bias, which are essential elements of responsible AI. Understanding AI's strategic potential does not

inherently equip leaders to evaluate the fairness, transparency, or compliance of AI systems. This underscores a disconnect between strategic vision and ethical implementation.

The respondents are not novices in AI but are strategically engaged individuals who recognise its value and potential. Their expressed interest in further education indicates a readiness to assume leadership responsibilities in this field. However, the findings also reveal gaps in both technical expertise and ethical understanding, which currently limit their capacity to exercise such roles responsibly.

Nearly 90% of the respondents affirmed the ethical imperative of fairness and human dignity in AI decision-making. Such overwhelming agreement suggests that these principles are not only widely endorsed but are likely perceived as foundational to responsible AI adoption within organizational contexts and ethical concerns such as fairness and dignity are no longer peripheral considerations; they are becoming mainstream expectations among professionals, executives, and leaders engaged in AI-related roles. Leaders who prioritize these values will be more closely aligned with prevailing ethical sentiments, thereby enhancing their legitimacy and trustworthiness. Organizations should incorporate these values into AI governance frameworks, codes of conduct, and impact assessments. Given the substantial support, training programs that emphasize these principles are likely to be well-received and may further institutionalize ethical awareness.

A significant majority of respondents indicate an understanding of the ethical implications of AI in business decision-making. This is a positive observation, suggesting that ethical discourse surrounding AI is increasingly prevalent in leadership and professional circles. While confidence levels are high, it is important distinguish this from competence. Leaders may be aware of general ethical concerns, such as bias,

privacy, or transparency, yet may lack the depth required to apply these principles in complex, real-world AI scenarios. Structured education or exposure to applied ethical frameworks, such as fairness audits, impact assessments, or explainability standards, may be beneficial for the 12% of respondents who exhibit a negative regard for the ethical implications of AI.

The implications for governance and strategy suggest that organizations should leverage existing ethical awareness by implementing case-based training in AI ethics, integrating ethics into the governance of the AI lifecycle, and involving ethics specialists in decision-making processes.

The findings offer substantial empirical evidence supporting the integration of ethical principles, specifically fairness and human dignity, into both strategic leadership development and the governance of AI operations. Disregarding ethical considerations in AI-related decisions may result in reputational harm or regulatory examination, especially given the heightened expectations of the public and employees.

Leadership training should encompass more than just the fundamentals of AI or overarching strategic considerations. It ought to incorporate practical tools for evaluating algorithmic fairness and transparency, offer scenario-based learning for the identification and mitigation of bias, and prepare leaders to ask relevant questions to data science and AI teams.

Leadership training should also extend beyond a focus on the fundamentals of AI or broad strategic considerations. It ought to incorporate practical tools for assessing algorithmic fairness and transparency, provide scenario-based learning to identify and mitigate bias, and prepare leaders to engage effectively with data science and AI teams by asking informed and relevant questions. The findings reinforce a central argument of this study: even among individuals who demonstrate a strategic understanding of AI,

there remains a critical need for targeted education in ethics, bias, and fairness. Effective leadership in responsible AI depends not only on visionary insight but also on actionable literacy. Encouragingly, the respondents in this study demonstrate both awareness of these gaps and a willingness to participate in initiatives designed to address them.

The findings suggest a strong foundational awareness of responsible AI leadership; however, this awareness must be operationalised through demonstrable competence, robust oversight mechanisms, and strategic foresight. This supports wider scholarship highlighting leadership literacy as a critical enabler of responsible AI adoption.

5.8 Leadership, Leadership Development and Leadership Frameworks

Burke and Litwin (1992) assert that leadership behaviour exerts a direct influence on organizational culture and the climate within work units, which ultimately affects individual organisational performance, as illustrated in Figure 5.4.

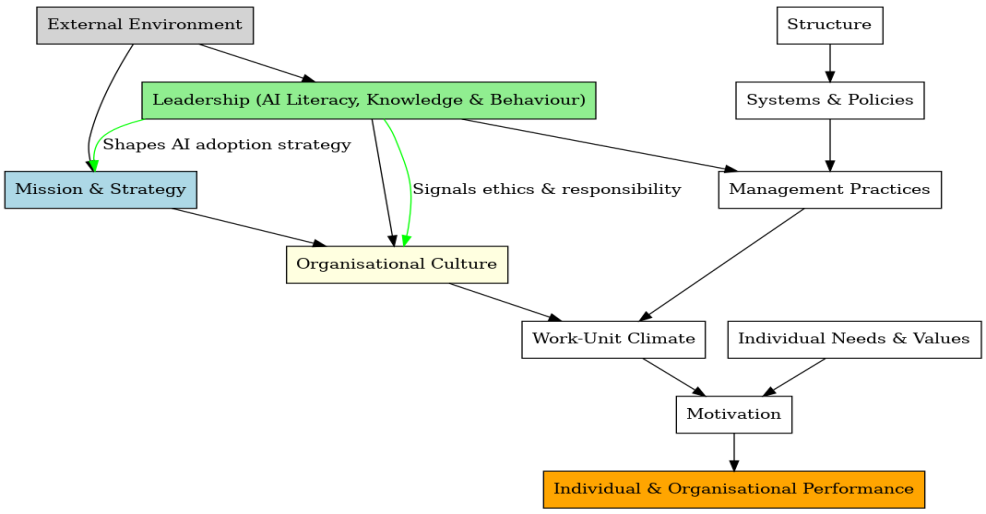


Figure 5.4: Burke-Litwin Model Adapted for AI Leadership Competency

The Burke Litwin model therefore positions leadership as a central determinant of organisational transformation. Within this framework, leadership provides the link

between external pressures and internal alignment, shaping mission, strategy, and organisational culture. As AI becomes embedded in organisational processes, the significance of leadership capability intensifies. AI demands not only technical understanding but also the integration of governance, ethics, and accountability into strategic decision making. Leaders therefore play a decisive role in determining whether AI becomes a source of responsible innovation or a point of organisational vulnerability.

A core component of this capability is AI literacy. More than a technical attribute, AI literacy encompasses the knowledge and judgement required to understand how AI systems function, their potential benefits, and the risks they introduce. Leaders lacking this competency risk creating organisational blind spots that weaken resilience, compliance, and trust. By contrast, leaders who combine AI literacy with strategic insight are able to direct AI adoption in ways that align with organisational values, regulatory expectations, and long-term performance.

Leadership behaviour also shapes organisational culture. Within the Burke Litwin model, behaviour is understood as a mechanism through which leaders signal priorities, reinforce values, and influence the climate of work units. In the context of AI, behaviours that demonstrate ethical awareness, openness to innovation, and attentiveness to societal implications encourage a culture that supports responsible AI. Such cultures not only strengthen employee motivation and commitment but also enhance the organisation's capacity to anticipate and address ethical, technical, and regulatory challenges.

These conceptual foundations are reflected in the empirical findings of this study. More than four fifths of respondents who reported high levels of strategic AI literacy indicated that their roles had already expanded to include responsibilities for AI integration and deployment. This demonstrates that strategic understanding is not merely conceptual; it is increasingly operationalised through direct involvement in AI related

initiatives. Leaders are no longer expected simply to endorse AI adoption but to participate actively in its implementation. This shift signals the emergence of a hybrid leadership model in which strategic foresight is combined with practical engagement in AI driven processes.

However, the findings also reveal a risk. Earlier results showed that many respondents identified significant gaps in their training related to AI ethics, governance, fairness, and risk management. When these gaps are considered alongside the expanding responsibilities described above, it becomes apparent that the evolution of leadership roles may be progressing more rapidly than the development of the competencies required to support them. In the absence of structured development, on the job support, and clear governance frameworks, leaders may find themselves responsible for complex AI decisions without adequate preparation. This creates the possibility of inconsistent oversight and increases the likelihood of ethical or regulatory lapses.

Organisations therefore face an urgent need to formalise leadership development in AI. Training must extend beyond introductory material or broad strategic themes. Leaders require practical tools to evaluate algorithmic fairness and transparency, engage effectively with data science teams, and interrogate AI driven outputs in ways that support accountability and informed decision making. For organisations operating internationally, it is also essential to develop fluency in regulatory requirements across jurisdictions and maintain awareness of cross border ethical considerations.

The findings thus reinforce a central argument of this thesis: strategic AI awareness is closely associated with the evolution of leadership responsibilities, but this evolution must be supported by targeted education and well-defined governance structures. Effective leadership in responsible AI depends on more than visionary intent. It requires actionable literacy, ethical competence, and operational readiness.

Encouragingly, respondents demonstrated awareness of these needs and expressed willingness to engage in training initiatives designed to address them. This suggests a strong foundation upon which organisations can build comprehensive programmes for responsible AI leadership.

The results highlight a clear demand for leadership development in the context of AI. More than 70% of participants acknowledge a personal deficiency in training related to leading organizational change involving AI, suggesting a prevalent lack of confidence or experience in this domain. This observation highlights the pressing necessity to equip leaders with change management skills specifically designed for AI integration. This includes understanding the organizational implications of AI, facilitating digital transformation, managing cross-functional teams, and addressing ethical and governance challenges.

Generic leadership programmes are often insufficient to address the specific organisational demands associated with AI adoption. Organisations should therefore develop tailored leadership development initiatives that emphasise the strategic alignment of AI with business objectives, cultural and behavioural transformation, regulatory and ethical responsibility, and the development of AI literacy and fluency among leaders without technical backgrounds.

The survey findings indicate that executives and business leaders recognise substantive gaps in their readiness to manage AI driven transformation. Across both public and private sector organisations, this represents both a strategic opportunity and an operational imperative to invest in targeted, forward looking leadership development. In the absence of structured capability building, leaders may struggle to guide AI adoption in ways that are responsible, effective, or competitively sustainable.

There is a prevailing consensus that possessing technical or strategic AI capabilities alone is insufficient for leadership that is prepared for the future. This finding indicates a belief that leadership education must evolve to not only incorporate AI fluency but also to prioritize the ethical and responsible application of AI technologies. Absent this focus, educational programs risk producing leaders who are digitally proficient yet ethically unprepared. Leadership-AI education should also encompass a value-driven curriculum that emphasizes fairness, accountability, transparency, bias mitigation, and human dignity. It should integrate regulatory and governance awareness and promote ethical decision-making in AI strategy and deployment. This serves as a strong indication to business schools, executive programs, and internal learning and development teams that ethical AI literacy should be central, not peripheral. Programs should transition from theoretical compliance to practical ethical leadership. Ethical behaviours and responsible AI use must be widely regarded as essential components of modern leadership education, particularly by those who are already critical of existing capability gaps. The results substantiate the urgent call for reform in leadership curricula, ensuring that future leaders are equipped not only to comprehend AI but to employ it with integrity and accountability.

The findings further indicate a prevalent perception among leaders and professionals that leadership education is not keeping pace with the rapid advancements in AI pointing towards a widespread recognition of a capability deficit. The perceived skills gap suggests a potential misalignment between leadership decision-making and the risks associated with AI deployment, which could expose organizations to operational, ethical, or reputational vulnerabilities. A redesign of the curriculum may therefore be necessary to include strategic comprehension of AI, governance frameworks, risk management, and regulatory environments. To ensure continued relevance, leadership

development programs must integrate not only AI literacy, ethical awareness, and technical proficiency, but also a methodology or process to keep leaders' knowledge up to date.

The findings strongly advocate for urgent investment in AI-focused leadership training, cross-disciplinary educational models that integrate business strategy with AI governance and targeted interventions to equip current and future leaders with the competencies necessary for responsible AI adoption.

The overwhelming agreement among respondents underscores a critical need for development and education. As AI systems become increasingly integral to strategic decision-making, organizational success will depend not only on technical implementation but also on leadership capable of navigating the ethical, regulatory, and practical dimensions of AI.

The findings indicate that governance, regulation, ethics, and legal compliance are widely perceived as interconnected dimensions of AI oversight. Governance proficiency must therefore incorporate ethical competence, as ethical and lawful decision making constitutes a core requirement rather than an optional enhancement. The results underscore the need for leadership frameworks that integrate legal obligations, ethical standards, organisational accountability, and practical oversight mechanisms. In the absence of such integration, organisations remain exposed to avoidable risks, with leaders who lack ethical operationalisation potentially delaying implementation, overlooking harms, or responding in inconsistent and reactive ways.

Across organisational practices and leadership attitudes, the evidence points to an urgent need for structured and actionable approaches to responsible AI. Although there is clear awareness of ethical considerations and support for innovation, the lack of coherent leadership pathways and integrated governance mechanisms represents a significant

vulnerability. Compliance, while necessary, is not by itself sufficient. Effective leadership must cultivate cross functional literacy, accountability and a culture of trust and human centred design, thereby ensuring that AI is deployed in ways that are both responsible and aligned with organisational purpose.

Most leaders who identified gaps in their understanding of regulatory frameworks also reported a need for greater guidance in the ethical and legal oversight of AI systems. This reinforces a central conclusion of this study: responsible AI leadership depends on comprehensive development that extends beyond technical awareness to include ethical reasoning and regulatory literacy. Embedding these dimensions within leadership training, organisational policies and governance structures will be essential in advancing AI maturity and reducing risk.

5.9 Leadership Education in AI: Emergent Key themes

Several key themes have emerged from the research regarding the evolution of leadership education in the context of AI. These themes collectively underscore the urgent need for a more holistic approach to preparing leaders for the AI era.

5.9.1 The Rapid Evolution of Leadership Roles and the Corresponding Readiness Gap

Over 80% of respondents who demonstrate strategic literacy in AI report an expansion of their roles to encompass responsibilities related to AI integration and deployment. This indicates that strategic understanding actively drives direct involvement in AI transformation, leading to an emerging hybrid leadership model combining strategic vision with applied technology management. This trend is particularly evident in international organisations, where over 75% of respondents note their roles have evolved to include AI integration and deployment. However, a significant concern highlighted is that this role expansion may be outpacing capability development, especially concerning

AI risks, fairness, ethics, and governance. Leaders may be assuming complex AI-related duties without adequate preparation, potentially leading to inconsistent oversight or ethical missteps.

5.9.2 The Interconnectedness of AI Governance, Regulation, Ethics, and Legal Compliance

For many professionals, governance, regulation, ethics, and legal compliance are perceived as interconnected and interdependent components of AI oversight. Proficiency in governance must encompass ethical competency, as support for legal and ethical decision-making is essential. The findings advocate for the development of integrated leadership frameworks that do not treat ethics, compliance, and governance as distinct disciplines. Instead, training and policy should unify legal obligations, ethical standards, organisational accountability, and practical tools for AI oversight.

A significant majority (over 70%) of respondents require support in understanding AI governance and regulatory frameworks, indicating a prevalent knowledge deficiency. This demand is understandable given the intricate and rapidly advancing nature of AI regulations, such as the EU AI Act, NIST RMF, and OECD guidelines. Insufficient comprehension of AI risks and regulations can lead to non-compliance, reputational damage, and ethical violations. Addressing this 'Governance Gap' is crucial for ensuring responsible AI principles are legally and strategically integrated into practice.

5.9.3 Critical Skills Gaps and the Need for Practical AI Literacy

Even among leaders confident in AI's strategic value, nearly three-quarters recognise a need to learn how to assess AI systems for fairness and bias, indicating a notable skills gap in AI ethics. Comprehending strategic potential does not inherently equip leaders to evaluate the fairness, transparency, or compliance of AI systems.

Leadership training, therefore, needs to extend beyond fundamentals and include practical tools for assessing algorithmic fairness and transparency, scenario-based learning for mitigating bias, and equipping leaders to ask pertinent questions to data science and AI teams. Furthermore, approximately 75% of respondents indicate a need for enhanced training on the limitations and risks associated with AI, including bias, lack of explainability, data privacy, and overreliance on automated outputs. While ethical intent may be present, leaders often lack the practical understanding of how AI risks manifest and how to effectively mitigate them.

Leadership education should emphasise real-world case studies of AI failures and provide practical tools for risk identification and mitigation, incorporating cross-functional perspectives. More than 70% of participants also acknowledge a personal deficiency in training related to leading organisational change involving AI adoption, suggesting a widespread lack of confidence or experience in this domain. Generic leadership programmes are unlikely to be adequate; customised training modules focusing on strategic alignment, cultural transformation, and AI literacy for non-technical leaders are required.

