Doctoral theses at NTNU, 2024:184

Doctoral thesis

Emmanouil Papagiannidis

Responsible AI Governance in practice

The strategic impact of Responsible AI Governance on business value and competitiveness

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of Science and Technology

Emmanouil Papagiannidis

Responsible AI Governance in practice

The strategic impact of Responsible AI Governance on business value and competitiveness

Thesis for the Degree of Philosophiae Doctor

Trondheim, May 2024

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-7960-7 (printed ver.)
ISBN 978-82-326-7959-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2024:184

Printed by NTNU Grafisk senter

Dedicated to my beloved parents

“Δόξα τω Θεώ πάντων ένεκεν.”

- ιερός Χρυσόστομος

“We are what we repeatedly do. Excellence, then, is not an act, but a habit.”

- Aristotle

“Have I played the part well? Then applaud as I exit.”

- Augustus

Abstract

Artificial Intelligence (AI) comprises a set of technologies with vast potential applications, which can range from autonomous vehicles and chatbots to fraud detection and medical diagnosis. While the domain of AI research has a history of over 50 years, recent technological advances have facilitated their utilization and deployment in real-world applications. Despite growing rates of AI deployment, many organizations struggle to fully realize business value from such technologies. Additionally, although AI offers numerous advantages, it is not exempt from potential negative or unintended consequences. Rising concerns regarding AI usage and instances of failed AI applications, some of which resulted in fatalities, job displacement, or racial biases, have underscored the urgent need for responsible AI governance (RAIG). Therefore, a gap exists between the design, deployment, and use of AI and its business value.

The thesis employs a sequential multiple methods research design, commencing with an exploratory approach to uncovering key aspects of RAIG and responsible principles. Initially, a comprehensive literature review was conducted to acquire a holistic understanding of AI use and its business value. A second literature review followed, exploring RAIG, aimed at grasping the principles and methods for ensuring responsible AI utilization. Following the two literature reviews, we conducted a series of empirical studies. We start with a multi-case study with three organizations to examine how responsible AI governance is implemented in practice and promotes the development of robust AI applications that do not introduce negative effects. Next, an in-depth case study with 14 expert interviews was conducted to explore the importance of RAIG and the dark side effects that might occur if RAIG is not present. After that, research was carried out to construct a conceptual framework, which forms the main processes involved in AI resource orchestration. This framework aims to explain the different activities used to orchestrate resources strategically, thereby generating business value. Finally, a quantitative study with 329 responses from Europe and the USA was undertaken to investigate whether RAIG yields tangible value and, if so, through which mechanisms and processes this value is realized.

The results contribute to our understanding of how RAIG is implemented in organizations, and what its resulting business value is: firstly, through a conceptual model by exploring the fundamental dimensions relevant to RAIG within organizations and unveiling the underlying practices supporting them; secondly, by identifying the negative or unintended consequences of AI in the absence of RAIG, categorized into three clusters related to the nature of work, conflicts and effects, and responsibility; and thirdly, through a conceptual model by presenting and elucidating how firms manage their RAIG practices to improve competitiveness. Finally, the research discusses implications for research, practice, and policy, while also highlighting avenues for future investigation.

Preface

This doctoral thesis was presented to the Norwegian University of Science and Technology (NTNU) in partial fulfillment of the requirements for the degree of philosophiae doctor. Professor Patrick Mikalef (NTNU) served as the primary supervisor, with Professor John Krogstie (NTNU) and Professor Kieran Conboy (University of Galway) acting as co-supervisors.

The research was financially supported by NTNU, and the majority of the doctoral work was conducted at the Department of Computer Science, NTNU, located in Trondheim, Norway. The research spanned from 2020 to 2024.

Acknowledgements

It has been four long years…we survived through COVID-19; almost two years without seeing people, and yet here we are.

First, I would like to start by expressing my deepest gratitude to my family for their support, encouragement, and understanding throughout this journey. Their love and belief in me have been my source of strength and inspiration. At this point, the reader will excuse me because I will write in my mother tongue language (Greek) so my parents can understand this part.

Ευχαριστώ τους αγαπημένους μου γονείς και τα αδέρφια μου, που με την αγάπη και την υποστήριξή τους με βοήθησαν σε αυτό το ταξίδι της διδακτορικής μου διατριβής. Είμαι βαθύτατα ευγνώμων για την ανέκφραστη στήριξή τους καθώς έφθασα στο τέλος αυτού του σημαντικού αλλα ταυτόχρονα περίπλοκου ταξιδιού που αν και τελίωσε σηματοδοτεί ένα καινούργιο με άγνωστο προορισμό. Οι συμβουλές τους, οι νουθεσίες τους και η απέραντη αγάπη τους με έκαναν πάντα να προσπαθώ και να επιδιώκω το καλύτερο, και για αυτόν τον λόγο τους οφείλω έναν μεγάλο ευχαριστώ και θα είμαι πάντα ευγνώμων. Η παρουσία τους στη ζωή μου είναι ο πραγματικό μου πλούτος και η συνεχής πηγή έμπνευσής μου. Επίσης θα ήθέλα να εκφράσω την ειλικρινή μου ευγνωμοσύνη στον πατέρα Βαρνάβα, τον πατέρα Ιωάννη και τον πατέρα Αλέξανδρο για τις προσευχές τους για μένα την υπομονή τους και την υποστήριξή τους. Οι σοφές τους συμβουλές και η ενθάρρυνσή τους με βοήθησαν να παραμείνω δυνατός κατά τις δύσκολες στιγμές. Είμαι πραγματικά ευγνώμων για την πνευματική τους καθοδήγησή και τον τρόπο που με βοήθησαν να γίνω ένα καλύτερος άνθρωπος.

I am profoundly thankful to my three supervisors, Patrick Mikalef, John Krogstie, Kieran Conboy for their invaluable guidance, mentorship, and scholarly insights. Especially, I would like to thank Patrick who was the main supervisor as his guidance and experience helped me tremendously through this voyage. All three supervisors' expertise, dedication, and constructive feedback have been instrumental in shaping this research and guiding me through the challenges of doctoral studies. I am grateful to my Department and NTNU for providing a stimulating academic environment and the appropriate resources for conducting my research. I would also like to extend my appreciation to my colleagues for their discussions, and support and many hilarious moments especially during lunch.

At this point, I would like to express my thanks to all my friends in Trondheim with whom I have shared a lot of memorable moments together, especially Athanasios who was always kind, cheerful, and willing to assist you in any moment.

Special thanks to my dear friend Markos for his unwavering friendship, positive encouragement, and occasional distractions that helped me going through these four years. Our late-night conversations were most appreciated!

Last but not least, to myself, I extend heartfelt gratitude for the determination, perseverance, and resilience that I have shown through this doctoral journey. As Snoop Dogg once said, "*I wanna thank me for believing in me*".

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence

EU European Union

IS Information Systems

KMC Knowledge Management Capabilities

LT Legitimacy Theory

ML Machine Learning

PLS Partial Least Squares

RAIG Responsible Artificial Intelligence Governance

RAI Responsible Artificial Intelligence

ROT Resource Orchestration Theory

SEM Structural Equation Modelling

Part I

1 Introduction

1.1 Problem Statement

Artificial Intelligence (AI) has emerged as a set of technologies with far-reaching implications across various industries and domains. Its applications have showcased remarkable capabilities, ranging from task automation and cost reduction to personalized recommendations and virtual assistants (Alsheibani et al., 2020; Frank et al., 2019; Gregory et al., 2020). Notably, the global AI market is now worth billions of dollars, with projections estimated to exceed one trillion US dollars by 2030 (Plastino & Purdy, 2018; Ramadoss et al., 2018). However, to fully realize the benefits of AI investments, organizations should also invest time, effort, and resources in order to adopt and diffuse AI (Enholm et al., 2021). Organizations recognize the importance of building trust and reducing the risks and unintended consequences associated with AI (Eitel-Porter, 2021), yet only 11% of risk leaders feel adequately prepared to measure the risk of the effects of AI (Accenture, 2019). Consequently, there is a clear and growing tension between technological capabilities and the human and social systems within which these technologies operate. Because of this tension, research into the extent to which technology enhances productivity, capacity, and well-being is required. In addition, potential side effects need to be considered, including adverse effects on human health, emotional states, and overall exhaustion (Conboy, 2019; Tarafdar et al., 2015). Recent studies have indicated that AI may lead to a number of unintended or negative effects, such as technology addiction and technostress (Turel & Ferguson, 2020), security and privacy concerns (Aqeel et al., 2022), and the spread of fake news by using deepfake technologies (Al-Asadi & Tasdemir, 2022). Of particular interest is the attention AI has gained concerning its dark side implications, encompassing ethical and societal considerations (Mikalef et al., 2022).

Highlighting the complexity and potential risks associated with the advancement of AI, recent studies (Mikalef et al., 2022; Papagiannidis et al., 2023; Sharma et al., 2022; Sun et al., 2022) have underscored the critical need for comprehensive measures to mitigate these challenges, leading to the emergence of AI governance, defined as *"a variety of tools, solutions, and levers that influence AI development and applications"* (Butcher & Beridze, 2019; Mäntymäki et al., 2022). We selected this definition as it aligns closely with our perspective on AI governance, particularly in terms of governing AI in a manner that prioritizes processes and tools for successfully building AI applications. Furthermore, the existing landscape of AI governance presents a number of critical challenges that require comprehensive and integrated solutions (Ghallab, 2019). However, one of the primary challenges is the lack of clear guidelines and regulations regarding AI governance (Hagendorff, 2020), and because AI technologies rapidly evolve, organizations face challenges in navigating the regulatory environment and determining appropriate mechanisms to effectively govern AI systems

(Jelinek et al., 2021). The lack of consensus about what AI governance could entail has resulted in fragmented regulations and impeded innovation since AI governance plays a huge role in the effective utilization of AI technologies within organizational settings (Jelinek et al., 2021).

What is more, as businesses increasingly integrate AI into their operations, it has become extremely important to establish clear guidelines and frameworks to govern the development, deployment, and usage of these technologies (Dwivedi, 2021). Effective responsible AI governance (RAIG) involves creating policies, procedures, and structures that guide the responsible and ethical use of AI within an organization (Gianni et al., 2022). Robust governance practices help organizations realize the benefits of AI while managing potential risks and ensuring compliance with relevant regulations. Hence, complete RAIG frameworks are essential for achieving optimal outcomes, mitigating risks, and ensuring ethical and responsible AI practices. This includes addressing issues related to data privacy, algorithmic transparency, accountability, and overall ethical considerations. For example, transparency and accountability are important aspects of RAIG because the uncertainty of AI outputs raises concerns about how decisions are made when AI is involved in decision-making processes due to the potential for discriminatory outcomes (Akter et al., 2021). Thus, understanding and explaining AI system decisions are vital to ensure transparency, accountability, and the ability to address potential errors (Haibe-Kains et al., 2020). However, striking the right balance between transparency and the protection of proprietary algorithms or sensitive information is a challenge that needs to be overcome. Another critical challenge in RAIG is ensuring fairness and mitigating biases in AI systems (Bellamy et al., 2019). AI algorithms and datasets can preserve existing social biases and lead to discriminatory outcomes, thereby worsening social disparities. That means developing mechanisms to detect and address biases, promoting diversity in training data, and establishing frameworks for auditing and certifying AI systems for fairness are crucial steps towards RAIG (Arrieta et al., 2020).

Other considerations of RAIG are privacy and data protection. AI systems often rely on vast amounts of personal data, raising questions about individuals' privacy rights and the potential for misuse or unauthorized access. Establishing comprehensive data protection frameworks, consent mechanisms, and privacy-preserving AI techniques are essential to safeguarding individuals' privacy while harnessing the benefits of AI technologies (Papagiannidis et al., 2021). The rapid evolution of AI technologies also raises concerns about accountability and liability. However, traditional notions of accountability may not be applicable when AI systems make autonomous decisions or have a significant impact on individuals or society. Additionally, ensuring cross-domain collaboration and interdisciplinary approaches to RAIG is crucial. AI technologies have implications in various sectors, including healthcare, finance, transportation, and more (Kuziemski & Misuraca, 2020).

As a result of these challenges, the concept of RAIG has emerged, where robust and comprehensive frameworks are needed. Ideally, these frameworks should incorporate AI governance with responsible AI principles. It is vital that frameworks are adaptable to the evolving nature of AI technologies and promote global cooperation and standardization. In addition, guidelines and best practices should be developed for assessing, certifying, and auditing AI systems in order to ensure compliance with ethical requirements. By investigating the current challenges and proposing practical solutions for RAIG, this study aims to contribute to the advancement of AI governance. Using existing governance frameworks, ethical considerations, and emerging best practices, this study aims to provide organizations, decision makers, managers and other stakeholders with insights and recommendations. These insights revolve around optimizing AI integration, understanding how RAIG aligns with strategic objectives to enhance competitive performance, and identifying practices that yield actionable insights for decision-making while addressing potential barriers and unintended AI consequences. Moreover, the insights aim to uncover the drivers behind RAIG practices—structural, relational, and procedural—and their impact on both businesses and society. Lastly, this study aims to establish the correlation between RAIG and firm performance. The ultimate goal is to facilitate an environment in which AI technologies can be exploited to their full potential while minimizing potential risks and ensuring societal benefit.