5.9.4 The Urgent Imperative for Holistic AI Leadership Education, Beyond Technical Proficiency

There is a prevailing consensus that possessing technical or strategic AI capabilities alone is insufficient for leadership that is prepared for the future. Leadership education must evolve to not only incorporate AI fluency but also to prioritize the ethical and responsible application of AI technologies. Failing to do so risks producing leaders who are digitally proficient yet ethically unprepared. The results strongly advocate for a value-driven curriculum that emphasizes fairness, transparency, bias mitigation, and

human dignity, positioning ethical AI literacy as central, not peripheral. Programs must transition from theoretical compliance to practical ethical leadership.

There is a widespread perception that current leadership education is not keeping pace with AI advancements, highlighting a significant capability deficit. Therefore, there is an urgent call for reform in leadership curricula, ensuring future leaders can not only comprehend AI but also employ it with integrity and accountability.

5.9.5 Formalising AI Leadership Development and Strategic Talent Planning

AI literacy is increasingly recognised as a core leadership competency, with over 90% of respondents, particularly within international organisations, expressing agreement. This consensus reflects a shift in the criteria for leadership effectiveness, positioning AI literacy as a strategic and competitive necessity rather than a specialised technical skill. The findings indicate that organisations must formalise AI focused leadership development and provide structured, practice-based support. This includes integrating ethical and regulatory competence alongside technical and strategic training, and embedding leadership and talent development within formal strategic planning processes. Such approaches involve aligning existing AI related responsibilities with targeted development pathways and prioritising regulatory fluency and cross border ethical awareness. Overall, the data provides strong empirical support for embedding ethical principles, notably fairness and respect for human dignity, within both leadership development and operational AI governance.

Figure 5.5 synthesises the principal themes arising from the research, illustrating the interdependence of AI governance, regulatory frameworks, ethical considerations, and legal compliance, and demonstrating that these domains cannot be addressed in isolation. It also highlights the rapid evolution of leadership requirements and the resulting readiness gap among many current leaders, compounded by persistent skills

shortages and limited practical AI literacy. Collectively, these findings underscore the need for comprehensive AI leadership education that extends beyond technical proficiency and supports the formalisation of AI focused leadership development and strategic talent planning as vital organisational responses.

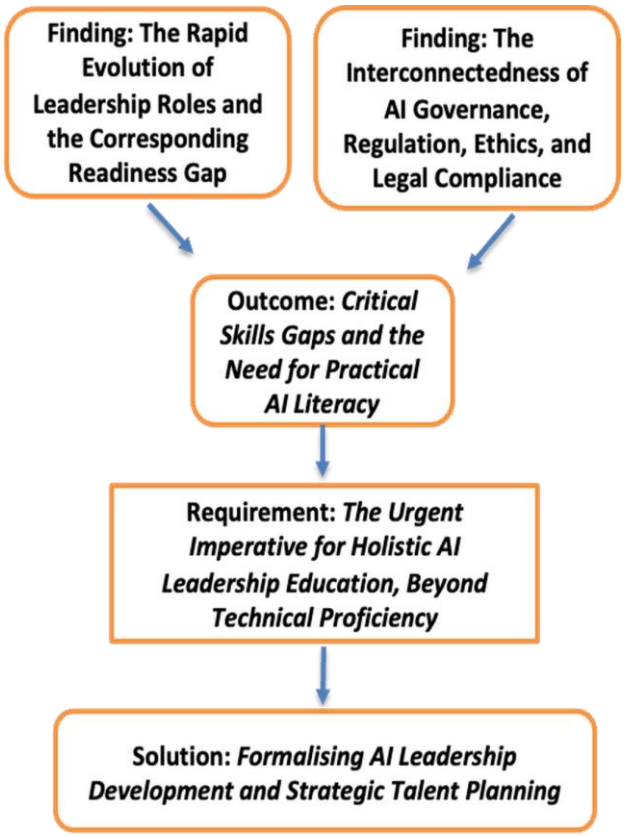


Figure 5.5: Interconnectedness of the Key Themes from the Findings

5.10 The Gap: Leadership Responsibilities and AI Capabilities

The survey responses illuminate a striking consensus: AI is not only redefining operational processes but is also reshaping the competencies and expectations of contemporary leadership. Overwhelming support for AI literacy as a core leadership skill suggests that the integration of AI into business strategy is no longer peripheral, it is fundamental. Yet this enthusiasm for strategic engagement is tempered by self-reported

deficiencies in ethics, risk awareness, governance comprehension, and change leadership capabilities.

The analysis uncovered several pivotal insights into the dynamic landscape of AI leadership. A notable consensus has emerged regarding AI literacy as an essential leadership competency, with over 90% of respondents, particularly those within international organizations, recognizing its competitive necessity.

Simultaneously, the findings underscore a substantial shift in leadership responsibilities, with numerous roles now incorporating tasks related to the integration and deployment of AI. Nonetheless, a significant concern arises that this swift evolution of roles may surpass the development of the necessary capabilities, thereby indicating a readiness gap.

Furthermore, the findings highlight specific deficiencies in leaders' preparedness across several critical domains. There is a pronounced need for enhanced training in understanding the limitations and risks associated with AI systems, alongside a recognized necessity for support in navigating complex AI governance and regulatory frameworks, such as the EU AI Act (European Union, 2024) and NIST RMF (Dotan, 2024). Despite self-reported ethical awareness, a significant skills gap persists in the practical assessment of AI systems for fairness and bias, even among leaders confident in AI's strategic value. This disconnect underscores the urgency for a more comprehensive approach to leadership education that transcends mere technical proficiency, emphasizing ethical and responsible AI application, and integrating legal and ethical considerations into governance. The widespread acknowledgment of a gap between current leadership education and the AI capabilities required in practice reinforces the imperative for targeted, forward-looking development initiatives.

This study's findings reinforce and extend existing literature on the critical role of leadership in ensuring the responsible adoption of AI. Scholars such as Dignum (2019) and Eitel-Porter (2021) argue that ethical AI cannot be achieved through technical safeguards alone; instead, it requires embedded organisational values, effective governance, and informed leadership. Participants in this study consistently pointed to a lack of clarity and confidence in operationalising ethical principles, echoing the gap between high-level ethical aspirations and practical implementation noted in the work of Mittelstadt et al. (2016). Furthermore, while frameworks such as the EU AI Act and the OECD Principles on AI (OECD, 2021) provide guidance at a policy level, there remains limited empirical insight into how these principles are interpreted and enacted by business leaders.

The findings suggest that AI literacy, defined here as a leader's capacity to understand, interrogate, and guide AI-related decisions, is unevenly distributed across senior roles, a pattern consistent with observations by Scharmer and Kaufer (2013) on systemic blind spots in leadership. This study therefore contributes to the literature by highlighting the interdependency between ethical governance structures and leadership capability, while also exposing the limitations of treating AI readiness as an IT or compliance function, rather than as a strategic leadership concern.

The findings also support and extend the existing literature on responsible AI by emphasising the pivotal role of leadership in guiding ethical and governance practices. Scholars such as Dignum (2019) and Morley et al. (2021) have argued that effective AI governance requires not only robust technical safeguards, but also strong organisational leadership grounded in ethical principles. This study corroborates that claim, revealing that while many leaders express ethical concern regarding AI systems, they lack the structured knowledge or frameworks to act with confidence. The concept of AI literacy,

emerging in the work of Long and Magerko (2020), becomes especially relevant here, as participants highlighted gaps in understanding key AI concepts, such as bias, explainability, and algorithmic accountability, at senior decision-making levels. Furthermore, although policy instruments like the EU AI Act (European Commission, 2021) aim to enforce responsible AI practices, this research suggests that compliance-driven approaches alone are insufficient without concurrent leadership development and cross-functional collaboration. This aligns with findings by Fjeld et al. (2020), who argue that ethical principles must be operationalised within organisations through culture, training, and leadership support. This study contributes to this discourse by offering empirical insight into the disconnect between high-level awareness and everyday decision-making in AI implementation, underscoring the need for leadership-centred frameworks for responsible AI.

In response, the next chapter presents a practical framework to guide leaders through the complexities of AI adoption. Grounded in regulation, ethics, governance, and organisational learning, it offers a roadmap for building the competencies, oversight structures, and cultural foundations required for responsible and sustainable AI implementation.

5.11 The TRAIL Framework

5.11.1 Introduction – Framework Chapter

Building on the findings discussed in the previous chapters, this section introduces a practical framework for leadership in the era of responsible AI. The framework is designed to address the critical gaps identified in leadership competency, AI literacy, and organisational governance, particularly in relation to the ethical design, development, deployment, and usage of AI systems. It draws directly from the empirical insights of this study, which revealed a prevalent recognition of AI's potential and risks,

but an equally widespread lack of structured guidance, cross-functional engagement, and strategic oversight. By integrating concepts from AI ethics, regulatory readiness, and leadership development literature, the framework aims to provide leaders with an actionable model for embedding responsible AI into their organisations.

The primary objective of the framework is to offer practical utility and actionable guidance for senior decision-makers. It is designed not as a rigid prescription but as a flexible tool that can be adapted across various industries and levels of maturity.

This chapter first outlines the guiding principles underpinning the framework, then examines its core components, and concludes with recommendations for implementation and evaluation. The chapter culminates with a force field diagram (Figure 5.8) illustrating the dynamic interplay of forces that influence an organisation's capacity to develop and sustain responsible AI capability and stewardship.

5.11.2 Background and Purpose

The TRAIL (Trustworthy and Responsible AI Leadership) framework is a structured model for developing responsible AI leadership. It provides a holistic approach to leadership development that enables senior professionals in the commercial sector to design, deploy, and oversee AI systems in ways that are ethically grounded, compliant with evolving regulations, and supported by effective governance structures.

The need for such a framework is evident from research findings that highlight the growing importance of AI literacy as a defining feature of contemporary leadership. Despite this, a substantial readiness gap persists, as many leaders are assuming responsibility for AI adoption more quickly than their level of preparation permits. TRAIL addresses this challenge by embedding best practice principles into leadership education, ensuring that transparency, accountability, and fairness are consistently

integrated into organisational approaches to AI. Its ultimate purpose is to strengthen public confidence, reduce risk, and support ethical and sustainable implementation.

5.11.3 Guiding Principles of the TRAIL Framework

The TRAIL framework is anchored by a set of guiding principles derived directly from the empirical findings of the study and informed by established global standards for responsible AI (e.g., the OECD AI Principles and the NIST Risk Management Framework). These principles are intended to operationalise the core value-based orientation required for responsible AI leadership and guide the implementation of the subsequent competency pillars.

The foundational principles that guide the structure and application of TRAIL are:

1. **Ethical Primacy and Human-Centred Focus** The framework advocates for value-driven content in leadership development, emphasising ethical, human-centred, and socially responsible approaches to AI. Its ultimate purpose is to strengthen public confidence and support ethical and sustainable implementation. Ethical capability must extend beyond conceptual awareness to encompass the practical skills required for oversight, ensuring that AI systems support human-centred outcomes and human dignity.
2. **Accountability and Transparency** The framework is designed to embed best practice principles that consistently integrate transparency and accountability into organisational approaches to AI. Leaders are supported in developing ethical and human-centred AI practices that reinforce these core values. For implementation to be effective, organisations must embed clear mechanisms of accountability that ensure responsible practices are consistently enacted through transparent processes. Global convergence around principles includes upholding accountability and transparency as shared norms.

3. **Fairness and Non-Discrimination** Fairness is recognised as a critical principle that leaders must uphold and embed into decision-making processes. The framework focuses on equipping leaders with the ability to identify, evaluate, and mitigate risks such as algorithmic bias and discrimination. This is achieved by implementing best-practice models that emphasise fairness throughout the data lifecycle.
4. **Integrated Governance and Oversight** Effective oversight requires an integrated approach that recognises the interdependence of governance, regulation, ethics, and legal compliance, which operate as mutually reinforcing domains. The framework advocates for integrated governance frameworks that unify legal obligations, ethical standards, and accountability mechanisms. Furthermore, responsibility for AI governance must not be siloed, but must be addressed collectively at senior and board level to ensure holistic oversight.
5. **Actionability and Contextual Adaptability** TRAIL is designed as an actionable model intended not as a rigid prescription, but as a flexible tool that can be adapted across industries and maturity levels. The evidence indicates increasing demand for leadership development programmes that are contextualised to organisational needs, sectoral risks, and maturity levels, rather than being generic. The framework structure is adaptable to situational variables, acknowledging differences in markets, jurisdictions, and regions.
6. **Continuous Capability Building** The framework assumes a need for continuous learning and adaptation among leaders. It promotes the institutionalisation of AI leadership development through continuous education, ongoing capability building, and the utilization of longitudinal reviews to monitor progress and

maturity. This principle ensures that leadership remains adaptive in the face of rapid technological advancement.

5.11.4 Scope and Target Audience

The TRAIL framework is designed for senior decision makers within commercial organisations, including those with international responsibilities. These leaders increasingly face direct accountability for AI development, integration, and deployment, making dedicated leadership training essential. The educational scope of the framework recognises the urgency of holistic AI leadership development that extends beyond technical competence. It advocates the formalisation of structured leadership pathways and strategic talent planning that address the ethical, regulatory, and cultural dimensions of AI. Specifically, the framework supports:

- Strategic leadership in AI adoption and oversight
- Compliance with national and international regulations and guidelines (including the EU AI Act, the NIST Risk Management Framework, and the OECD AI Principles)
- The development and application of internal governance mechanisms
- Ethical and human-centred AI practices that reinforce trust, fairness, and accountability
- Organisational capacity building, cultural development, and talent planning to support AI maturity

The analysis highlights several domains that require sustained leadership attention to support responsible AI adoption within organisations.

1. **AI Literacy and Strategic Fluency:** Leaders require both foundational and applied knowledge of AI technologies, together with an appreciation of their strategic, operational, and socio-technical implications. Such literacy enables

informed decision making and supports the integration of AI into broader organisational priorities.

2. **Risk and Limitation Literacy:** A critical competency involves understanding the inherent risks and limitations of AI, including algorithmic bias, opacity, privacy and security vulnerabilities, misuse, and the organisational hazards associated with overreliance on automated systems. This literacy forms the basis for risk-aware leadership and responsible deployment.
3. **Governance and Regulatory Readiness:** Leaders must be able to interpret and operationalise emerging regulatory frameworks, ensuring that organisational practices remain compliant while also aligning internal policies, processes, and controls with legal and ethical obligations.
4. **Practical Ethical Competence:** Ethical capability must extend beyond conceptual awareness to encompass the practical skills required for oversight. This includes the ability to interrogate AI systems for fairness, transparency, accountability, and human-centred outcomes, and to embed these considerations within governance mechanisms.
5. **Change Management:** Effective leadership for AI adoption demands the capacity to integrate AI within wider programmes of digital transformation. This includes orchestrating cross-functional collaboration, supporting workforce adaptation, and managing the organisational change dynamics that accompany technological disruption.

5.11.5 Assumptions of the Framework

The TRAIL framework is founded on several assumptions derived from empirical findings and the broader literature.

1. **AI Literacy as a Leadership Imperative:** The research indicates that AI competence has become a core dimension of contemporary leadership, influencing organisational competitiveness, strategic capability, and the capacity to exercise responsible oversight.
2. **The need for Value-Driven Content:** Effective leadership development must extend beyond technical instruction to emphasise ethical, human-centred, and socially responsible approaches to AI. Technical proficiency, while necessary, is insufficient without an accompanying value-based orientation.
3. **Leadership as a Driver of Transformation:** Consistent with organisational change models (Burke and Litwin, 1992), leaders shape strategy, mission, culture, and behaviour. Their influence is therefore central to the responsible adoption, governance, and integration of AI systems.
4. **A Willingness to Learn:** The findings suggest that senior executives recognise gaps in their AI knowledge and ethical readiness and are prepared to engage in targeted development to strengthen these competencies.
5. **The Interconnectedness of Oversight Domains:** Governance, regulation, ethics, and legal compliance operate as mutually reinforcing domains rather than discrete areas of responsibility. Effective oversight requires an integrated approach that recognises their interdependence.
6. **The Inevitability of Evolving Regulation:** Although regulatory requirements will continue to develop, there is growing global convergence around principles such as fairness, accountability, transparency, and human protection. Leaders must therefore remain adaptive while upholding these shared norms.
7. **The Need for Tailored Education:** The evidence indicates increasing demand for leadership development programmes that are contextualised to organisational

needs, sectoral risks, and maturity levels, rather than generic or universally applied training.

5.11.6 Constraints

Implementation of the TRAIL framework must be situated within a set of significant organisational and environmental constraints that shape leaders' capacity to enact responsible AI practices. These constraints, identified through the empirical study, introduce systemic complexity, reduce capability, and limit the organizational capacity for comprehensive oversight. Successfully navigating these structural challenges requires an adaptive and comprehensive leadership model that explicitly addresses the sources of friction identified below:

- Regulatory complexity and fluidity: Global regulatory environments remain highly complex and continue to evolve at pace, requiring leaders to navigate shifting compliance expectations and emerging governance obligations.
- Fragmented regulatory standards: Cross-jurisdictional inconsistencies create uncertainty for multinational organisations and complicate the development of unified governance practices.
- Rapid technological advancement: The speed of AI innovation necessitates adaptive leadership frameworks capable of responding to emerging tools, risks, and socio-technical disruptions.
- Gaps between ethical awareness and practical competence: Although ethical awareness is increasingly common, many leaders lack the practical skills necessary to evaluate systems for fairness, transparency, accountability, and human impact.

- Siloed responsibility for AI oversight: In many organisations, AI governance is confined to technical, legal, or compliance functions, limiting board-level engagement and constraining holistic oversight.
- Limitations of generic leadership programmes: Traditional or generalised leadership development initiatives often fail to address the specific knowledge, competencies, and judgement required for responsible AI adoption.
- Subjectivity in ethical interpretation: Ethical reasoning is influenced by demographic, cultural, and experiential factors. This variation can produce divergent judgments about acceptable risk, fairness, and responsibility.
- Resource constraints: Many organisations face limitations related to specialist expertise, training budgets, and the capacity to embed responsible AI practices at scale.

These constraints collectively contribute to the substantial readiness gap that persists across the commercial sector, necessitating a framework like TRAIL which advocates for formalised leadership pathways and strategic talent planning to address the ethical, regulatory, and cultural dimensions of AI.

5.11.7 Risks Associated with Non-Implementation

The absence of a structured and holistic framework such as TRAIL exposes organizations to a variety of significant and multifaceted risks, which are the direct consequence of the organizational and leadership gaps identified in the research. These risks extend beyond financial penalties to encompass long-term erosion of stakeholder trust, compromised organizational integrity, and misalignment between decision-making and ethical imperatives. Addressing these potential negative consequences underscores the urgent need for structured guidance and strategic oversight provided by the TRAIL framework:

- **Inadequate leadership preparedness:** In the absence of targeted development, leaders may assume responsibility for AI systems without possessing the necessary knowledge or ethical foundation, leading to inconsistent oversight, poor judgment, or ethical lapses.
- **Regulatory and risk mismanagement:** A limited understanding of regulatory obligations or AI-related risks heightens the probability of non-compliance, financial penalties, reputational damage, and erosion of stakeholder trust.
- **Imbalanced leadership capability:** An overemphasis on technical or strategic competence, without accompanying ethical literacy, may result in leaders who are digitally proficient but ill-equipped to assess societal, organizational, and human impacts.
- **Misalignment between decision-making and ethical risk:** An inadequate appreciation of the ethical, legal, or reputational consequences of AI deployment can lead to decisions that compromise organizational integrity and long-term performance.

The structured approach offered by the TRAIL framework is specifically designed to mitigate these risks by integrating ethical principles, promoting continuous capability building, and establishing robust governance structures. The implementation of TRAIL, as detailed in the subsequent sections, provides the structured practices required, including candid self-assessment of leadership readiness, regular evaluation of outcomes, and the use of independent advisory expertise to support responsible adoption.

5.11.8 Implementation Process

The implementation of TRAIL is organized into four distinct phases:

Phase 1 involves conducting a thorough diagnostic assessment of organizational AI preparedness and literacy to determine the scope and resource requirements necessary for implementation.

Phase 2 focuses on formalizing leadership development pathways, redesigning curricula to incorporate governance and risk management, establishing governance structures, and introducing AI literacy programs supported by baseline capability assessments.

Phase 3 aims to integrate ethics and governance into leadership curricula and promote cross-functional approaches to AI oversight.

Phase 4 seeks to institutionalize AI leadership development through continuous education, alignment with responsible AI values, and the utilization of longitudinal reviews to monitor progress and maturity.

5.11.9 Framework Structure: Key Competency Pillars

The analysis of the empirical findings confirms a crucial challenge: while leaders recognize AI literacy as a competitive necessity and exhibit strong ethical awareness, this strategic insight is consistently undermined by structural and competency deficits. The research revealed that the swift evolution of leadership roles is outpacing the development of necessary capabilities, resulting in a substantial readiness gap and vulnerability to ethical or regulatory lapses.

Specifically, the synthesis highlighted five critical deficiencies that must be addressed through a structured intervention:

1. The expansion of leadership roles without corresponding formal training creates inconsistent oversight.
2. The governance gap persists due to insufficient knowledge of complex regulatory frameworks (e.g., EU AI Act, NIST RMF).

3. Despite ethical intent, critical skills gaps limit the practical ability to assess fairness and mitigate bias in AI systems.
4. Responsibility for AI is often siloed, hindering the holistic, integrated governance required for effective compliance.
5. Compliance-driven approaches alone are insufficient without concurrent leadership development and cross-functional collaboration.

Therefore, to operationalise ethical principles and close these persistent gaps, the current chapter transitions from analysis to solution. The TRAIL (Trustworthy and Responsible AI Leadership) framework is introduced as the empirically derived, comprehensive, and actionable model specifically designed to translate strategic awareness into measurable competence, embedding ethical accountability and governance structures across the commercial organisation. This framework is grounded in regulation, ethics, governance, and organisational learning, offering the roadmap required for building the necessary competencies and cultural foundations for responsible and sustainable AI implementation.

The core structure of the TRAIL framework is defined by four interdependent Key Competency Pillars (Figure 5.6), which represent the strategic, governance, ethical, and data-related domains critical for responsible AI leadership. These pillars are derived directly from the empirical findings of this study, explicitly addressing the critical gaps identified in leadership competency, AI literacy, and organisational governance. As shown in Figure 5.6, the four pillars correlate directly with the urgent leadership imperatives revealed by the research, translating generalized readiness deficits into specific developmental domains.

As an actionable model intended to provide specific guidance rather than a rigid prescription, the framework's detailed content is presented primarily via a comprehensive

table (Table 5.1). This structured format is utilised to ensure clarity and immediate comprehension of the components necessary for implementation.

The table disaggregates the requirements of each Pillar across three essential and integrated dimensions:

1. **Competency Focus:** Defining the specific knowledge, skills, and judgement required of senior leaders within that domain.
2. **Best Practice Guidance:** Providing systemic, organisational, and policy-level recommendations necessary to embed responsible practices (e.g., integrated governance frameworks, ethical oversight mechanisms).
3. **Development Techniques:** Outlining practical methods and intervention strategies for professional growth and capability building (e.g., scenario-based simulations, cross-functional collaboration training).

By presenting the Pillars in this tripartite structure, the framework ensures that ethical intent and regulatory fluency are translated into practical competence and measurable organizational practices, which is central to achieving sustainable responsible AI adoption. The following table details the comprehensive requirements of these four pillars.

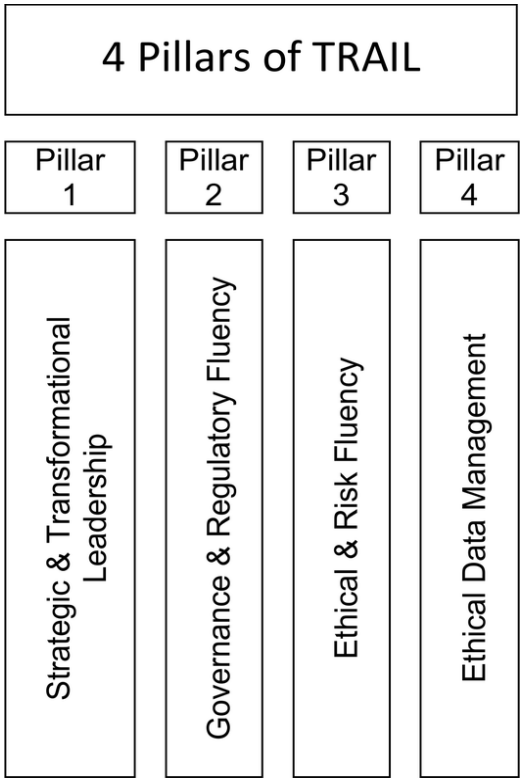


Figure 5.6: The Four Pillars of TRAIL

The model is structured as a framework that, while consistently emphasizing competency, best practices, and development techniques, is adaptable to situational variables. It acknowledges that no two organizations are identical, nor are the markets, jurisdictions, or regions in which they operate.

Table 5.1: The Four Pillars Related to the Readiness Gap

Pillar	Competency Focus:	Best Practice Guidance:	Development Techniques:
Pillar I: Strategic and Transformational Leadership	This pillar centres on the leadership behaviours and change management skills required to guide organisations through the transformational adoption of AI. Leaders must align AI strategy with organisational objectives, anticipate structural and cultural implications, and drive digital transformation in ways that foster long-term value creation.	Effective strategic leadership requires integrating ethical principles directly into the governance of the AI lifecycle. Leaders should model responsible behaviour, visibly embedding accountability and fairness into decision-making processes. By aligning AI initiatives with organisational values, leaders can cultivate a culture that embraces responsible innovation and sustains trust.	Training should prioritise cross-functional collaboration and provide leaders with the tools to manage diverse teams involved in AI adoption. Programmes should incorporate practical modules on digital transformation, case studies of successful AI integration, and exercises in aligning technological initiatives with broader strategic goals.
Pillar II: Governance and Regulatory Fluency	This pillar focuses on developing leaders' ability to navigate the complex legal, compliance, and governance challenges associated with AI, such as the EU AI Act, the NIST Risk Management Framework, and the OECD AI Principles. Leaders must acquire fluency in regulatory frameworks across jurisdictions and apply governance tools to ensure AI systems are both legally compliant and ethically aligned.	Organisations should adopt integrated governance frameworks that unify legal obligations, ethical standards, and accountability mechanisms. Leaders are encouraged to establish context-specific AI policies and consult with compliance or ethics officers to ensure oversight. Embedding governance into strategic and operational routines helps mitigate regulatory and reputational risks.	Programmes should provide executive briefings on emerging regulations, complemented by tailored modules on governance practices. Training should combine legal, ethical, and operational perspectives to ensure leaders can apply frameworks in practice, supported by scenario-based exercises that prepare them for regulatory challenges.
Pillar III: Ethical and	This pillar emphasises the need to translate ethical intent into practical competence by equipping leaders with the ability to identify, evaluate, and mitigate risks such as bias, discrimination,	Leadership education must adopt a values-driven curriculum that embeds fairness, transparency, bias mitigation, and respect for human dignity.	Learning interventions should include tools for assessing algorithmic fairness and explainability, along with case-based training that examines real-world

Risk Literacy	lack of explainability, and overreliance on automation.	Training should reflect cross-functional input, integrating legal, technical, and organisational perspectives to ensure a balanced approach to ethical risk management.	ethical failures. Scenario-based simulations should help leaders recognise and respond to ethical dilemmas, ensuring ethical considerations move beyond theoretical compliance to practical application.
Pillar IV: Ethical Data Management	This pillar focuses on the ethical acquisition, use, storage, and governance of data as a foundational requirement for responsible AI. Leaders must develop fluency in addressing risks such as data misuse, privacy breaches, and weak stewardship, while also recognising the strategic advantage of ethical and transparent data practices.	Ethical data management requires embedding ethical principles into the everyday handling of data. Leaders should implement best-practice models that emphasise transparency, accountability, and fairness throughout the data lifecycle. This involves scrutinising provenance and consent, codifying policies, conducting impact assessments, and establishing governance structures that maintain organisational resilience.	Training should provide leaders with practical tools such as fairness audits, scenario-based exercises around privacy and misuse, and case studies on data governance challenges. Education should also introduce stewardship models, encourage cross-functional oversight, and build confidence in applying ethical principles consistently across the organisation.

As shown in Table 5.1, each of the four Pillars is designed to close a specific dimension of the leadership and organisational readiness gap identified in the research:

- Pillar I (Strategic and Transformational Leadership) addresses the Rapid Evolution of Leadership Roles and the Corresponding Readiness Gap, focusing on Change Management and Strategic leadership in AI adoption and oversight.
- Pillar II (Governance and Regulatory Fluency) is structured to address the Interconnectedness of AI Governance, Regulation, Ethics, and Legal Compliance, thereby building necessary Governance and Regulatory Readiness to ensure compliance with national and international guidelines.

- Pillar III (Ethical and Risk Fluency) directly closes the Critical Skills Gaps and the Need for Practical AI Literacy. It equips leaders with the Practical Ethical Competence and Risk and Limitation Literacy required to move beyond conceptual ethical awareness.
- Pillar IV (Ethical Data Management) addresses the Urgent Imperative for Holistic AI Education, Beyond Technical Proficiency, ensuring foundational leadership capability in the ethical acquisition, use, storage, and governance of data.

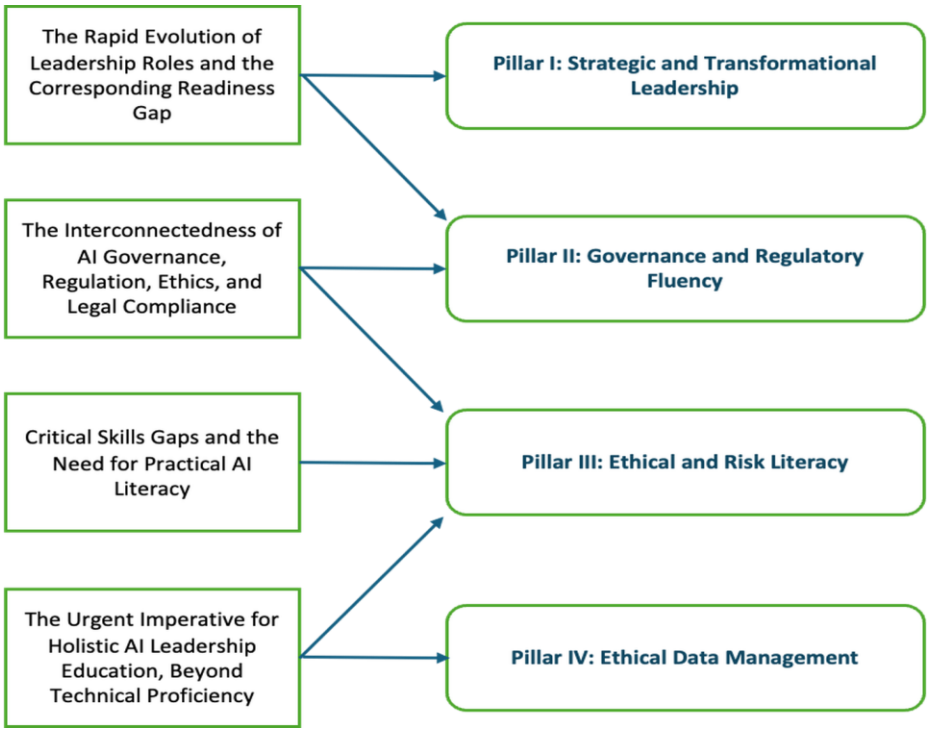


Figure 5.7: Relationship Between Key Findings and TRAIL Key Pillars

5.11.10 Supporting Mechanisms of the TRAIL Framework

The TRAIL framework is designed as a holistic and actionable model that integrates the four core competency Pillars with essential organizational mechanisms required for successful implementation and sustained effectiveness. While the Pillars define *what* leaders must know and do, these Supporting Mechanisms define the systemic structures and

processes necessary to embed responsible AI practices organization-wide, ensuring that ethical intent translates into consistent operational outcomes.

These supporting components are vital for moving beyond conceptual development to achieve measurable, accountable, and culturally resilient AI leadership. The integration of these mechanisms helps to address critical constraints identified in the research, such as the gap between ethical awareness and practical competence, and the tendency toward siloed responsibility for AI oversight.