1.2 Research Motivation

Recognizing the challenges that organizations encounter while pursuing competitive advantage to increase business value, our research is primarily driven by the aim to provide knowledge of how companies can effectively manage their AI systems in a responsible way. We also seek to provide guidelines and evidence highlighting how RAIG contributes to business value. Numerous sources have emphasized the significance of AI governance and RAI practices, merging them to gain substantial business benefits. For instance, the European Commission (2019), the Singapore Government (2020), and Google (2019) have published extended work on responsible AI principles by providing guidelines for organizations. Additionally, scholars have investigated topics encompassing ethical guidelines regarding AI policies (Gianni et al., 2022), failed cases such as Tay (Mark Van Rijmenam & Schweitzer, 2018), and the connection of AI governance and digital responsibility (Thelisson et al., 2019).

Nevertheless, there is a gap between theory and practical applications, necessitating research to provide concrete evidence on how RAIG enhances performance and increases business value. Given the fact that RAIG is relatively new, there is limited empirical work grounded on established IS theories. The frameworks mentioned above do not provide any support for systematically incorporating, building, and maintaining RAIG practices and procedures. In our research, we utilize Legitimacy Theory (LT) to investigate how organizations aim to align their AI practices with societal expectations (norms and ethics), regulatory requirements, and stakeholder demands, thus bolstering their legitimacy and credibility when they develop AI products. However, what is considered ethical AI use lacks robust mechanisms to enforce normative claims (Hagendorff, 2020). While there may be potential ethical consequences, such as reputational losses or restrictions on professional memberships, these mechanisms are generally weak and do not always pose an immediate threat (Hagendorff, 2020). Hence, we use LT to check the link between RAIG, corporate reputation, and performance.

The responsible part of RAIG is an attractive concept for many AI companies and institutions. Companies and research institutes often formulate their own responsible AI guidelines, frequently include ethical considerations in their public relations efforts, or adopt responsible motivated "self-commitments." RAI focuses more on the practical implementation and operational aspects of ensuring that AI systems conform to established guidelines and principles, while ethics goes into the broader moral considerations regarding the development, deployment, and impact of AI on society, individuals, and the environment. These serve to discount the need for a binding legal framework and suggest to policymakers that internal self-governance within the scientific and industrial sectors is sufficient to mitigate potential technological risks and prevent abuse (Hickok & Maslej, 2023). Even when more concrete regulations concerning AI systems are demanded, as seen in recent calls by Google (2019), these demands tend to remain vague and superficial. Bourgon (2007) argues that the promotion of academia-led or industry-led ethics guidelines and other forms of self-governance can create an illusion that accountability can be transferred from state authorities and democratic institutions to the respective sectors of academia or industry. Moreover, responsibility can be used as a means to pacify public voices while maintaining low criticism within the organization (Zhang & Yang, 2021). A good example in this context is the "Partnership on AI" association (2018), which brings together companies such as Amazon, Apple, Baidu, Facebook, Google, IBM, and Intel. Companies often emphasize their membership of such associations to convey a sense of commitment to regulating business activities while simultaneously denying the need for more strict legal regulations.

RAIG should not be limited to the scope of governance alone, but it should necessitate the implementation of mechanisms that can overcome various challenges, including the alignment between business users and business owners (Adam Cutler, 2020; Arrieta et al., 2020; Fadler & Legner, 2021). The governance of AI projects can be interpreted differently depending on individual perspectives, and it is crucial to consider the implications of AI applications (de Laat, 2021; Mökander & Floridi, 2023). The impact of AI on key organizational processes, such as delegation, coordination, and decision-making, is determined by the extent to which institutional frameworks enable them to assume managerial roles (Papagiannidis et al., 2022). In contrast, researchers at Microsoft (Amershi et al., 2019, May) approach RAIG from a technical standpoint, while the European Commission (EC) (Smuha, 2019) and the Singapore Government (2020) take a human and ethics-centric approach. The Microsoft researchers emphasize the technical aspects of AI governance and share insights on best practices employed by Microsoft teams to establish a coherent workflow incorporating software engineering processes. They also shed light on several crucial engineering challenges that organizations may encounter when developing large-scale AI solutions for the marketplace. Their findings highlight key aspects of AI governance that deal with the complexity of Machine Learning (ML) applications, which surpasses that of typical software applications. They identify the diverse skills required for constructing and customizing models depending on the project, and they recognize the potential challenges in managing AI components when distinct modules and models exhibit nonmonotonic error behaviour. The EC and the Singapore government view AI governance as a means of fostering trustworthy AI through the establishment of guidelines. These guidelines were used for creating policies that go beyond mere ethical principles to provide principles which promote ethical AI (Smuha, 2019), while maintaining or improving the corporate reputation of the organization (Dignam, 2020; Sharma, 2022).

Nonetheless, a significant disparity exists between the utilization of AI and the implementation of responsible AI. While some initiatives have been undertaken, as discussed earlier, it is evident that these efforts are insufficient. The low adoption rate of these guidelines by organizations may be attributed to either the overly ambiguous nature of the guidelines or their excessive specificity, making them impractical for many firms to follow. This creates a paradox where regulations and guidelines have to strike a balance between being abstract enough to accommodate future AI technologies and strict enough to ensure compliance, albeit within reasonable bounds. Consequently, a research gap has emerged in RAIG that necessitates attention. Therefore, it has become vital to identify effective practices for RAIG, to explore any undesired effects associated with them, and to investigate the relationship between RAIG and organizational performance—since the ultimate objective of technology adoption is to enhance the overall organizational effectiveness in a manner that aligns with the societal norms and ethics of society.

1.3 Research Questions

In response to the existing gaps and research needs identified in the literature, this thesis aims to address the following main research question (MRQ):

***MRQ**: What are the key factors for leveraging AI value and achieving competitive advantage through responsible AI governance?*

To provide a systematic framework for investigation, this thesis falls within the Information Systems (IS) field and focuses on exploring the contribution of RAIG to competitive advantage. The main research question is further divided into four sub-questions, which aim to examine the existing gap between AI utilization and RAIG practices:

> ***RQ1**: How do organizational factors influence the adoption and implementation of AI technologies?*
>
> ***RQ2**: What are the key drivers and mechanisms for generating value from AI in organizations?*
>
> ***RQ3**: What are the key antecedents and effects for generating value from RAIG in organizations?*

RQ4*: How does RAIG impact organizational outcomes, including performance gains?*

1.4 Research Outcomes

A collection of articles was peer-reviewed; six primary research papers and three secondary research papers were authored and published in conferences and journals. The research papers have made contributions to the advancement of knowledge in the fields of IS and RAIG, thereby enhancing the development of a comprehensive body of knowledge.

1.4.1 Research Papers

The main research papers presented in this study address the research questions at hand. Table 1 provides a map that illustrates their connection to the respective research questions.