Specifically, the Supporting Mechanisms cover four interdependent domains:

1. **Measurement and Evaluation:** This mechanism ensures the framework's effectiveness through systematic evaluation that assesses practical competence and organisational outcomes, utilizing both quantitative and qualitative indicators. Continuous monitoring ensures the framework remains relevant and effective.
2. **Operational Accountability and Reporting:** This requires embedding clear mechanisms that ensure responsible practices are consistently enacted, fostering transparent processes and requiring collective responsibility for AI governance at senior levels.
3. **Cultural and Behavioural Transformation:** Recognizing that culture shapes decision-making, this mechanism focuses on actively modelling ethical responsibility and embedding fairness, transparency, and accountability into organizational strategy and daily routines.
4. **Global and Cultural Nuances:** This provides the necessary sensitivity and regulatory fluency for leaders navigating diverse cultural contexts and jurisdictional differences, balancing local relevance with globally coherent ethical standards.

Taken together, these mechanisms equip leaders with the necessary tools, behaviours, and insights to guide their organizations in deploying AI in ways that are ethical, resilient, and sustainable. The following table details the Competency Focus, Best Practice Guidance, and Development Techniques associated with these critical Supporting Mechanisms.

Table 5.2: Supporting Mechanisms

Element	Description	Competency Focus	Best Practice Guidance	Development Techniques
1. Measurement and Evaluation	The effectiveness of the TRAIL framework depends on systematic evaluation that moves beyond self-reported awareness to the assessment of practical competence. Measurement should be multi-dimensional, combining qualitative and quantitative indicators that capture both leadership behaviours and organisational outcomes.	Evaluation must determine whether leaders are able to apply ethical principles, governance knowledge, and risk management skills in practice. The focus is not only on confidence or awareness but on demonstrated capability to question, challenge, and guide AI development and deployment effectively.	Organisations should establish structured mechanisms to track the impact of leadership education on decision-making and culture. This includes the use of operational metrics such as fairness audits, algorithmic impact assessments, and explainability standards. Continuous monitoring of compliance incidents, resilience indicators, and organisational maturity in responsible AI provides further insight into effectiveness.	Evaluation should incorporate longitudinal studies to track changes in ethical attitudes and behaviours over time. Organisations should identify trends in leadership practice and culture as AI adoption matures. Training programmes should be reviewed through systematic feedback and outcome analysis. Assessments should combine individual competency reviews with organisational performance indicators, creating a balanced picture of progress.
	For the TRAIL framework to be effective, organisations must	Leaders must demonstrate the ability to establish and oversee	Organisations should formalise reporting systems that allow	Training should equip leaders to design, implement, and

2. Operational Accountability and Reporting Mechanisms	embed clear mechanisms of accountability that ensure responsible practices are not only promoted but consistently enacted. This requires leadership commitment to transparent processes, ethical oversight, and systems that enable continuous monitoring of behaviour and outcomes.	mechanisms that support ethical decision-making across the AI lifecycle. Responsibility for AI governance must not be siloed within technical or compliance functions but addressed collectively at senior and board level.	employees to raise concerns about unethical practices without fear of reprisal. Independent ethics specialists should be consulted to ensure balanced oversight. Governance structures must integrate cross-functional perspectives, strengthening internal accountability and public trust.	evaluate transparent reporting mechanisms. Programmes should include case-based exercises on responding to misconduct reports and guidance on integrating ethics committees into decision-making structures. Development efforts should support leaders in fostering a culture that normalises accountability.
3. Cultural and Behavioural Transformation	The success of the TRAIL framework depends on embedding responsible AI values into organisational culture and leadership behaviour. Culture shapes decision-making and signals priorities, making it essential that leaders actively model ethical responsibility and create conditions where responsible practice is expected and rewarded.	Leaders must foster a culture prioritising fairness, transparency, and accountability. This requires moving beyond technical safeguards toward cultivating behaviours and values that promote ethical awareness and integrity across the organisation.	Organisations should integrate responsible AI principles into mission, strategy, and operations. Structures and routines must support ethical decision-making at all levels, recognising the complexity of cultural dynamics and how organisational structures influence behaviour.	Training should encourage leaders to reflect on their behaviours and signals to teams. Scenario exercises can illustrate how leadership actions shape culture. Programmes should equip leaders with strategies to embed ethical principles in processes, performance management, and strategic planning.
4. Global and Cultural Nuances in Ethics	Responsible AI leadership requires sensitivity to cultural and regulatory diversity. Ethical principles vary across societies, and leaders must	Leaders must navigate diverse cultural contexts where norms and expectations vary. This includes regulatory fluency across jurisdictions and	Organisations should embed guidance on cultural variation into training and policy. This includes recognising geographical and	Programmes should offer case studies from global contexts, highlighting how ethical challenges differ. Scenario-based exercises should help

	balance these differences while maintaining consistent standards of fairness, accountability, and transparency.	ensuring organisational practices remain locally relevant and globally coherent.	demographic influences on ethical perceptions and building flexibility into governance structures while aligning with international standards.	leaders reconcile conflicting cultural or legal expectations, supported by dialogue with peers from diverse backgrounds.

5.11.11 Theoretical Alignment: Justifying the Intervention Strategy

The observed deficiency in leadership readiness, where high ethical awareness coexists with a lack of practical competence, mandates that the intervention strategies proposed by the TRAIL framework be grounded in organizational learning and change theory. Effective development must move beyond simple knowledge transfer to challenge the underlying values and assumptions that shape AI use, consistent with the concept of double-loop learning articulated by Argyris and Schön (1978).

5.12 Ethical Learning and Strategic Alignment

The majority of leaders acknowledge the necessity of acquiring the ability to assess AI systems for fairness and bias. This critical skills gap cannot be closed through passive training; it requires interventions that compel leaders to question the established organizational norms and governance mechanisms that may inadvertently perpetuate risk or bias.

The development techniques defined within the TRAIL framework, particularly in Pillar III (Ethical and Risk Literacy) and Pillar IV (Ethical Data Management), mandate the use of scenario-based simulations and case-based learning. These Double-Loop Learning techniques compel leaders to engage in active, applied ethical reasoning, thereby simulating real-world trade-offs and challenging the automatic response of merely complying with minimum standards. This structure provides the necessary

practice to translate theoretical ethical principles (such as fairness and human dignity) into practical, interrogative competence.

For the implementation of Responsible AI to be sustainable, the intervention must be systemic, aligning leadership behaviour with organisational culture, as modelled by the Burke-Litwin framework.

The TRAIL framework is designed as a holistic and strategic intervention. Pillar I (Strategic and Transformational Leadership) explicitly requires leaders to model responsible behaviour and align AI strategy with organizational values, thereby influencing culture and mitigating the risk of ethical principles being siloed within technical departments. Consistent with organizational change theory, the framework's emphasis on cross-functional collaboration and cultural transformation ensures that accountability is distributed, rather than being confined to individual technical competency.

By embedding ethical competence into governance structures and strategic planning, TRAIL addresses the finding that the rapid evolution of leadership roles is outpacing capability development. This theoretical alignment ensures that the framework provides not just a list of required skills, but a structured pathway to institutionalize AI leadership development, ensuring its impact is systemic and enduring.

Ultimately, the TRAIL framework's intervention strategies are deliberately designed to move leadership education from passive awareness to active, reflective practice, thereby serving as the necessary systemic structure required to manage the complex, adaptive challenge of responsible AI adoption.

5.13 TRAIL: A Structured, Empirically Derived Approach

The TRAIL framework represents the primary contribution of this research, offering a structured, empirically derived approach to resolving the critical gaps

identified in leadership competency, AI literacy, and organisational governance related to responsible AI adoption. The framework moves beyond mere conceptual awareness by providing senior professionals in the commercial sector with a holistic, actionable model to design, deploy, and oversee AI systems in ways that are ethically grounded and compliant with evolving regulations.

The framework is intentionally structured to integrate two essential components: Four Key Competency Pillars and comprehensive Supporting Mechanisms. The Pillars, covering Strategic and Transformational Leadership, Governance and Regulatory Fluency, Ethical and Risk Fluency, and Ethical Data Management, ensure that leaders align AI with organisational strategy, navigate complex regulatory environments, and embed ethical literacy into practice. The underlying purpose of this pillar structure is to provide structured guidance for the competencies necessary to address the substantial readiness gap currently persisting among leaders.

The managerial utility of TRAIL is significantly enhanced by its Supporting Mechanisms, which delineate the systemic structures essential for ensuring that ethical intentions are consistently translated into operational outcomes. These mechanisms emphasize measurable outcomes through rigorous evaluation, incorporate transparent accountability and collective oversight, facilitate cultural transformation, and consider global and cultural nuances in ethical interpretation.

Designed as a flexible tool rather than a rigid prescription, the TRAIL framework is adaptable across various industries and maturity levels. By integrating specific Competency Focus, Best Practice Guidance, and Development Techniques within each component, TRAIL directly equips leaders with the requisite skills, behaviours, and organizational insights. The implementation process, organised into four distinct phases,

provides a clear roadmap for institutionalizing AI leadership development, ensuring continuous education and the utilization of longitudinal reviews to monitor progress.

Ultimately, the TRAIL framework is positioned to mitigate the significant risks associated with non-implementation, including inadequate leadership preparedness and regulatory mismanagement, thereby fulfilling its purpose to strengthen public confidence, reduce risk, and support ethical, resilient, and sustainable AI deployment within organisations. The chapter now concludes with a Force Field diagram (Figure 5.8) depicting the dynamic interplay of forces that shape an organisation's ability to build and sustain Responsible AI capability and stewardship.

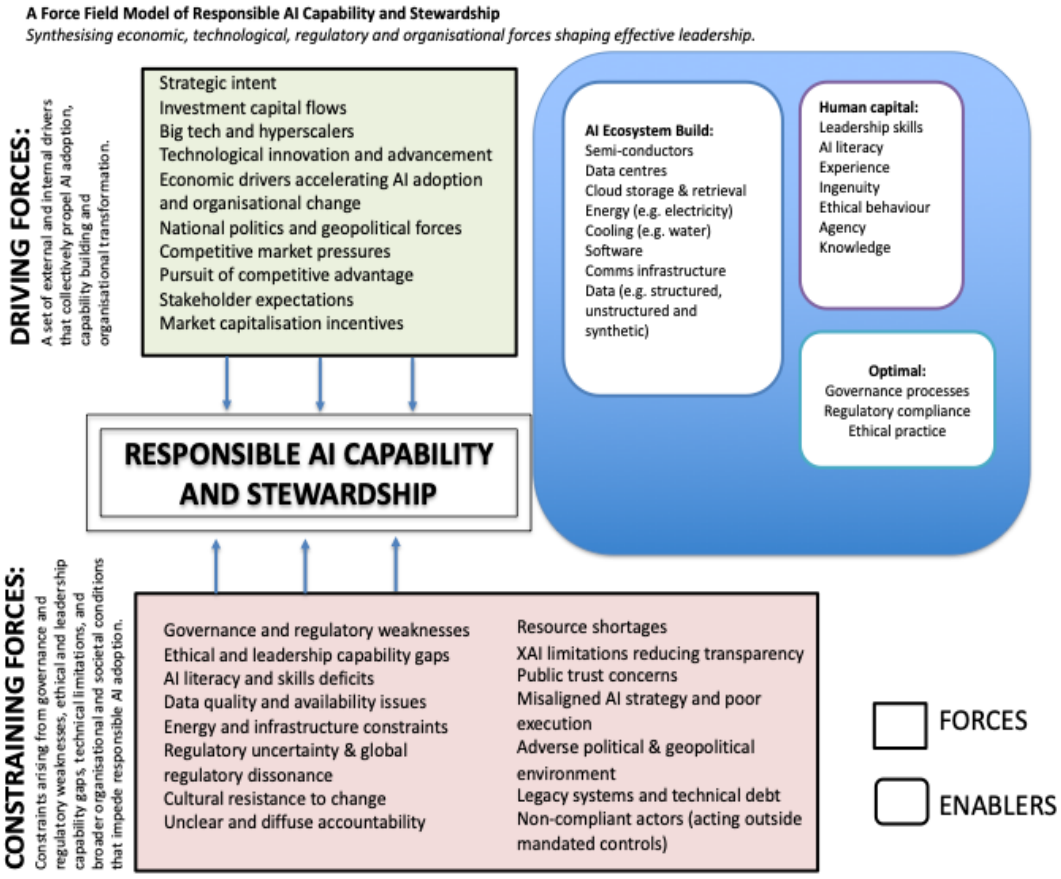


Figure 5.8: A Force Field Model of Responsible AI Capability and Stewardship

Diagram 5.8 is a force-field model providing a synthesised view of the contextual dynamics that shape an organisation's ability to achieve Responsible AI Capability and Stewardship. Positioned at the centre of this model, responsible AI is portrayed not as a technical end state but as a leadership and organisational maturity outcome. Its realisation depends on how effectively leaders navigate the opposing pressures that accelerate or hinder progress. This reflects one of the core arguments of the thesis: that responsible AI is fundamentally a question of organisational readiness, shaped by leadership competence, governance fluency, and ethical judgement.

The Driving Forces identified in the model align with the opportunities highlighted throughout the thesis. Innovation cycles, investment capital, competitive

pressures, and the influence of hyperscalers demonstrate the strong momentum toward AI adoption. These drivers create the strategic imperative for organisations to enhance their AI capability and, crucially, to embed responsible practices from the outset. In doing so, they reinforce TRAIL's assertion that leadership must proactively interpret external incentives and translate them into coherent strategies, structures, and behaviours that enable responsible and accountable AI deployment.

By contrast, the Constraining Forces correspond to the persistent vulnerabilities that limit organisational readiness. Skills shortages, misaligned strategies, poor execution, regulatory fragmentation, and inherent AI risks illustrate the practical and ethical challenges that leaders must confront. These constraints underscore the leadership deficits identified in the empirical findings, particularly in relation to governance literacy and ethical capability. The model therefore serves to highlight that the barriers to responsible AI are not primarily technological but organisational and systemic. It is here that TRAIL's contribution becomes most visible: by offering a structured leadership framework capable of addressing these internal gaps through improved competence, oversight, and cross-functional literacy.

The dual presence of politics and geopolitics as both drivers and constraints is particularly significant. Their ambivalent role reflects the wider complexity facing organisations that operate across multiple regulatory and geopolitical environments. This duality reinforces the conclusion that responsible AI stewardship requires leaders who can interpret and manage uncertainty, anticipate regulatory divergence, and embed ethical considerations into decision-making even when external pressures pull in conflicting directions. TRAIL equips leaders with precisely this capability by integrating ethical reasoning, governance awareness, and strategic adaptability.

Ultimately, the model offers a clear visual synthesis of the thesis's central claim: achieving Responsible AI Capability is contingent upon a leadership approach that can strengthen enabling forces while mitigating or strategically managing constraints. It reinforces that organisational readiness is dynamic rather than static, requiring continuous alignment between strategy, governance, ethics, and execution. By situating TRAIL within this landscape, the model emphasises the practical value of the framework as both a diagnostic and developmental tool for leaders seeking to navigate the accelerating pressures and emerging risks associated with AI adoption.

In this way, the model contributes to the broader argument of the thesis: that responsible AI will not be secured through technology alone. It will depend on leaders who possess the capacity, literacy, and ethical orientation to guide their organisations through a rapidly evolving environment. TRAIL provides a structured pathway toward that capability, offering a foundation for organisations seeking to develop the maturity, resilience, and stewardship required to adopt AI responsibly and sustainably.

CHAPTER VI

SUMMARY, IMPLICATIONS AND RECOMMENDATIONS

This research commenced with the premise that AI represents a paradigm shift comparable to the Industrial Revolution, offering profound economic opportunities alongside significant ethical and systemic risks. The investigation confirms that while the strategic potential of AI is widely recognised by commercial leaders, a critical 'readiness gap' persists between high-level aspirational commitment and the practical organisational capabilities required to operationalise responsible AI. The findings unequivocally demonstrate that ethical intent, absent structured governance and applied literacy, is insufficient to safeguard organisations against the complexities of algorithmic bias, regulatory fragmentation, and the erosion of human agency.

The leadership capability gaps identified in this study carry immediate regulatory implications. As AI regulation shifts from principles-based guidance to enforceable, risk-based regimes, particularly under frameworks such as the European Union AI Act, insufficient governance fluency at leadership level translates directly into compliance exposure. Where leaders lack confidence in regulatory interpretation, accountability structures, and oversight mechanisms, organisations become vulnerable not only to non-compliance but to enforcement actions, sanctions, and mandated system withdrawal. In this context, RAI is no longer a future compliance concern but an active leadership risk that requires urgent capability development.

Beyond regulatory exposure, the findings indicate a significant risk of strategic failure arising from the disconnect between high strategic awareness and low organisational preparedness. While leaders broadly recognise the transformative potential of AI, the absence of governance capability, applied ethical literacy, and structured guidance undermines execution. This gap risks reducing AI strategy to aspirational

rhetoric rather than operational reality, leading to stalled initiatives, inconsistent deployment, or unintended negative outcomes. The results suggest that leadership readiness, rather than technological ambition, is now the critical determinant of whether AI investments translate into sustainable value creation.

The reputational consequences of leadership shortcomings in Responsible AI are equally significant. The strong ethical expectations expressed by respondents, particularly in relation to fairness, transparency, and data stewardship, indicate that failures in AI governance are likely to be interpreted as failures of leadership judgement rather than technical error. In an environment of heightened public scrutiny and stakeholder awareness, incidents of AI misuse, bias, or opacity risk eroding organisational trust and legitimacy. These findings underscore that RAI governance is central to maintaining a social licence to operate, positioning leadership competence as a critical safeguard against reputational harm.

This study has shown that responsible AI cannot be achieved through technical safeguards or regulatory compliance alone. The empirical evidence highlights a persistent leadership readiness gap, in which ethical awareness and strategic intent are not matched by governance capability or operational preparedness. The TRAIL framework responds to this gap by translating principles of Responsible AI into a structured leadership capability model grounded in empirical findings. In doing so, this research establishes that the responsible use of AI is ultimately a question of leadership competence, exercised through governance, ethical judgement, and sustained organisational stewardship.

The central contribution of this study is the reconceptualization of responsible AI not as a peripheral compliance function, but as a core determinant of organisational resilience. By proposing Responsible AI Capability and Stewardship as a sixth strategic force

within Porter's framework, this research establishes that ethical integrity is now a competitive differentiator essential for long-term value creation. This theoretical advancement is operationalised through the TRAIL framework (Trustworthy and Responsible AI Leadership), which provides the necessary scaffolding to bridge the identified skills gap. By integrating Strategic Leadership, Governance Fluency, Ethical Literacy, and Data Management, TRAIL equips decision-makers to move beyond passive awareness toward active, competent oversight.

Ultimately, this study asserts that the trajectory of the AI era will be defined less by the sophistication of algorithms than by the quality of the human leadership guiding them. As organisations navigate an increasingly complex regulatory and socio-technical landscape, the imperative for leaders is to cultivate a culture where technology serves human dignity rather than diminishing it. Responsible AI is not a technological challenge to be solved, but a leadership responsibility to be exercised.

The future of AI is therefore not a question of technological possibility, but of leadership accountability and capability requiring immediate and structured intervention. It is the choices, competence, and ethical foresight of today's leaders that will determine whether AI remains a source of sustainable innovation or becomes a systemic liability.

REFERENCES

Aguilera, R.V. and Crespi-Cladera, R. (2016) ‘Global corporate governance: On the relevance of firms’ ownership structure’, *Journal of World Business*, 51(1), pp. 50–57.

Ahmed, I., Mia, R. and Shakil, N.A.F. (2023) ‘Mapping blockchain and data science to the cyber threat intelligence lifecycle’, *JACAIDMS* [online]. Available at: Link (Accessed: 30 October 2024).

Ajiga, D. (2024) ‘Review of ai techniques in financial forecasting: applications in stock market analysis’, *Finance & Accounting Research Journal* [online], 6(2), pp. 125-145. Available at: https://doi.org/10.51594/farj.v6i2.784 (Accessed: 30 October 2024).

Akbar, M., Khan, A., Mahmood, S., Rafi, S. and Demi, S. (2023) ‘Trustworthy artificial intelligence: a decision-making taxonomy of potential challenges’, *Software Practice and Experience* [online], 54(9), pp. 1621-1650. Available at: https://doi.org/10.1002/spe.3216 (Accessed: 30 October 2024).

Al, T. (2023) ‘Transparency in AI-driven marketing: A consumer perspective’, *Journal of Business Ethics and Technology*, 18(3), pp. 241–259.

Alborough, K. and Hansen, H. (2012) ‘Organisational development as a practice of becoming: A dialogic approach to the study of change’, *Journal of Organizational Change Management* [online], 25(2), pp. 232–248. Available at: https://doi.org/10.1108/09534811211213966 (Accessed: 30 October 2024).

Almeida, F. (2017) ‘Experience with entrepreneurship learning using serious games’, *Cypriot Journal of Educational Sciences* [online], 12(2), pp. 69–80. Available at: https://doi.org/10.184844/cjes.v12i2.1939 (Accessed: 30 October 2024).

Alvesson, M. and Einola, K. (2019) ‘Warning for excessive positivity: Authentic leadership and other traps in leadership studies’, *Leadership*, 15(4), pp. 405–423.

Alzghoul, A. and Alrawahna, A. S. (2025) ‘The Impact of Artificial Intelligence on Public Sector Decision-Making’, *International Review of Administrative Sciences*.

Ameen, N., Tarhini, A., Reppel, A. and Anand, A. (2021) ‘Customer experiences in the age of artificial intelligence’, *Computers in Human Behavior* [online], 114, 106548. Available at: https://doi.org/10.1016/j.chb.2020.106548 (Accessed: 30 October 2024).

Ananny, M. and Crawford, K. (2018) ‘Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability’, *New Media & Society*, 20(3), pp. 973–989.

Anfara, V. and Mertz, N. (2014) *Theoretical Frameworks in Qualitative Research*. Thousand Oaks, CA: SAGE.

Argyris, C. and Schön, D.A. (1978) *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

Argyris, C. and Schön, D.A. (1996) *Organizational Learning II: Theory, Method and Practice*. Reading, MA: Addison-Wesley.

Aristotle (2020) *Nicomachean Ethics*. Translated by A. Beresford. London: Penguin Classics.

Arora, A. and Sharma, C. (2016) 'Corporate governance and firm performance in developing countries: evidence from India', *Corporate Governance* [online], 16(2), pp. 420-436. Available at: https://doi.org/10.1108/cg-01-2016-0018 (Accessed: 30 October 2024).

Avolio, B.J. and Gardner, W.L. (2005) 'Authentic leadership development: Getting to the root of positive forms of leadership', *The Leadership Quarterly*, 16(3), pp. 315-338.

Axios (2025) *AI in education's potential privacy nightmare* [online]. Available at: https://www.axios.com/2025/08/14/ai-education-privacy (Accessed: 22 August 2025).

Ayeola, F. and Okunola, C. (2025) 'Data Governance in Smart Cities', *ResearchGate*.

Baran, B. E. and Woznyj, H. M. (2020) 'Managing VUCA: The role of agility and resilience in the 21st century', *Human Resource Management Review*, 30(2), 100705.

Bart, C. and Deal, K. (2006) 'The governance–performance relationship: The impact of board structure on firm outcomes', *Corporate Governance: An International Review*, 14(6), pp. 865–876.

Bashir, D. and Frank, C. (2024) 'The EU AI Act's Enforcement and Self-Taught Evaluators', *The Gradient* [online], 13 August. Available at: https://thegradientpub.substack.com/p/update-81-the-eu-ai-actsenforcement (Accessed: 31 October 2024).

Bass, B.M. and Avolio, B.J. (1994) *Improving organisational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage.

Bietti, E. (2020) 'From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy', *Minds and Machines*, 30(2), pp. 195–211.

Binns, R. (2018) 'Fairness in machine learning: Lessons from political philosophy', *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, pp. 149–159.

Birhane, A. (2021) ‘Algorithmic injustice: A relational ethics approach’, *Patterns*, 2(2), 100205.

Blackman, R. (2020) ‘A practical guide to building ethical AI’, *Harvard Business Review* [online]. Available at: https://academichelptoday.com/assets/documents/A_Practical_Guide_to_Building_Ethical_AI_-_HBR.pdf (Accessed: 27 October 2025).

Blackman, R. (2022) *Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI*. Dallas: HarperCollins Leadership.

Blackman, R. and Euchner, J. (2024) ‘Ethics in the Age of AI: A Conversation with Reid Blackman’, *Research-Technology Management* [online]. Available at: https://www.tandfonline.com/doi/abs/10.1080/08956308.2024.2280481 (Accessed: 27 October 2025).

Bodie, M.T., Cherry, M.A., McCormick, M.L. and Tang, J. (2020) ‘The law and policy of people analytics’, *University of Colorado Law Review*, 91(6), pp. 131–187.

Bonache, J. (2020) ‘The challenge of using a ‘non-positivist’ paradigm and getting through the peer-review process’, *Human Resource Management Journal* [online], 31(1), pp. 37-48. Available at: https://doi.org/10.1111/1748-8583.12319 (Accessed: 30 October 2024).

Bradford, A. (2021) *The Brussels Effect: How the European Union Rules the World*. Oxford: Oxford University Press.

Bradford, A. (2023) *Digital Empires: The Global Battle to Regulate Technology*. Oxford: Oxford University Press.

Bradford, A. (2024a) *Digital Empires: The Global Battle to Regulate Technology*. London: Penguin Press.

Bradford, A. (2024b) ‘Regulation and innovation: Reassessing the European approach to digital governance’, *Journal of European Public Policy* [online], 31(4), pp. 625–642. Available at: https://doi.org/10.1080/13501763.2024.2295181 (Accessed: 30 October 2024).

Bradford, A. (2024c) ‘The False Choice Between Digital Regulation and Innovation’, *Northwestern University Law Review* [online], 118(2), 6 October. Available at: https://ssrn.com/abstract=4753107 (Accessed: 30 October 2024).

Bryman, A. and Bell, E. (2015) *Business Research Methods*. 4th edn. Oxford: Oxford University Press.

Buiten, M.C. (2019) ‘Towards intelligent regulation of artificial intelligence’, *European Journal of Risk Regulation*, 10(1), pp. 41–59.

Burke, W.W. (2018) *Organization Change: Theory and Practice*. 6th edn. Thousand Oaks, CA: Sage.

Burke, W.W. and Litwin, G.H. (1992) 'A causal model of organisational performance and change', *Journal of Management*, 18(3), pp. 523–545.

Burnes, B. and Cooke, B. (2012) 'The past, present and future of organization development: Taking the long view', *Human Relations* [online], 65(11), pp. 1395-1429. Available at: https://journals.sagepub.com/doi/abs/10.1177/0018726712450058 (Accessed: 30 October 2024).

Burnes, B., Hughes, M. and By, R.T. (2018) *Reimagining Organisational Change*. London: Routledge.

Bushe, G. (2021) 'The generative change model: creating the agile organization while dealing with a complex problem', *The Journal of Applied Behavioural Science* [online], 57(4), pp. 530-533. Available at: https://doi.org/10.1177/00218863211038119 (Accessed: 30 October 2024).

Bushe, G. and Marshak, R. (2014a) 'Dialogic organization development', *Research in Organizational Change and Development*, pp. 193-211.

Bushe, G.R. and Marshak, R.J. (2014b) 'The dialogic mindset in organization development', *Research in Organizational Change and Development*, 22, pp. 55–97.

Cadwalladr, C. and Graham-Harrison, E. (2018) 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach', *The Guardian* [online], 17 March. Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (Accessed: 15 August 2025).

Caroli, L. (2024) *Will the EU AI Act work? Lessons learned from past legislative initiatives, future challenges* [online]. Available at: https://iapp.org/news/a/will-the-eu-ai-act-work-lessons-learned-from-past-legislative-initiatives-future-challenges (Accessed: 31 October 2024).

Carreyrou, J. (2018) *Bad Blood: Secrets and Lies in a Silicon Valley Startup*. New York: Knopf.

Cath, C. (2018) 'Governing artificial intelligence: Ethical, legal and technical opportunities and challenges', *Philosophical Transactions of the Royal Society A* [online], 376(2133), 20180080. Available at: https://doi.org/10.1098/rsta.2018.0080 (Accessed: 30 October 2024).

Cath, C. (2021) 'Governing artificial intelligence: Ethical, legal and technical opportunities and challenges', *Philosophical Transactions of the Royal Society A*, 379(2202), 20200360.

Cath, C. (2022) ‘Governing artificial intelligence: ethical, legal and technical opportunities and challenges’, *Philosophy & Technology*, 35(2), pp. 1–18.

Chelliah, J., Boersma, M. and Klettner, A. (2015) ‘Corporate governance and sustainability: An ethical perspective’, *Journal of Business Ethics*, 127(2), pp. 385–396.

Chen, J., Ramanathan, L. and Alazab, M. (2021) ‘Holistic big data integrated artificial intelligent modelling to improve privacy and security’, *Microprocessors and Microsystems* [online], 81. Available at: https://doi.org/10.1016/j.micpro.2021.103683 (Accessed: 30 October 2024).

Cheung Judge, M.Y. and Holbeche, L. (2015) *Organization development: A practitioner’s guide for OD and HR*. London: Kogan Page.

Chukwurah, N., Ige, A.B. and Idemudia, C. (2024) ‘Managing data lifecycle effectively: Best practices for data retention and archival processes’, *International Journal of Research*.

Chundru, S. and Mudunuri, L.N.R. (2025) ‘Developing sustainable data retention policies: A machine learning approach’ in *Driving Business Success through Eco-Innovation*. Hershey, PA: IGI Global.

Cihon, P., Maas, M.M. and Kemp, L. (2021) ‘Should artificial intelligence governance be centralised? Design lessons from history’, *Global Policy* [online], 12(S1), pp. 5–17. Available at: https://doi.org/10.1111/1758-5899.12928 (Accessed: 30 October 2024).

Collis, J. and Hussey, R. (2014) *Business Research: A Practical Guide for Undergraduate and Postgraduate Students*. 4th edn. Basingstoke: Palgrave Macmillan.

Colomo, P. (2021) ‘The draft digital markets act: a legal and institutional analysis’, *Journal of European Competition Law & Practice* [online], 12(7), pp. 561-575. Available at: https://doi.org/10.1093/jeclap/lpab065 (Accessed: 30 October 2024).

Consumer Financial Protection Bureau (2022) *CFPB acts to protect the public from black-box credit models using complex algorithms* [online]. Available at: https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/ (Accessed: 5 September 2025).

Cooke, B. (1998) ‘Participation, 'process' and management: Lessons for development in the history of organization development’, *Journal of International Development*, 10(1), pp. 35-54.

Copeland, M.K. (2014) ‘The emerging significance of values-based leadership: A literature review’, *International Journal of Leadership Studies*, 8(2), pp. 105-135.

Couldry, N. and Mejias, U.A. (2019) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press.

Crane, A. and Matten, D. (2021) *Business Ethics: Managing Corporate Citizenship and Sustainability in the Age of Globalization*. 6th edn. Oxford: Oxford University Press.

Creswell, J.W. and Plano Clark, V.L. (2018) *Designing and Conducting Mixed Methods Research*. 3rd edn. Thousand Oaks, CA: Sage.

Cummings, T.G. and Worley, C.G. (2019) *Organization Development and Change*. 11th edn. Boston, MA: Cengage Learning.

Dataguard (2024) *The growing data privacy concerns with AI: What you need to know* [online]. Available at: https://www.dataguard.com/blog/growing-data-privacy-concerns-ai (Accessed: 22 August 2025).

Day, D. and Dragoni, L. (2015) 'Leadership development: An outcome-oriented review based on time and levels of analyses', *Annual Review of Organizational Psychology and Organizational Behavior* [online], 2(1), pp. 133-156. Available at: https://doi.org/10.1146/annurev-orgpsych-032414-111328 (Accessed: 30 October 2024).

Day, D.V., Bastardoz, N., Bisbey, T., Reyes, D. and Salas, E. (2021) 'Unlocking human potential through leadership training & development initiatives', *Behavioral Science & Policy* [online], 7(1), pp. 41-54. Available at: doi:10.1177/237946152100700105 (Accessed: 30 October 2024).

Daza, L. and Ilozumba, P. (2022a) 'AI in business: Benefits, risks, and ethical imperatives', *Business Horizons*, 65(5), pp. 589–598.