1. **MRP1:** Enholm, Ida Merete; Papagiannidis, Emmanouil; Mikalef, Patrick; Krogstie, John. (2021) **Artificial intelligence and business value: a literature review.** Information Systems Frontiers. *-Published).*
   My contribution: I was involved in conducting a thorough analysis of existing literature, identifying pertinent research studies, and synthesizing the main findings. I played a key role in carrying out systematic searches to locate relevant academic articles, books, and other valuable sources of information, and conceptualizing a framework for AI in business. Additionally, I actively participated in critically assessing the quality and significance of the literature, pointing out any gaps that could be addressed in further research, and paper writing.
   Relevance to the thesis: This paper serves to provide an overview of the research context and establish the current state of AI use in the business domain and the value-generating mechanisms. By conducting an extensive review of the existing literature, the paper identifies gaps, limitations, and unresolved issues in the current knowledge landscape. In terms of research questions, it contributes to RQ1 and RQ2 by constructing a research framework that examines the drivers, enablers, and inhibitors of AI adoption in business settings. Notably, during the identification of research gaps, it became apparent that the responsible aspect of AI use, specifically RAIG, plays a significant role in the development, implementation, and use of AI products and services. This realization prompted me to focus my research on the responsible practices associated with AI utilization.

2. **MRP2:** Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran. (2023) **Responsible AI governance: a systematic literature review.** The Journal of Strategic Information Systems. - *In second round of review.*
   My contribution: I was the leading author and, in a similar manner to the first literature review, my involvement encompassed conducting a comprehensive analysis of the existing literature, identifying relevant research studies, and synthesizing the primary findings. I played a key role in systematically searching for relevant academic articles, publications and other valuable information. Together with the team, we conceptualized a framework for AI in the business context, and I actively contributed to critically evaluating the quality and relevance of the literature, highlighting any gaps that could be addressed in future research, and collaborating in the process of writing the paper. **Relevance to the thesis:** This paper serves as a foundation for RAIG, aiming to enhance the understanding of responsible practices and the essential attributes and principles that AI applications and services should possess to be deemed trustworthy and contribute to business value within an organization. The paper addresses RQ3 and RQ4 by offering definitions and exploring various themes related to responsible AI, emphasizing the significance of responsible AI principles and discussing the identification of factors that influence RAIG practices, including structural, relational, and procedural aspects. Furthermore, it delves into the effects of responsible AI (RAI) on businesses and society, highlighting their implications.

3. **MRP3:** Papagiannidis, Emmanouil; Enholm, Ida Merete; Dremel, Christian; Mikalef, Patrick; Krogstie, John. (2022) **Toward AI governance: identifying best practices and potential barriers and outcomes.** Information Systems Frontiers. – *Published.*
   My contribution: I was the leading author, and I developed interview protocols to ensure that the interviews would cover the necessary information, and I identified relevant research questions that would guide our investigation. During the data collection phase, I conducted the interviews, carefully recording participants' responses. Once the data was gathered, I undertook the analysis process, searching for recurring themes and patterns within the information by comparing participants' responses, and I contributed to the writing process.
   Relevance to the thesis: This paper follows a positive incline approach towards AI and introduces a model that examines the application of AI governance in fostering the development of reliable AI applications without adverse effects. This is achieved by conducting a comparative case analysis, and 15 semi-structured expert interviews, involving three different firms. The paper addresses RQ2 and RQ3. Firstly, it illustrates the practices implemented by these firms to generate valuable knowledge that aids in decision-making. These practices help overcome barriers, and recommend actions, that lead to desired outcomes. Secondly, it goes

through the key barriers that are pertinent to AI governance within organizations and sheds light on the underlying practices that support these dimensions.

4. **MRP4:** Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran; Rogier Van de Wetering. (2023) **Uncovering the dark side of AI-based decision-making: A case study in a B2B context.** Industrial Marketing Management. – *Published.*
My contribution: I was the leading author and in a similar manner to the previous paper, I played an active role in the research project by developing interview protocols to ensure that we captured the necessary information during the interviews, identifying relevant research questions, collecting data, conducting the interviews, and ensuring that participants' responses were accurately recorded. Once all the data was collected, I analyzed it, searching for themes and patterns that emerged from the information provided by the participants. By comparing participants' responses, I was able to interpret the data in a meaningful and insightful way. In the subsequent stages, I played a vital role in synthesizing the research findings, integrating them to present a comprehensive and cohesive picture of our study's outcomes, and actively contributed to the writing process.
Relevance to the thesis: This paper follows a negative incline approach towards AI and examines its potential negative implications, in contrast to the previous paper. Specifically, through 14 semi-structured expert interviews, the paper explores the impact of AI trading bots on the relationship between traders and AI developers, as well as how organizations adapt to this new reality. By addressing RQ1, the paper identifies three clusters of negative or unintended consequences associated with AI use. These clusters refer to the nature of the work, conflicts and effects arising from AI implementation, and issues of responsibility.

5. **MRP5:** Papagiannidis, Emmanouil; Mikalef, Patrick; Krogstie, John; Conboy, Kieran. (2022) **From responsible AI governance to competitive performance: the mediating role of knowledge management capabilities.** Lecture Notes in Computer Science (LNCS). – *Published.*
My contribution: I was the leading author, and I played a crucial role in designing the survey questionnaire, ensuring that it effectively captured the relevant variables and constructs related to our research. I contributed to determining the appropriate sampling technique and sample size for our study by considering the characteristics of the target population and the research objectives and applying appropriate statistical techniques to analyze the survey data we collected. This involved conducting hypothesis testing and exploring relationships between variables using the selected statistical methods, playing a significant role in interpreting the end results in a meaningful and insightful manner, and actively contributing to the writing process of the paper.
Relevance to the thesis: This paper introduces a conceptual model that explores the relationship between RAIG, knowledge management capabilities (KMC), strategic alignment, and competitive performance and it is empirically validated

through a survey from 144 Nordic firms using partial least squares structural equation modelling (PLS-SEM). The model aims to provide valuable insights for companies planning to incorporate AI into their overall strategy. By examining the research questions related to the role of RAIG and its impact on competitive performance, this paper contributes to the existing knowledge in the field. The conceptual model presented in this paper establishes a foundation for understanding how KMC, when amplified through strategic alignment, can influence the adoption and implementation of RAIG practices within an organization. The model further highlights the potential benefits of effective RAIG in enhancing a company's competitive performance by addressing RQ4. This research situates itself within the field of RAIG and provides a framework for conceptualizing the interplay between RAIG and competitive performance. The insights gained from this study could be instrumental in guiding organizations as they navigate through the challenges and opportunities associated with AI adoption and AI governance.