Daza, M. and Ilozumba, U. (2022b) 'A survey of ai ethics in business literature: maps and trends between 2000 and 2021', *Frontiers in Psychology* [online], 13. Available at: https://doi.org/10.3389/fpsyg.2022.1042661 (Accessed: 30 October 2024).

De Silva, D. and Alahakoon, D. (2022) 'An artificial intelligence life cycle: From conception to production', *Patterns* [online], 3(6), pp. 1–12. Available at: https://doi.org/10.1016/j.patter.2022.100503 (Accessed: 30 October 2024).

Demidenko, E. and McNutt, P. (2010) 'The ethics of enterprise risk management as a key component of corporate governance', *International Journal of Social Economics*, 37(10), pp. 802–815.

Department for Science, Innovation and Technology (2023) *A pro-innovation approach to AI regulation*. London: HM Government.

Dignum, V. (2019) *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham: Springer.

Ding, J. (2021) *Deciphering China's AI dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI*. Oxford: Center for the Governance of AI.

Dotan, R. (2024) 'Unifying governance and regulation: A comparative analysis of the EU AI Act and the NIST AI Risk Management Framework', *AI and Society*, 39(1), pp. 112–130.

Drummond, J. J. (2002) 'Introduction: The Phenomenological Tradition and Moral Philosophy', *Springer Netherlands* [online], pp. 1–13. Available at: https://doi.org/10.1007/978-94-015-9924-5_1 (Accessed: 30 October 2024).

Duan, Y., Edwards, J. and Dwivedi, Y. (2019) 'Artificial intelligence for decision making in the era of big data – evolution, challenges and research agenda', *International Journal of Information Management* [online], 48, pp. 63-71. Available at: https://doi.org/10.1016/j.ijinfomgt.2019.01.021 (Accessed: 30 October 2024).

Easterby-Smith, M., Thorpe, R. and Jackson, P. (2021) *Management and Business Research*. 7th edn. London: Sage.

Ebers, M. (2022) 'Standardising AI – The case of the European Commission's proposal for an Artificial Intelligence Act', *Computer Law and Security Review* [online], 45, 105681. Available at: https://doi.org/10.1016/j.clsr.2022.105681 (Accessed: 30 October 2024).

EDPB (2025) *AI Privacy Risks & Mitigations – Large Language Models (LLMs)* [online]. Available at: https://edpb.europa.eu (Accessed: 22 August 2025).

Eitel-Porter, R. (2021) 'Ethical AI – An imperative for trust', *Digital Transformation Journal*, 2(1), pp. 5–10.

Ess, C. (2020) *Ethics and Information Technology: A Case-Based Approach to a Healthier Information Society*. 2nd edn. London: Routledge.

European Commission (2018) *General Data Protection Regulation (GDPR)* [online]. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj (Accessed: 19 June 2025).

European Commission (2021) *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Brussels: European Commission.

European Commission (2024) *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*. Brussels: European Commission.

European Union (2022a) *Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act)*. *Official Journal of the European Union*, L265, pp. 1–66. Available at: https://eur-lex.europa.eu/eli/reg/2022/1925/oj (Accessed: 23 December 2025).

European Union (2022b) *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act)*. *Official Journal of the European Union*, L277, pp. 1–102. Available at: https://eur-lex.europa.eu/eli/reg/2022/2065/oj (Accessed: 23 December 2025).

European Union (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, L series. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj (Accessed: 17 December 2025).

Favaretto, M., Clercq, E., Gaab, J. and Elger, B. (2020) 'First do no harm: an exploration of researchers' ethics of conduct in big data behavioral studies', *Plos One* [online], 15(11), e0241865. Available at: https://doi.org/10.1371/journal.pone.0241865 (Accessed: 30 October 2024).

Finocchiaro, A. (2024a) 'The geopolitics of the EU Artificial Intelligence Act: Regulation as strategy', *Journal of European Integration*, 46(2), pp. 211–229.

Finocchiaro, G. (2024b) 'The regulation of artificial intelligence', *AI & Society*, Springer.

Fisher, R.J. (1993) 'Social Desirability Bias and the Validity of Indirect Questioning', *Journal of Consumer Research*, 20(2), pp. 303–315.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020) 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI', *Berkman Klein Center Research Publication*, (2020-1), pp. 1–51.

Flavián, C., Pérez-Rueda, A., Belanche, D. and Casaló, L. (2021) 'Intention to use analytical artificial intelligence (ai) in services – the effect of technology readiness and awareness', *Journal of Service Management* [online], 33(2), pp. 293-320. Available at: https://doi.org/10.1108/josm-10-2020-0378 (Accessed: 30 October 2024).

Fletcher, A. (2023) 'International pro-competition regulation of digital platforms: healthy experimentation or dangerous fragmentation?', *Oxford Review of Economic Policy*

[online], 39(1), pp. 12-33. Available at: https://doi.org/10.1093/oxrep/grac047 (Accessed: 30 October 2024).

Floridi, L. (2019) 'Translating principles into practices of digital ethics: Five risks of being unethical', *Philosophy & Technology* [online]. Available at: link.springer.com (Accessed: 18 October 2024).

Floridi, L. (2021) *Digital ethics online and off* [online]. Available at: philarchive.org (Accessed: 18 October 2024).

Floridi, L. (2022) *Ethics, Governance, and Policies for Artificial Intelligence*. Oxford: Oxford University Press.

Floridi, L. (2023) *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018) 'AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations', *Minds and Machines*, 28(4), pp. 689–707.

Fortes, P., Santos, L. and Diniz, E. (2022) 'Algorithmic governance: Risks, regulation and public administration', *Government Information Quarterly*, 39(4), 101740.

French, S. (2013) 'Cynefin, statistics and decision analysis', *Journal of the Operational Research Society*, 64(3), pp. 336–343.

FTC (2019a) *FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook* [online]. Available at: https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook (Accessed: 25 October 2025).

FTC (2019b) *FTC, CFPB and States announce settlement with Equifax related to 2017 data breach* [online]. Available at: https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-cfpb-states-announce-settlement-equifax-related-2017-data-breach (Accessed: 9 December 2025).

FTC (2020) *Using artificial intelligence and algorithms.* Washington, DC: Federal Trade Commission. Available at: https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms
(Accessed: 27 November 2025).

FTC (2023) *Keep your AI claims in check* [online]. Available at: https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check (Accessed: 15 August 2025).

Ganapathi, S. and Duggal, S. (2023) 'Exploring the experiences and views of doctors working with artificial intelligence in english healthcare; a qualitative study', *Plos One* [online], 18(3), e0282415. Available at: https://doi.org/10.1371/journal.pone.0282415 (Accessed: 30 October 2024).

Garrigues (2024) Regulating AI in the EU, US and OECD: the difficult balance between security and driving innovation. Available at: https://www.garrigues.com/ (Accessed: 31 December 2025).

Gasser, U. and Almeida, V. (2017) 'A layered model for ai governance', *IEEE Internet Computing* [online], 21(6), pp. 58-62. Available at: https://doi.org/10.1109/mic.2017.4180835 (Accessed: 30 October 2024).

Godenhjelm, S. (2016) *Project organisations and governance: Processes, actors, actions, and participatory procedures*. Helsinki: University of Helsinki.

Greene, D., Hoffmann, A.L. and Stark, L. (2019) 'Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning', *Proceedings of the 52nd Hawaii International Conference on System Sciences*, pp. 2122–2131.

Greenwood, M. (2022) 'Stakeholder engagement: Beyond the rhetoric', *Journal of Business Ethics*, 178(2), pp. 315–329.

Grieves, J. (2000) 'The origins of organisational development', *Journal of Management Development*, 19(5), pp. 345–447.

Gromova, E., Juchnevicius, E. and Cyman, D. (2021a) 'Regulation of artificial intelligence in BRICS and the European Union', *Brics Law Journal*.

Gromova, E., Sinyukova, A. and Shkalenko, A. (2021b) 'Comparative legal analysis of AI regulation in BRICS and the European Union', *Journal of Law, Technology and Public Policy*, 7(2), pp. 134–151.

Gudivada, V.N., Apon, A. and Ding, J. (2017) 'Data quality considerations for big data and machine learning', *International Journal on Advances in Software* [online], 10(1–2), pp. 1–20. Available at: Link (Accessed: 30 October 2024).

Guo, W. (2023) 'Exploring the value of ai technology in optimizing and implementing supply chain data for pharmaceutical companies', *Innovation in Science and Technology*

[online], 2(3), pp. 1-6. Available at: https://doi.org/10.56397/ist.2023.05.01 (Accessed: 30 October 2024).

Hacker, P. (2023) ‘AI regulation in the European Union: Balancing innovation, fairness and fundamental rights’, *Common Market Law Review*, 60(2), pp. 341–372.

Hair, J.F., Page, M. and Brunsveld, N. (2020) *Essentials of Business Research Methods*. 2nd edn. London: Routledge.

Hamon, R. (2024) ‘Three challenges to secure ai systems in the context of ai regulations’, *IEEE Access* [online], 12, pp. 61022-61035. Available at: https://doi.org/10.1109/access.2024.3391021 (Accessed: 30 October 2024).

Hanna, M.G. (2025) *Ethical and Bias Considerations in Artificial Intelligence*. London: Routledge.

Hanna, N. and Kazim, E. (2021) ‘A dignitarian approach to artificial intelligence ethics’, *AI and Ethics*, 1(1), pp. 1–13.

Hasselbalch, G. (2022) *Data Ethics of Power: A Human Approach in the Big Data and AI Era*. Cheltenham: Edward Elgar Publishing.

Häußermann, J. and Lütge, C. (2021a) ‘Community-in-the-loop: towards pluralistic value creation in ai, or—why ai needs business ethics’, *AI and Ethics* [online], 2(2), pp. 341-362. Available at: https://doi.org/10.1007/s43681-021-00047-2 (Accessed: 30 October 2024).

Häußermann, J.J. and Lütge, C. (2021b) ‘Business ethics and AI: Ethical challenges of algorithmic decision-making’, *AI and Society*, 36(1), pp. 49–59.

Hill, K. (2023) ‘Facial recognition leads to mistaken arrest in Louisiana’, *The New York Times* [online], 7 January. Available at: https://www.nytimes.com (Accessed: 14 September 2025).

Hinckley, S. R. Jr. (2014) ‘A history of organization development’ in *The NTL Handbook of Organization Development and Change*. Hoboken, NJ: Wiley, pp. 50-62.

Hoffmann-Riem, W. (2020) ‘Regulating artificial intelligence – conceptual perspectives’, *The Hague Journal on the Rule of Law*, 12, pp. 5–32.

Holistic AI (2022) *The five pillars of responsible AI*. London: Holistic AI.

Holzhausen, J. (2024) ‘Legal accountability and AI: Challenges for business governance’, *Journal of Law, Innovation and Technology*, 16(1), pp. 78–93.

Homeyer, A., Lotz, J., Schwen, L., Weiss, N., Romberg, D. and Höfener, H. (2021) 'Artificial intelligence in pathology: from prototype to product', *Journal of Pathology Informatics* [online], 12(1), 13. Available at: https://doi.org/10.4103/jpi.jpi_84_20 (Accessed: 30 October 2024).

Huang, M. and Rust, R. (2018) 'Artificial intelligence in service', *Journal of Service Research* [online], 21(2), pp. 155-172. Available at: https://doi.org/10.1177/1094670517752459 (Accessed: 30 October 2024).

Huang, M.H., Rust, R.T. and Maksimovic, V. (2020) 'The Feeling Economy: How Artificial Intelligence Is Creating a New Era of Empathy', *Journal of the Academy of Marketing Science*, 48, pp. 855–876.

Hunkenschroer, A. and Luetge, C. (2022a) 'Ethical challenges of AI in human resource management', *AI and Ethics*, 2(1), pp. 13–23.

Hunkenschroer, A. and Luetge, C. (2022b) 'Ethics of ai-enabled recruiting and selection: a review and research agenda', *Journal of Business Ethics* [online], 178(4), pp. 977-1007. Available at: https://doi.org/10.1007/s10551-022-05049-6 (Accessed: 30 October 2024).

Huriye, B. (2023) 'Accountability and the opacity of artificial intelligence systems: Legal and ethical challenges', *AI and Ethics*, 3(2), pp. 245–259.

IDC (2022) *Worldwide Global DataSphere Forecast, 2022–2026*. Framingham, MA: International Data Corporation.

Infosys Knowledge Institute (2022) *Responsible AI: Enabling resilience, growth and positive outcomes* [online]. Available at: https://www.infosys.com/iki (Accessed: 15 September 2025).

Infosys Knowledge Institute (2025) *Responsible enterprise AI in the agentic era* [online]. Available at: https://www.infosys.com/iki/research/responsible-enterprise-ai-agentic.html (Accessed: 30 October 2025).

Institute for Organisation Development (2024) *Dialogic OD: Integrating Emergent Change Practices* [online]. Available at: https://instituteod.com (Accessed: 28 October 2025).

Jacobides, M.G., Cennamo, C. and Gawer, A. (2018) 'Towards a theory of ecosystems', *Strategic Management Journal*, 39(8), pp. 2255–2276.

Jobin, A. and Ienca, M. (2019) 'The global landscape of ai ethics guidelines', *Nature Machine Intelligence* [online], 1(9), pp. 389-399. Available at: https://doi.org/10.1038/s42256-019-0088-2 (Accessed: 30 October 2024).

Joint Authorities Technical Review (2019) *Boeing 737 MAX Flight Control System: Findings and Recommendations*. Washington, DC: Federal Aviation Administration.

Justo-Hanani, R. (2022) 'The European Union's emerging approach to artificial intelligence regulation', *Policy and Internet* [online], 14(3), pp. 648–666. Available at: https://doi.org/10.1002/poi3.281 (Accessed: 30 October 2024).

Kamau, G. (2022) 'Ict4d research in developing countries: a call for pragmatism approach', *International Journal of Computer and Information System (Ijcis)* [online], 3(2), pp. 51-55. Available at: https://doi.org/10.29040/ijcis.v3i2.67 (Accessed: 30 October 2024).

Kant, I. (2019) *Groundwork of the Metaphysics of Morals*. London: Oxford World's Classics.

Kaptein, M. (2022) 'The effectiveness of ethics programmes: The role of scope, composition, and sequence', *Journal of Business Ethics*, 176(3), pp. 567–585.

Kate, A. (2025) 'Designing a Risk Management Framework for Algorithmic Governance', *ResearchGate*.

Katz, E. (1991) 'The call of the wild: The struggle against domination in nature philosophy', *Environmental Ethics*, 13(3), pp. 265–273.

Kazim, E. and Koshiyama, A. (2020) 'The interrelation between data and AI ethics in the context of impact assessments', *AI and Ethics*.

Kiggins, R.D. (2023) *The Political Economy of Artificial Intelligence: Power, Politics and Policy in the Digital Age*. London: Routledge.

Kirsch, L.J. (2021) 'Making sense of complexity: A critical reflection on the Cynefin framework', *Management Learning*, 52(4), pp. 489–506.

Kwon, C. and Kwon, K. (2023) 'A conceptual framework for practicing inclusive dialogic organization development in times of uncertainty and complexity', *European Journal of Training and Development* [online], 48(5/6), pp. 592-608. Available at: https://doi.org/10.1108/ejtd-11-2022-0120 (Accessed: 30 October 2024).

Löfström, E., Poom-Valickis, K., and Heikonen, L. (2021) 'The ethics of care in higher education: A moral imperative or professional burden?', *Ethics and Education*, 16(2), pp. 167–183.

Long, D. and Magerko, B. (2020) 'What is AI Literacy? Competencies and Design Considerations', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–16.

Longe, T. (2023) ‘A qualitative exploration of effective leadership development programs and leadership processes on succession planning and organisational performance’, *International Journal of Business Administration* [online], 14(4), p. 33. Available at: https://doi.org/10.5430/ijba.v14n4p33 (Accessed: 30 October 2024).

Macintyre, A. (2003) *A Short History of Ethics*. London: Routledge.

Majumder, T. (2024) ‘The evaluating impact of artificial intelligence on risk management and fraud detection in the commercial bank in bangladesh’, *IJANS* [online], 1(1), pp. 67-76. Available at: https://doi.org/10.61424/ijans.v1i1.75 (Accessed: 30 October 2024).

Mancosu, M. and Vegetti, F. (2020) ‘What you can scrape and what is right to scrape: a proposal for a tool to collect public facebook data’, *Social Media + Society* [online], 6(3). Available at: https://doi.org/10.1177/2056305120940703 (Accessed: 30 October 2024).

Marcus, G. (2022) ‘The Next Decade in AI: Why Common Sense Matters More than Big Data’ in Fjeld, J. and Rahwan, I. (eds.) *Responsible AI*. Cambridge, MA: MIT Press.

Marcus, G. and Davis, E. (2020) *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Vintage.

Marshak, R. and Bushe, G. (2021) ‘A compelling beginning and more to uncover’, *The Journal of Applied Behavioral Science* [online], 58(1), pp. 149-152. Available at: https://doi.org/10.1177/00218863211060884 (Accessed: 30 October 2024).

Marshak, R. J. and Grant, D. (2008) ‘Organizational discourse and new organization development practices’, *British Journal of Management*, 19(3), pp. 7-19.

Marshak, R.J. and Grant, D. (2011) ‘The new OD: A critical examination of organisation development’, *Journal of Applied Behavioral Science,* 47(4), pp. 507–531.

Martin, K. (2022) ‘Ethical AI and organisational practice: From principles to implementation’, *Journal of Business Ethics,* 179(4), pp. 851–870.

Matta, V. (2015) ‘A Critical Review of Design Science in Information Systems Research’, *Journal of Information Technology Theory and Application*, 16(2), pp. 5–20.

McDermid, J.A., Jia, Y., Porter, Z. and Habli, I. (2021) ‘Artificial intelligence explainability: the technical and ethical dimensions’, *Philosophical Transactions of the Royal Society A* [online], 379, 20200363. Available at: https://doi.org/10.1098/rsta.2020.0363 (Accessed: 30 October 2024).

McIntyre, R. (2023) ‘Decentralisation and recentralisation: an institutional analysis of eu competition law and the digital markets act’, *LSELR* [online], 8(3), pp. 227-285. Available at: https://doi.org/10.61315/lselr.516 (Accessed: 30 October 2024).

McKinsey & Company (2023) *The State of AI in 2023: Generative AI’s Breakout Year* [online]. Available at: https://www.mckinsey.com (Accessed: 22 August 2025).

McKinsey & Company (2024) *The economic potential of generative AI: The next productivity frontier.* New York: McKinsey Global Institute. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (Accessed: 28 November 2025).

McKinsey & Company (2025) *The state of AI: How organisations are rewiring to capture value* [online]. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai (Accessed: 23 December 2025).

Mejias, U.A. and Couldry, N. (2024) *Data Grab: The New Colonialism of Big Tech (and How to Fight Back).* London: WH Allen.

Mill, J.S. (1998) *Utilitarianism*. London: Oxford Philosophical Texts.

Minkkinen, M., Niukkanen, A. and Mäntymäki, M. (2022) ‘What about investors? esg analyses as tools for ethics-based ai auditing’, *AI & Society* [online], 39(1), pp. 329-343. Available at: https://doi.org/10.1007/s00146-022-01415-0 (Accessed: 30 October 2024).

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B. and Gebru, T. (2019) ‘Model cards for model reporting’, *Proceedings of the Conference on Fairness, Accountability, and Transparency* [online]. Available at: https://doi.org/10.1145/3287560.3287596 (Accessed: 30 October 2024).

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) ‘The ethics of algorithms: Mapping the debate’, *Big Data & Society*, 3(2), pp. 1–21.

Mökander, J. and Floridi, L. (2023a) ‘A unified framework for AI governance: Integrating regulatory compliance and ethical oversight’, *Computers and Society*, 9(3), pp. 215–233.

Mökander, J. and Floridi, L. (2023b) ‘Operationalising AI governance through ethics-based auditing: an industry case study’, *AI and Ethics*, 3(1), pp. 1–14.

Mökander, J., Axente, M., Casolari, F. and Floridi, L. (2021) ‘Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations’, *Science and Engineering Ethics*, 27(4), pp. 1–23.

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2019) 'From what to how: an initial review of publicly available ai ethics tools, methods and research to translate principles into practices', *Science and Engineering Ethics* [online], 26(4), pp. 2141-2168. Available at: https://doi.org/10.1007/s11948-019-00165-5 (Accessed: 30 October 2024).

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2021) 'From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices', *Science and Engineering Ethics*, 27(1), pp. 1–31.

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2023) 'Operationalising AI ethics: Barriers, solutions and future directions', *Science and Engineering Ethics*, 29(1), pp. 1–19.

Nagbøl, P., Müller, O. and Krancher, O. (2021) 'Designing a risk assessment tool for artificial intelligence systems', *Lecture Notes in Computer Science* [online], pp. 328-339. Available at: https://doi.org/10.1007/978-3-030-82405-1_32 (Accessed: 30 October 2024).

Naik, N., Hameed, B., Shetty, D., Swain, D., Shah, M., Paul, R. and Somani, B. (2022) 'Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?', *Frontiers in Surgery* [online], 9. Available at: https://doi.org/10.3389/fsurg.2022.862322 (Accessed: 30 October 2024).

Naz, M. (2024) 'Artificial intelligence in marketing: Balancing innovation with ethical accountability', *Journal of Business and Technology*, 12(1), pp. 75–89.

New York City Government (2023) *Local Law 144 of 2021: Automated Employment Decision Tools – Final Rule* [online]. Available at: https://rules.cityofnewyork.us/rule/automated-employment-decision-tools/ (Accessed: 9 December 2025).

New York Post (2025) *AI-powered hiring tools favour Black and female job candidates over White and male applicants: study* [online]. Available at: https://nypost.com/2025/06/24/business/ai-hiring-tools (Accessed: 22 August 2025).

Nurhayati, Y. (2023) 'Regulatory analysis digital markets act (dma) european union in business competition', *International Journal of Law Environment and Natural Resources* [online], 3(1), pp. 107-174. Available at: https://doi.org/10.51749/injurlens.v3i1.46 (Accessed: 30 October 2024).

OECD (2021) *OECD AI principles: Recommendations of the Council on Artificial Intelligence* [online]. Available at: https://oecd.ai (Accessed: 15 September 2025).

Oladele, A. (2024) 'Ethical deployment of AI in business organisations: A governance perspective', *Journal of Corporate Responsibility and Leadership*, 11(2), pp. 121–137.

Olatoye, F. (2024) ‘Ai and ethics in business: a comprehensive review of responsible ai practices and corporate responsibility’, *International Journal of Science and Research Archive* [online], 11(1), pp. 1433-1443. Available at: https://doi.org/10.30574/ijsra.2024.11.1.0235 (Accessed: 30 October 2024).

OpenForum (2025) *We are at an inflection point with AI. Which way will we go?* [online]. Available at: https://www.sfchronicle.com/opinion/openforum/article/artificial-intelligence-data-trust (Accessed: 22 August 2025).

Ospina, S. and Foldy, E.G. (2009) ‘A critical review of race and ethnicity in the leadership literature: Surfacing context, power, and the collective dimensions of leadership’, *The Leadership Quarterly*, 20(6), pp. 876-896.

Oyeniyi, L. (2024) ‘The influence of ai on financial reporting quality: a critical review and analysis’, *World Journal of Advanced Research and Reviews* [online], 22(1), pp. 679-694. Available at: https://doi.org/10.30574/wjarr.2024.22.1.1157 (Accessed: 30 October 2024).

Paley, J. (2010) ‘The appropriation of complexity theory in health care’, *Journal of Health Services Research & Policy*, 15(1), pp. 59-61.

Panchal, T. and Panchal, D. (2025) ‘AI Ethics and Governance in Business Management’, *AI and Ethics*, Springer.

Park, Y., Konge, L. and Artino, A.R. (2020) ‘The positivism paradigm of research’, *Academic Medicine* [online], 95(5), pp. 690–694. Available at: https://doi.org/10.1097/ACM.0000000000003093 (Accessed: 30 October 2024).

Pearce, C.L. and Conger, J.A. (2003) *Shared Leadership: Reframing the Hows and Whys of Leadership*. Thousand Oaks, CA: Sage Publications.

Peddle, J., Thorne, S., McIntyre, L. and Canam, C. (2020) ‘Integrating quantitative and qualitative phases in research design: methodological considerations’, *Journal of Mixed Methods Research*, 14(3), pp. 367–381.

Pell, D. (2016) ‘The Cynefin Framework: Applying decision-making concepts to public policy’, *Public Policy and Administration*, 31(2), pp. 132-150.

Plato (2020) *Republic*. Humboldt: Prairie Publishing Ltd.

Porter, M.E. (1979) ‘How competitive forces shape strategy’, *Harvard Business Review*, 57(2), pp. 137–145.

Porter, M.E. (2008) *The Five Competitive Forces That Shape Strategy*. Boston: Harvard Business School Publishing.

Presthus, W. and Sønslien, K. (2021) 'An analysis of violations and sanctions following the gdpr', *International Journal of Information Systems and Project Management* [online], 9(1), pp. 38-53. Available at: https://doi.org/10.12821/ijispm090102 (Accessed: 30 October 2024).

Price, W. and Cohen, I. (2019) 'Privacy in the age of medical big data', *Nature Medicine* [online], 25(1), pp. 37-43. Available at: https://doi.org/10.1038/s41591-018-0272-7 (Accessed: 30 October 2024).

Purdy, N. (2016) 'Impact of a leadership development institute on professional lives and careers', *Canadian Journal of Nursing Leadership* [online], 29(2), pp. 10-30. Available at: https://doi.org/10.12927/cjnl.2016.24811 (Accessed: 30 October 2024).

PwC (2017) *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* London: PricewaterhouseCoopers. Available at: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf (Accessed: 27 October 2025).

PwC (2023) *Sizing the prize: What's the real value of AI for your business and how can you capitalise?* London: PricewaterhouseCoopers. Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html
(Accessed: 15 November 2025).

Qualys (2025) *AI and Data Privacy: Mitigating Risks and Ensuring Protection* [online]. Available at: https://blog.qualys.com/product-tech/2025/02/07/ai-and-data-privacy (Accessed: 22 August 2025).

Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K. (2020) 'Mitigating bias in algorithmic hiring: Evaluating claims and practices', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT)**, pp. 469–481.

Rahwan, I. et al. (2025) 'Responsible artificial intelligence governance: a review and research agenda', Technological Forecasting and Social Change. Available at: https://www.sciencedirect.com/ (Accessed: 31 December 2025).

Reason, P. and McArdle, K.L. (2006) 'Action research and organisation development' in Clegg, S.R., Hardy, C., Lawrence, T.B. and Nord, W.R. (eds.) *The Sage Handbook of Organization Studies*. 2nd edn. London: Sage, pp. 533–553.

Reddy, S., Reddy, B., Dodda, S. and Raparthi, M. (2020) ‘Ethical considerations in AI enabled Big Data research: Balancing innovation and privacy’, *International Journal of Control and Automation* [online]. Available at: https://doi.org/10.52783/ijca.v13i03.38350 (Accessed: 5 September 2025).

Remenyi, D., Money, A., Price, D. and Bannister, F. (2010) ‘The Concept of the Methodology in Information Systems Research’ in Remenyi, D. (ed.) *A Guide to the Methodology of Information Systems*. Reading: Academic Conferences Limited, pp. 1–15.

Reuters (2025) *State AGs fill the AI regulatory void* [online]. Available at: https://www.reuters.com/legal/legalindustry/state-ags-fill-ai-regulatory-void-2025-05-19 (Accessed: 22 August 2025).

Robson, C. and McCartan, K. (2016) *Real World Research*. 4th edn. Chichester: Wiley.

Rothwell, W.J. and Stavros, J.M. (2015) *Practicing organization development: Leading transformation and change*. 4th edn. Hoboken, NJ: Wiley.

Sacchetti, S. and Borzaga, C. (2021) ‘Stakeholder governance: Concepts and issues’, *Annals of Public and Cooperative Economics,* 92(3), pp. 423–446.

Sackett, P. R. and Lievens, F. (2025) ‘Hiring People in Organizations’, *Annual Review of Organizational Psychology*.

Saunders, M. and Lewis, P. (2014) *Doing Research in Business and Management: An Essential Guide to Planning Your Project*. Harlow: Pearson Education.

Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students*. 8th edn. Harlow: Pearson Education.

Sawah, S., Tharwat, A. and Rasmy, M. (2008) ‘A quantitative model to predict the Egyptian ERP implementation success index’, *Business Process Management Journal* [online], 14(3), pp. 288–306. Available at: https://doi.org/10.1108/14637150810876643 (Accessed: 30 October 2024).

Scharmer, O. and Kaufer, K. (2013) *Leading from the Emerging Future: From Ego-System to Eco-System Economies*. San Francisco: Berrett-Koehler Publishers.

Scharre, P. (2023) *Four Battlegrounds: Power in the Age of Artificial Intelligence*. New York: W.W. Norton & Company.

Schein, E.H. (2017) *Organizational culture and leadership*. 5th edn. Hoboken, NJ: Wiley.

Schiff, D., Biddle, J., Borenstein, J. and Laas, K. (2020) ‘What's next for ai ethics, policy, and governance? a global overview’, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* [online]. Available at: https://doi.org/10.1145/3375627.3375804 (Accessed: 30 October 2024).

Schönberger, M. (2023) ‘Artificial intelligence for small and medium-sized enterprises: identifying key applications and challenges’, *Journal of Business Management* [online], 21, pp. 89-112. Available at: https://doi.org/10.32025/jbm23004 (Accessed: 30 October 2024).

Schwab, K. (2016) *The Fourth Industrial Revolution*. Geneva: World Economic Forum.

Shahid, S. et al. (2025) ‘Artificial Intelligence Applications in Retail Banking’, *Academia International Journal*.

Shahriar, S., Allana, S., Hazratifard, S.M. and Dara, R. (2023) ‘A survey of privacy risks and mitigation strategies in the artificial intelligence life cycle’, *IEEE Access* [online], 11. Available at: https://ieeexplore.ieee.org/document/10155147 (Accessed: 30 October 2024).

Shneiderman, B. (2020) ‘Human-centered artificial intelligence: Reliable, safe & trustworthy’, *International Journal of Human-Computer Interaction*, 36(6), pp. 495–504.

Shreve, J., Khanani, S. and Haddad, T. (2022) ‘Artificial intelligence in oncology: current capabilities, future opportunities, and ethical considerations’, *American Society of Clinical Oncology Educational Book* [online], 42, pp. 842-851. Available at: https://doi.org/10.1200/edbk_350652 (Accessed: 30 October 2024).

Shukla, P. (2025) ‘Understanding and Evaluating Computer Vision Models through the Lens of Counterfactuals’, *arXiv preprint arXiv:2508.20881*.

Simarmata, H., Karo, A., Sari, L., Hendrawan, D. and Aditya, R. (2022) ‘The motivation of fitness center members in doing fitness at NR Gym Fitness Center Medan’, *Acpes Journal of Physical Education Sport and Health (Ajpesh)* [online], 2(1), pp. 24–31. Available at: https://doi.org/10.15294/ajpesh.v2i1.55227 (Accessed: 30 October 2024).

Simply Sustainable (2025) AI series: governance of AI – accountability and transparency. Available at: https://simplysustainable.com/ (Accessed: 31 December 2025).

Sinclair, C. (2012) ‘Ethical Principles, Values, and Codes for Psychologists: An Historical Journey’, *Oxford University* [online], pp. 3–18. Available at: https://doi.org/10.1093/oxfordhb/9780199739165.013.0001 (Accessed: 30 October 2024).

Singh, R. (2010) *Ethics in Research*. New Delhi: A.P.H. Publishing Corporation.

Smuha, N.A. (2021) 'Beyond a human rights-based approach to AI governance: Promise or peril?', *Philosophy and Technology* [online], 34, pp. 1607–1630. Available at: https://doi.org/10.1007/s13347-021-00480-8 (Accessed: 30 October 2024).

Snowden, D. J. (2003) 'Complex acts of knowing: Paradox and descriptive self-awareness', *Journal of Knowledge Management*, 6(2), pp. 100-111.

Snowden, D. J. and Boone, M. E. (2007) 'A leader's framework for decision making', *Harvard Business Review*, 85(11), pp. 68-76.

Stahl, B. C. (2022) 'From computer ethics and the ethics of AI towards an ethics of digital ecosystems', *AI and Ethics* [online]. Available at: springer.com (Accessed: 18 October 2024).