6. **MRP6:** Papagiannidis, Emmanouil; Mikalef, Patrick. (2024) **Exploring the link between responsible AI governance, legitimacy, and firm performance.** - *Completed*
   My contribution: I was the leading author and in a similar manner to the previous paper, I was responsible for designing the survey questionnaire, ensuring that it encompassed the necessary variables and constructs based on our research objectives. This involved careful consideration of the factors that needed to be considered to address our research questions effectively, and I helped establish a reliable and representative sample that would yield meaningful insights. Once the survey data was collected, I applied suitable statistical techniques to analyze the data. This involved conducting hypothesis testing and exploring relationships between variables using the selected statistical methods, and I actively contributed to the writing process of the paper.
   Relevance to the thesis: This paper can be seen as a continuation of the previous paper, using a different angle. This paper empirically validates a conceptual model through a survey of 329 Scandinavian firms using PLS-SEM and applying LT. The paper aims to help companies realize the connection between responsible AI use and competitive advantage over their competition. The paper contributes to RQ4 by presenting clear evidence and showing how RAIG practices enhance corporate reputation, especially when businesses communicate their responsible AI use, both externally to the public and internally within the organization, and by establishing a solid foundation for understanding the relationship between RAIG and firm performance by focusing on the legitimacy practices associated with RAIG.

Table 1: Mapping of main research papers and research questions.

	P1	P2	P3	P4	P5	P6
RQ1	•			•		
RQ2	•		•			
RQ3		•	•			
RQ4		•			•	•

Additionally, three conference papers were produced of a secondary nature:

1. **SRP1:** Papagiannidis, Emmanouil; Enholm, Ida Merete; Mikalef, Patrick; Krogstie, John. (June 2021) **Structuring AI resources to build an AI capability: a conceptual framework**. Proceedings of the European Conference on Information Systems (ECIS) 2021. – *Published.*
2. **SRP2:** Papagiannidis, Emmanouil; Enholm, Ida Merete; Dremel, Christian; Mikalef, Patrick; Krogstie, John. (June 2021) **Deploying AI governance practices: a revelatory case study**. Proceeding of 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society (I3E 2021). – *Published.*
3. **SRP3:** Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran; Rogier van de Wetering. (September 2022) **The dark side of AI-based decision-making: a study of B2B trading.** Proceedings of the Conference: 21st IFIP Conference I3E2022 e-Business, e-Services, and e-Society. – *Published.*

All the secondary papers provided complementary perspectives to this thesis. First, SRP1 develops a conceptual framework and draws upon the principles of ROT. This framework emphasizes the difference between the ideation and implementation of AI capabilities, and it highlights activities related to ROT within the context of AI deployments. The paper puts forth a set of propositions that clarifies the key processes involved in resource orchestration for AI. This paper's relevance to the thesis lies in its goal to assist companies that have made substantial investments in AI by providing insights into the process through which AI can deliver business value. Second, SRP2, which has been extended and incorporated into MRP3, adopts a single case study approach and explores the implementation of AI governance. The primary objective is to facilitate the development of robust AI applications that do not introduce negative impacts on companies. By examining how AI governance is effectively put into practice, this paper sheds light on the main dimensions relevant to AI governance within organizations. Furthermore, it uncovers the underlying practices that contribute to successful AI governance. The paper's significance to the thesis lies in its ability to contribute to the understanding of AI governance and its implications. It provides valuable insights for organizations seeking to navigate through the complexities of AI adoption while ensuring responsible and beneficial outcomes. Third, SRP3, which has been further developed into MRP4, provides an exploration of the growing concerns regarding the negative and unintended consequences associated with AI technologies. The paper adopts a single case study approach based on eight semi-structured expert interviews, focusing on a Norwegian energy trading firm to delve into the dark aspects of AI. This paper's relevance to the thesis lies in its examination of the essential characteristics of AI trading within business-to-business (B2B) organizations. It sheds light on the potential negative implications of AI trading and proposes strategies to mitigate these negative effects by offering insights into the challenges and risks associated with AI trading.

For all secondary papers, I was the main author. My contributions encompassed various aspects, including generating ideas, conducting data collection, and playing a significant role in the core components of the reports. This involvement ranged from conceptualization and framework design to actively participating in the writing process and providing feedback on the draft versions. However, it is important to note that these secondary papers offer indirect contributions to the specific research questions addressed in this thesis. As a result, they have not been included in the main narrative of this thesis.

All co-authors played a significant role in the development of the above papers, contributing their valuable expertise and insights. Throughout the process, we engaged in fruitful discussions, exchanged ideas, and provided constructive feedback to refine the research. This collaborative approach helped ensure the accuracy and reliability of the analysis results. In terms of data collection, each co-author made valuable contributions, actively participating in the process of gathering the necessary information. This collaborative effort ensured high-quality data collection. What is more, during the peer-review process, all co-authors played an important role in critically reviewing the papers. Their input and feedback helped refine the content, strengthen the arguments, and ensure the accuracy and clarity of the research presented.

1.4.2 Research Contributions

The field of IS research has a history of incorporating theories from various disciplines like economics, computer science, psychology, and general management (Wade et al., 2004). Despite this thesis being focused on the IS field, it introduces a theory from political economy. Specifically, it draws from LT and provides novel insights into how RAIG improves the firm’s overall performance. Because of its emphasis on stakeholder interactions and social expectations, legitimacy theory offers a lens for understanding ethical AI governance. When dealing with AI, trust, ethical issues, and public perception are critical; thus, legitimacy theory provides a framework for navigating the challenges that arise. It emphasizes the need for AI systems to conform to cultural standards, comply with legislation, and earn stakeholder confidence through transparent, ethical actions. Using this idea, it can be seen how RAIG aspires not just to compliance but also to long-term societal acceptability and beneficial effects, ensuring long-term alignment with stakeholder expectations.

This thesis establishes an important and unexplored connection between RAIG and performance. For firms, it offers valuable insights, targeting decision makers, such as managers and board members, and helping them understand the significance of RAIG practices. By doing so, they can maintain corporate reputation, prevent fatal mistakes that may impact AI users and the environment, achieve remarkable firm performance, and gain a competitive advantage over the competition. For scholars, this concept can be seen as an opportunity to explore the link between RAIG and performance, validating it through empirical research. It also encourages further investigation and a deeper understanding of this relationship under different notions. Policymakers can benefit from this research by gaining practical insights into deploying RAI and aligning their policy and regulatory efforts. The research questions which contribute to the field can be shown in Table 2.

Table 2: Mapping of contributions and research questions.