Stahl, B.C. (2021) *Artificial Intelligence for a Better Future: An ecosystem perspective on the ethics of AI and emerging digital technologies*. Cham: Springer.

Stanar, C. (2023) 'Beyond the human: Environmental ethics in the age of AI', *Journal of Applied Philosophy,* 40(1), pp. 55–70.

Stanford University (2025) *AI Index Report 2025* [online]. Available at: https://aiindex.stanford.edu/report/ (Accessed: 23 November 2025).

Stöger, K., Schneeberger, D. and Kieseberg, P. (2021) 'Legal aspects of data cleansing in medical AI', *Computer Law & Security Review* [online], 41. Available at: https://doi.org/10.1016/j.clsr.2021.105531 (Accessed: 30 October 2024).

Streel, A., Crémer, J., Heidhues, P., Fletcher, A., Kimmelman, G., Monti, G. and Morton, F. (2023) 'The effective use of economics in the eu digital markets act', *SSRN Electronic Journal* [online]. Available at: https://doi.org/10.2139/ssrn.4526050 (Accessed: 30 October 2024).

Streitz, N. (2023) 'Sector-specific approaches to AI regulation: A comparative analysis of healthcare, finance, and transportation', *AI & Society* [online], 38(3), pp. 789–803. Available at: https://doi.org/10.1007/s00146-022-01398-5 (Accessed: 30 October 2024).

Suprapmanto, J. (2024) 'Utilisation of Powtoon platform as learning media and improving student achievement', *World Psychology* [online], 3(1), pp. 113–127. Available at: https://doi.org/10.55849/wp.v3i1.608 (Accessed: 30 October 2024).

Susskind, D. (2024) *Growth: A History and a Reckoning*. London: Allen Lane.

Taddeo, M. and Floridi, L. (2021) ‘How AI can be a force for good’, *Science* [online], 372(6547), pp. 124–126. Available at: https://doi.org/10.1126/science.abg3838 (Accessed: 30 October 2024).

Tanti, T., Astalini, A., Kurniawan, D., Darmaji, D., Puspitasari, T. and Wardhana, I. (2021) ‘Attitude for physics: The condition of high school students’, *Jurnal Pendidikan Fisika Indonesia* [online], 17(2), pp. 126–132. Available at: https://doi.org/10.15294/jpfi.v17i2.18919 (Accessed: 30 October 2024).

Tashakkori, A. and Teddlie, C. (2010) *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. 2nd edn. Thousand Oaks, CA: Sage.

TechRadar (2025a) ‘I am an AI expert and here’s why synthetic threats demand synthetic resilience’ [online]. Available at: https://www.techradar.com/pro/i-am-an-ai-expert-and-heres-why-synthetic-threats (Accessed: 22 August 2025).

TechRadar (2025b) ‘Tackling Shadow AI: how UK businesses can mitigate the risks’ [online]. Available at: https://www.techradar.com/pro/tackling-shadow-ai (Accessed: 22 August 2025).

Teece, D.J. (2018) ‘Business models and dynamic capabilities’, *Long Range Planning*, 51(1), pp. 40–49.

Teixeira, G., Silva, M. and Pereira, R. (2019) ‘The critical success factors of gdpr implementation: a systematic literature review’, *Digital Policy Regulation and Governance* [online], 21(4), pp. 402-418. Available at: https://doi.org/10.1108/dprg-01-2019-0007 (Accessed: 30 October 2024).

The Guardian (2025) *AI tools used by English councils downplay women’s health issues, study finds* [online]. Available at: https://www.theguardian.com/technology/2025/aug/11/ai-tools-used-by-english-councils (Accessed: 22 August 2025).

TIME (2025) ‘Cardell, S. On the role of the CMA in AI regulation’ [online]. Available at: https://time.com/7012813/sarah-cardell (Accessed: 22 August 2025).

Toward AI Governance (2022) ‘Toward AI governance: identifying best practices and potential barriers’, AI and Ethics. Available at: https://pmc.ncbi.nlm.nih.gov/ (Accessed: 31 December 2025).

Treviño, L.K., den Nieuwenboer, N.A. and Kish-Gephart, J.J. (2014) ‘(Un)ethical behaviour in organisations’, *Annual Review of Psychology*, 65, pp. 635–660.

TTMS (2025) *AI Security Risks Uncovered: What You Must Know in 2025* [online]. Available at: https://ttms.com/ai-security-risks-explained (Accessed: 22 August 2025).

Tula, N. (2024) ‘Sustainability through ethical AI innovation’, *Technology and Management Review*, 7(2), pp. 95–110.

Turner, J. and Baker, R. (2017) ‘Pedagogy, leadership, and leadership development’, *Performance Improvement* [online], 56(9), pp. 5-11. Available at: https://doi.org/10.1002/pfi.21734 (Accessed: 30 October 2024).

Ullah, G., Ahmed, S. and Jamshed, K. (2017) ‘Do multinational companies practice good corporate governance? empirical evidence from bangladesh’, *International Journal of Accounting and Financial Reporting* [online], 7(2), 96. Available at: https://doi.org/10.5296/ijafr.v7i2.11843 (Accessed: 30 October 2024).

Ullrich, D., Cope, V. and Murray, M. (2020) ‘Common components of nurse manager development programmes: A literature review’, *Journal of Nursing Management* [online], 29(3), pp. 360-372. Available at: https://doi.org/10.1111/jonm.13183 (Accessed: 30 October 2024).

US House of Representatives (2020) *Final Committee Report: The Design, Development and Certification of the Boeing 737 MAX*. Washington, DC: US Government Publishing Office.

Vakkuri, V., Kemell, K. and Abrahamsson, P. (2019) ‘Implementing ethics in ai: initial results of an industrial multiple case study’, *Lecture Notes in Computer Science* [online], pp. 331-338. Available at: https://doi.org/10.1007/978-3-030-35333-9_24 (Accessed: 30 October 2024).

Vallor, S. (2016) *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Veale, M. and Zuiderveen Borgesius, F.Z. (2021) ‘Demystifying the Draft EU Artificial Intelligence Act’, *Computer Law Review International*, 22(4), pp. 97–112.

Verschuere, B. and Beddeleem, E. (2013) ‘Innovating public sector governance: Linking governance and performance’, *Public Management Review*, 15(6), pp. 803–824.

Vidal, M. E. and Russo, M. (2025) ‘Towards an Ontology-Driven Approach to Document Bias’, *Journal of Artificial Intelligence Research*.

Vincent, J. (2023) ‘Algorithmic bias lawsuit filed against State Farm’, *The Verge* [online], 12 March. Available at: https://www.theverge.com (Accessed: 14 September 2025).

Walsham, G. (1995) ‘Interpretive case studies in IS research: nature and method’, *European Journal of Information Systems* [online], 4(2), pp. 74–81. Available at: https://doi.org/10.1057/ejis.1995.9 (Accessed: 30 October 2024).

Washington Post (2025) *AI systems 'ignorant' of sensitive data can be safer, but still smart* [online]. Available at: https://www.washingtonpost.com/newsletter/politics/2025/08/12/ai-systems-ignorant-sensitive-data (Accessed: 22 August 2025).

Weaver, J. (2018) 'The United States and the self-regulation of artificial intelligence: Innovation or abdication?', *Technology and Society Review,* 23(1), pp. 44–59.

Wijewantha, P. (2019) 'Developing the corporate global leadership bench strength through transformational leaders', *Asian Journal of Empirical Research* [online], 8(12), pp. 453-467. Available at: https://doi.org/10.18488/journal.1007/2018.8.12/1007.12.453.467 (Accessed: 30 October 2024).

Wilson, H.J. and Daugherty, P.R. (2018) *Human + machine: Reimagining work in the age of AI*. Boston, MA: Harvard Business Review Press.

Wirtz, B.W., Weyerer, J.C. and Geyer, C. (2019) 'Artificial Intelligence and the Public Sector—Applications and Challenges', *International Journal of Public Administration,* 42(7), pp. 596–615.

Wong, A. (2021) 'Ethics and regulation of artificial intelligence' in *Artificial Intelligence for Knowledge Management*. Cham: Springer International Publishing, pp. 1-18.

World Economic Forum (2022) *Harnessing artificial intelligence for the Earth.* Geneva: World Economic Forum. Available at: https://www.weforum.org/reports/harnessing-artificial-intelligence-for-the-earth/ (Accessed: 25 November 2025).

Xia, H. (2026) 'Exploring the impact of responsible AI governance on organisational performance', Technological Forecasting and Social Change. Available at: https://www.sciencedirect.com/ (Accessed: 31 December 2025).

Xu, X., Zhang, M., Zhang, H. and Guo, S. (2016) 'Research on the governance structure and corporate performance of listed companies', *Proceedings of the 2016 International Conference on Management Science and Management Innovation* [online]. Available at: https://doi.org/10.2991/msmi-16.2016.67 (Accessed: 30 October 2024).

Yazdani, A. (2023) 'The impact of ai on trends, design, and consumer behavior', *Aitechbesosci* [online], 1(4), pp. 4-10. Available at: https://doi.org/10.61838/kman.aitech.1.4.2 (Accessed: 30 October 2024).

Yerra, S. (2023) 'The impact of AI-driven data cleansing on supply chain data accuracy', *ResearchGate*.

Yeung, K. (2022) ‘Algorithmic regulation: A critical interrogation’, *Regulation & Governance*, 16(3), pp. 685–703.

Zajko, M. (2025) ‘The ethics of AI business practices: a literature review on AI ethics guidelines and governance’, Journal of Management and Social Responsibility. Available at: https://jmsr-online.com/ (Accessed: 31 December 2025).

Zeno, R. (2013) ‘Emergent change and accountability in dialogic OD: A systems perspective’, *Organisational Change Review*, 19(3), pp. 212–228.

Zeydan, B., Azevedo, C.J., Makhani, N., Cohen, M., Tutuncu, M., Thouvenot, E. and Lebrun Frenay, C. (2024) ‘Early disease modifying treatments for presymptomatic multiple sclerosis’, *CNS Drugs* [online], 38(12), pp. 973-983. Available at: doi:10.1007/s40263-024-01117-9 (Accessed: 30 October 2024).

Zhou, Y. (2025) ‘A systematic literature review on AI governance platforms’, World Journal of Advanced Engineering and Technology Studies. Available at: https://journalwjaets.com/ (Accessed: 31 December 2025)

Zikmund, W.G., Babin, B.J., Carr, J.C. and Griffin, M. (2013) *Business Research Methods*. 9th edn. Mason, OH: Cengage Learning.

APPENDIX A SURVEY COVER LETTER 1

Survey 1

Ethical Attitudes and Practices

This survey contributes to a Doctor of Business Administration (DBA) research thesis at the Swiss School of Business and Management (Geneva), with the working title '**Navigating the Ethical Landscape of Artificial Intelligence in the Commercial Sector.**'

All responses are confidential, so please feel confident in answering the questions sincerely, truthfully and accurately.

Before you begin the survey, **please review the following definitions** to ensure a common understanding of key terms:

Ethics: The principles and standards that govern the conduct of individuals and organisations, based on concepts of right and wrong, fairness, justice, and respect for others.

Ethical Behaviour: The practice of making decisions and acting in ways that are consistent with ethical principles, involving honesty, integrity, fairness, and respect for others.

Thank you so much for taking the time to participate in this survey. Your responses will provide valuable insights into attitudes towards ethics and ethical behaviour as it relates to the design, deployment and usage of Artificial Intelligence (AI) within the business sector.

The following section captures some basic demographic information about you. All information is confidential and is not shared.

APPENDIX B SURVEY COVER LETTER 2

Survey 2

Perspectives on Responsible AI: Governance, Ethics, Regulation, and Leadership Development.

Thank you for participating in this research on responsible AI. The survey takes 6-8 minutes to complete and uses the same straightforward Likert scale for each question.

I am Glenn Chilcott, a Doctor of Business Administration (DBA) candidate at the Swiss School of Business and Management (Geneva).

This study explores leadership and organisational readiness, responsible AI adoption, and executive development in the age of AI.

Who should participate?

This survey is intended for executives, policymakers, business leaders, managers, consultants, and professionals involved in AI strategy, development, deployment or usage including:

- Executives, consultants and professionals engaged in AI strategies, development or implementation;
- Business leaders overseeing AI strategies, implementation and usage;
- Policymakers shaping AI regulations and frameworks;
- Governance, compliance, and ethics specialists influencing responsible AI deployment.

How will your responses be used?

By sharing your perspectives, you will help advance leadership-focused AI literacy and shape best practice frameworks for responsible and sustainable AI adoption in business.

Your privacy matters.

All responses are **confidential and anonymous**. Please answer each question openly and honestly.

Thank you again for your participation! Your insights will be invaluable.

Key Definitions:

Artificial Intelligence (AI): The development and application of computer systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, problem-solving, and decision-making.

Ethics: The principles and standards governing conduct based on concepts of right and wrong, fairness, justice, and respect.

Governance: The frameworks, policies, and processes through which organisations, institutions, or societies are directed and controlled to ensure accountability and ethical compliance.

Regulation: Rules, laws, and guidelines established by authorities to oversee AI, ensuring fairness, compliance, and protection of public interests.

Responsible AI in Business Leadership refers to the strategic integration and oversight of artificial intelligence technologies in a manner that aligns with ethical standards, organisational values, and societal expectations.

It involves leaders actively ensuring that AI systems are:

- fair

- transparent

- accountable

- human-centred,

with a clear focus on enhancing trust, driving inclusive innovation, and delivering long-term value to all stakeholders—including employees, customers, shareholders, and society at large.

Key dimensions include:

Ethical decision-making: Embedding ethical principles into AI-driven business strategies and operations.

Transparent governance: Establishing clear policies, roles, and audit mechanisms to monitor AI use and outcomes.

Workforce and talent impact: Supporting workforce transition, upskilling, and responsible automation practices.

Stakeholder trust: Building and maintaining public and stakeholder confidence through open communication and responsible data use.

Sustainable innovation: Leveraging AI to achieve business goals while promoting social good, equity, and environmental responsibility.

APPENDIX C INFORMED CONSENT

Informed Consent Form – Research Participant Information

Study Title: Leadership Capability, AI Literacy, and Responsible AI Practices in Organisations

Researcher: Glenn Chilcott MBA MSc, Doctor of Business Administration (DBA) Candidate, Swiss School of Business & Management, Geneva (SSBM)

Supervisor: Dr Denis Buterin, Swiss School of Business & Management

---

Purpose of the Study

You are invited to take part in a research study that examines leadership capability, AI literacy, governance, and ethical considerations relating to the use of artificial intelligence (AI) in organisational contexts. This research forms part of a Doctor of Business Administration (DBA) programme.

Participation

Your participation in this study is entirely voluntary. You may decline to answer any question and may withdraw from the survey at any point before submission without providing a reason.

Eligibility

You are eligible to participate if you are currently in a leadership, management, or decision-making role, or have professional experience relevant to AI adoption, governance, or organisational leadership.

Data Collection and Use

The survey collects quantitative data through structured questions. No personally identifiable information is required unless explicitly stated. All responses will be analysed in aggregate form for academic research purposes only.

The data collected will be used to support doctoral research and may contribute to academic publications or conference presentations. Individual participants will not be identifiable in any outputs.

Confidentiality and Data Protection

All data will be handled in accordance with the UK General Data Protection Regulation (UK GDPR). Responses will be stored securely and accessed only by the researcher and, where necessary, academic supervisors. Data will be retained for a limited period in line with institutional research data retention policies and then securely deleted.

Risks and Benefits

There are no anticipated risks associated with participation beyond those encountered in everyday professional reflection. While there is no direct personal benefit, your participation will contribute to research on responsible AI leadership and organisational practice.

Right to Withdraw

You may withdraw from the study at any point before submitting the survey. Once the survey has been submitted, responses cannot be withdrawn as data are anonymised and cannot be linked to individual participants.

Contact Information

If you have any questions about the study or your participation, please contact: Glenn Chilcott: glenn@ssbm.ch

For concerns regarding research ethics, you may contact: Swiss School of Business & Management, Geneva

---

Consent Statement

Please confirm your consent by selecting **"I agree"** below.

- I have read and understood the information provided above.

- I understand that my participation is voluntary and that I may withdraw at any time prior to submission.
- I understand how my data will be used and stored.
- I consent to participate in this research study.

☐ **I agree to participate in this study**
☐ I do not agree to participate