	C1	C2	C3	C4
RQ1	•	•		
RQ2	•	•		
RQ3			•	•
RQ4			•	•

The contributions of this thesis are as follows:

> ***C1:*** *Improve understanding of AI in business, its overall business value, and how to gain its competitive advantage.* This includes findings on the potential value that AI brings to businesses, on how AI can be deployed to give a competitive edge in the market and future avenues of research.
>
> ***C2:*** *Identify enablers, inhibitors, and antecedents of AI.* This includes identifying key considerations, challenges and opportunities associated with AI implementation to support decision-making regarding AI adoption and utilization.
>
> ***C3:*** *New knowledge on RAIG and which RAIG practices and principles are considered essential.* This includes the dimensions of RAIG, theoretical and practical implications, the practices that are crucial to achieving RAIG and guidelines for governing AI systems ethically.

C4: *Explore the relationship between RAIG and firm performance.* This includes models that can explain the link between RAIG and firm performance and the benefits of adopting RAIG frameworks. The empirical validation contributes to the credibility and validity of the findings, which helps managers recognize the value of RAIG.

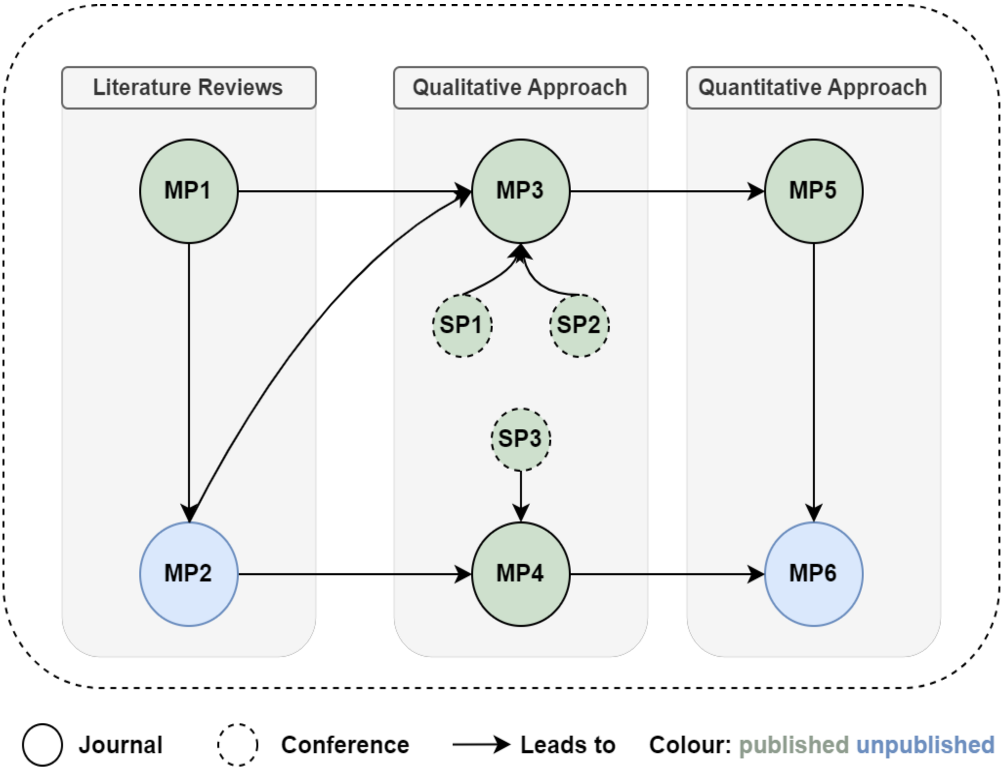

Figure 1: A schema of the research papers, their connection and scientific approach.

1.5 Thesis Structure and Outline

The thesis is structured into two main parts. Part I acts as the foundational pillar and encompasses the introduction to the research work. It offers an overview of the groundwork by delving into the background theories, methodological approaches, obtained results, and the contributions made by the thesis. Within Part I, each subsequent section unveils a facet of the research journey. After the introductory chapter 1, chapter 2 navigates through the landscape of AI and RAIG, providing a comprehensive understanding of the concepts crucial to this study. Chapter 3 comes next, where the methods and approaches employed are outlined. Moving forward, chapter 4 summarizes and evaluates the research questions posed while shedding light on the contributions made within this study. After that comes chapter 5 which examines the research findings and their implications for research on IS, the practical implications, and policies that could be based on the findings. This chapter not only shows the discoveries but also acknowledges limitations and paves the way for future explorations, outlining potential avenues for further research. As the last part of Part I, Chapter 6 concludes the thesis, offering final reflections and remarks that encapsulate the essence of the thesis, providing a summary of insights throughout this academic journey.

Part II presents the journal and conference papers that were produced for the purposes of this thesis. These papers combined encapsulate the original contributions, research methods, and findings of this thesis. Thus, reading Part II offers a deep understanding of Part I. The papers also serve as tangible evidence of the researcher’s ability to conduct independent investigations, demonstrating proficiency in data analysis, and suggesting new avenues for research since most of the papers were published or presented at conferences.

2 Theoretical Background

In this chapter, the main concepts of this thesis and the relevant work are presented so that the literature can be properly understood. The chapter begins with AI definitions and commonly used AI technologies in business. Next, the concept of RAIG and the principles that surround it are discussed. Finally, the chapter ends with the value of RAIG and the importance of compliance.

2.1 Artificial Intelligence and AI Technologies

2.1.1 Defining Artificial Intelligence

To understand the concept of AI, it is necessary first to understand the notions of "*artificial*" and "*intelligence*" separately. The term "*intelligence*" can be seen as involving mental activities, such as learning and reasoning (Lichtenthaler, 2019). The term "*artificial*" refers to an entity that is created by humans, rather than something natural (Mikalef et al., 2021). By merging these two aspects, AI can be understood as the creation of machines capable of simulating intelligence (Wamba-Taguimdje et al., 2020). Other similar definitions of intelligence are “*a person's ability to learn, cope with new situations, grasp and handle complicated concepts, and impact one's surroundings through knowledge*” (Demlehner and Laumer (2020a), or “*the ability to perceive and interpret information, transform that information into knowledge, and then apply that knowledge to goal-directed activities*” (Paschen et al., 2020). Hence, good intelligence adaptation entails tasks, such as problem solving, reasoning, learning, memory, and acting.

The popularity of AI and the attention it has received from firms, the media, and ordinary people is due to the recent advances in computers and, more specifically, to the hardware, internet network speed, the billions of bytes of available data, and the AI algorithms (Alsheibani et al., 2020), but still there is substantial uncertainty about what AI means as a concept. It can be defined as “*a system capable of interpreting external data, learning from such data, and using them to achieve specific goals and tasks through flexible adaption*” (Enholm et al., 2021). This uncertainty exists as AI contains several sub-disciplines and different AI approaches exist (Schmidt et al., 2020), and their terminology is often used synonymously to list a range of technologies and applications (Dwivedi, 2021). As a result, it is important to have a clear distinction between these core concepts and provide comprehensive definitions. In Table 3, a list of AI definitions is presented.

Table 3: Definitions of Artificial Intelligence.

Author(s) and date	**Definition**
Kolbjørnsrud (2017)	AI is defined as computers and applications that sense, comprehend, act, and learn.
Afiouni (2019)	AI is the general concept for computer systems able to perform tasks that usually need natural human intelligence, whether rule-based or not
Lee et al. (2019)	Artificial Intelligence: Intelligent systems created to use data, analysis, and observations to perform certain tasks without needing to be programmed to do so
Wang (2019)	AI is a broad concept that captures the intelligent behaviour of the machine
Makarius et al. (2020)	Artificial Intelligence: a system's capability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaption
Schmidt et al. (2020)	Artificial Intelligence: The endeavour to mimic cognitive and human capabilities in computers
Demlehner and Laumer (2020a)	Artificial Intelligence: a computer system having the ability to perceive, learn, judge, or plan without being explicitly programmed to follow predetermined rules or action sequences throughout the whole process.
Wamba-Taguimdje et al. (2020)	Artificial Intelligence: defined as a set of "theories and techniques" used to create machines capable of simulating intelligence. AI is a general term that involves the use of computers to model intelligent behaviour with minimal human intervention.
Mikalef et al. (2021)	AI is the ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals.

Based on the definitions in Table 3, it is clear that AI refers to giving the computers human-like capabilities, enabling them to perform tasks that only humans were able to do in the past. This emulation of AI agents requires inputs from multiple sources to understand the environment in which the agent acts (Eriksson, 2020), meaning that AI adopts and learns as a human would have done if exposed to a similar situation. This capability of learning is referred to as cognitive technology. Cognitive technologies imitate the human mind (Bytniewski et al., 2020), which is an attempt by the computer to function like a human being. This also implies that AI should not need to be explicitly programmed to perform an intelligent task (Demlehner & Laumer, 2020a). On the other hand, AI should have the ability to sense, interpret, learn, plan, comprehend, and act on its own (Demlehner & Laumer, 2020a; Kolbjørnsrud, 2017; Wang, 2019); thus, AI should be able capable of interpreting external data, learning from them, and accomplishing goals through flexible adaptation (Makarius et al., 2020) without following a rule-based system (Demlehner & Laumer, 2020a).

Furthermore, the definitions in Table 3 show two main approaches to defining AI. The first approach defines AI as a tool that solves tasks that could be impossible or very time-consuming for humans to complete (Demlehner & Laumer, 2020a; Makarius et al., 2020). The second approach defines AI as a system that mimics human intelligence and cognitive processes, such as interpreting, making inferences, and learning (Mikalef et al., 2021). While these two categories of definitions have differences, they also share a few important commonalities. A fundamental commonality is that AI is not created to replace humans, but instead, AI operates as an augmentation agent assisting in difficult and time-consuming tasks (Mikalef et al., 2021). In some ways, however, the two approaches differ.

While the first group of definitions regards AI as a tool only, not able to imitate human capabilities (Wamba-Taguimdje et al., 2020), the second group regards AI as capable of imitating human behaviour to a great extent (Kolbjørnsrud, 2017; Wang, 2019). Another noticeable difference is that some definitions refer to AI as a discipline of scientific study (Schmidt et al., 2020) or perceive AI as an applied capacity of a system (Afiouni, 2019; Lee et al., 2019). These definitions reveal that there are different aspects of what is considered important in AI and what really consists of the essence of an AI system. For this thesis, we assume that AI is an applied discipline with the goal of empowering systems for recognizing, interpreting, and learning from data to achieve organizational and societal objectives.

2.1.2 AI Technologies

One of the most commonly used methods in AI (if not the most common) is ML. ML became possible only after the advances in computational power (Afiouni, 2019). Similar to AI definitions, ML has a variety of definitions, as shown in Table 4. The main objective of ML is to train a model by learning from data and making inferences, predictions, and identifying patterns, which produces an outcome that is used for decision-making (Afiouni, 2019; Wang, 2019). ML models accomplish this by parsing big data (in most cases), learning patterns for these data, and coming to conclusions based on what has been learned (Wang, 2019). This is an inductive approach, where decisions are based on the collected data using statistics (Schmidt et al., 2020).

ML can be categorized into four main categories: *supervised*, *semi-supervised*, *unsupervised*, and *reinforcement learning* (Wang, 2019). In supervised learning, the training data includes the target value (Schmidt et al., 2020), and the system identifies patterns from the training data and creates its own rules from the labelled data (Afiouni, 2019). On the other hand, unsupervised learning approaches do not have a target value. Instead, the system analyzes the structure and the statistical properties of the data (Afiouni, 2019). Mainly, unsupervised learning is useful for discovering hidden patterns, creating clusters and detecting anomalies in a system (Schmidt et al., 2020). For example, email spamming or banking frauds are detected using unsupervised ML techniques. Semi-supervised learning is something between the two as it uses both labelled and unlabelled data (Harfouche, 2017). Reinforcement learning has a very different approach from the other categories as it does not need past data (Afiouni, 2019) since the system learns from continuous feedback, which is received in the form of rewards from an external environment (Harfouche, 2017). The idea is to maximize collective rewards by using trial and error techniques to make rational decisions in a dynamic environment based on feedback for each event that takes place (Afiouni, 2019).

Table 4: Sample Definitions of Machine Learning

Author(s) and date	**Definition**
(Wang, 2019)	Machine learning empowers the machine to "learn" without explicit programming. This learning process is accomplished by the machine itself through collecting data, analyzing data and making predictions.
(Wang, 2019)	The principle of machine learning incorporates training algorithms to enable machines to learn how to make accurate predictions. There are four training categories of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement.
(Afiouni, 2019)	Machine learning is that subset of AI that is capable of "learning from data and making predictions and/or decisions" without human dictated rules.
(Schmidt et al., 2020)	Machine learning uses an inductive approach in which decision rules are identified based on collected data using statistical methods.
(Wamba-Taguimdje et al., 2020)	Machine Learning - automatic learning: machines ’learn’ from the datasets offered to them

There are several ways to categorize ML applications, one of which is between *shallow* and *deep*. All four training categories are applicable to both types. Shallow-structured learning architectures are the most traditional, where AI learns from data described by pre-defined features (LeCun et al., 2015). Deep-machine learning, though, also known as deep-learning, can extract structure from data in a multi-layered manner (Wang, 2019). What sets deep learning apart from other ML techniques is the use of an artificial neural network (Wamba-Taguimdje et al., 2020), which aims to mimic the functionality of the human brain (Jelonek,

2020) by imitating human neurons (Schmidt et al., 2020). Deep learning works by creating a deep neural network with multiple hidden layers, where the higher layers learn complex concepts and the lower layers learn simple concepts (Harfouche, 2017); thus, it represents the world through a hierarchy of concepts, where each concept can be decomposed into simpler ones (Borges et al., 2021).

It is worth pointing out that while ML applications dominate the research interest in the IS field, there are also other AI technologies that have been examined in empirical studies and can be seen in Table 5. These technologies can work alongside ML or deep learning because they can provide better and more sophisticated solutions. A classic example would be chatbots, where both natural language processing (NLP) and ML are applied (Baby et al., 2017). NLP allows chatbots to understand and communicate using human language while, at the same time, the ML model learns and evolves as it gains access to more data (Castillo et al., 2021). Table 5 presents other AI technologies.

Table 5: Definition of other AI Technologies.

Technology	**Definition**	**Reference(s)**
Natural language processing (NLP)	NLP: the process through which machines can understand and analyze language as used by humans.	Jarrahi (2018)
Machine vision	Machine vision: Algorithmic inspection and analysis of images.	Jarrahi (2018)
Expert system	Expert systems are directed at imitating human decision-making by capturing and representing the expertise of experts for other organizational members to use, serving as a knowledge base.	Lichtenthaler (2019)
Planning and scheduling	The development of action strategies and sequences for subsequent execution	Lichtenthaler (2019)
Speech synthesis systems	Includes text-to-speech and speech-to-text solutions.	Lichtenthaler (2019)
	Text-to-speech: the production of speech by machines, by automatic conversion of text to a phonemic specification of the pronunciation of the sentences to utter.	Damper et al. (1999)
	Speech-to-text systems take a human speech utterance as an input and require a string of words as output	Ghadage (2016)

2.2 Responsible AI Governance

2.2.1 Definition of RAIG

RAIG can be defined as the structure of rules, practices, and processes used to ensure that the organization's AI technology sustains and extends the organization's strategies and objectives (Schneider et al., 2020). It is intended to empower individuals and organizations while maintaining fairness for all members of society, bestow trust in AI-use and expand AI capabilities in a responsible manner. It also involves taking special care about the use and maintenance of data (Brackett & Earley, 2017) while planning for AI implementation and decisions through data (Conboy et al., 2020). However, RAIG is not about data management but mostly about the procedures and mechanisms in the system that deal with gathering, managing, and using data. Individuals play a crucial role as they are responsible for the overall quality of the system (Benfeldt et al., 2020); thus, successful RAIG should hold all members who are part of data collection, administration, and implementation processes accountable. Moreover, RAIG relies on collaboration among the organizations and individuals that form the system and extend beyond the boundaries of a firm. For such multi-organizational environments, trusted frameworks are required to guarantee smooth transitions and operations between firms or customers while complying with the General Data Protection Regulation (GDPR) and other relevant laws and regulations. Based on the above, we define RAIG as:

> *A set of practices that documents the process involved in developing, applying, and monitoring AI applications and products while addressing all challenges that surround AI with a set of rules and authorities for (1) managing the appropriate functionality of AI, (2) assuring the trustworthiness of AI, and (3) overseeing the whole life cycle of data and algorithms within and between organizations and firms.*

It is worth noting that although there is a lack of clear and universally accepted definitions for RAIG, the existing literature uses terms like trustworthy AI or principled AI, which could be considered synonyms. The difference in terminology may be because the concept is not mature yet or because scholars are still focused on AI rather than its governance. As a result, scholars offer a range of different definitions and topics that surround responsible AI practices and principles. It is important to note that the work of "High Level Expert Group of Artificial Intelligence" (AI-HLEG) (European Commission, 2019) has influenced scholars to a great extent, and its impact is clear in most of the current academic work, including this thesis. Table 6 shows the different definitions of RAIG.

Table 6: Description of AI governance terms.

Name	Description	References
Principled AI	Principled AI consists of 8 themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, promotion of human values. The themes were derived from 35 papers.	(Clarke, 2019; Fjeld et al., 2020)
	Ethical framework of AI specifies five core principles: beneficence, nonmaleficence, autonomy, justice, and explicability.	(Floridi & Cowls, 2021; Thiebes et al., 2021)
	Beneficial AI consists of 20 principles, organized into three categories: research issues, ethics and values, and long-term issues. These are: Research Goals, Research Funding, Science-Policy Link, Research Culture, Race Avoidance, Safety, Failure Transparency, Judicial Transparency, Responsibility, Value Alignment, Human Values, Personal Privacy, Liberty and Privacy, Shared Benefit, Shared Prosperity, Human Control, Non-subversion, AI Arms Race, Capability Caution, Importance, Risks, Recursive Self-Improvement, and Common Good.	(Future of Life Institute, 2017; Pagallo et al., 2019)
Responsible AI	Responsible AI framework consists of 10 principles: well-being, respect for autonomy, privacy and intimacy, solidarity, democratic participation, equity, diversity inclusion, prudence, responsibility, and sustainable development.	(Dignum, 2017, 2019; Liu et al., 2022)
	Explainable AI is a suite of algorithmic techniques generating high-performance, explainable, and trustworthy models.	(Adadi & Berrada, 2018; Kaur et al., 2022; Li et al., 2021; Zou & Schiebinger, 2018)
Trustworthy AI	Trustworthy AI (TAI) consists of three complementary value-based principles for the responsible stewardship of AI: lawful, ethical, robust.	(Chatila et al., 2021; Liu et al., 2022; Theodorou & Dignum, 2020)
	TAI consists of seven key requirements for achieving technical trust: availability, reliability,	(Chatila et al., 2021; European

	safety, confidentiality, integrity, maintainability, security.	Commission, 2019; Mora-Cantallops et al., 2021; Wu et al., 2020; Zicari et al., 2021)

2.2.2 Principles of Responsible AI Governance

AI ethical standards and principles have been established by governments, researchers, and corporations. These concepts contain various aspects of responsible AI, including AI interpretation, safety and testing, and ethics in existing AI systems (Wu et al., 2020). Previous research focused on specific aspects of responsible AI, such as bias elimination (Brighton & Gigerenzer, 2015), explainability of AI outcomes (Arrieta et al., 2020), and safety and security (Srivastava et al., 2017). Recently, there has been a shift towards a deeper understanding of what responsible AI really entails (Theodorou & Dignum, 2020). The European Commission has taken steps in this direction by forming an independent expert body, the AI-HLEG, with the goal of developing an integrated framework for responsible and trustworthy AI (European Commission, 2019). The AI HLEG promotes Trustworthy-AI, with three key criteria in mind: (1) lawful, complying with all applicable laws and regulations; (2) ethical, ensuring compliance with ethical principles and values; and (3) robust, from a socio-technical perspective, which means having the ability to withstand and adapt to different challenges and disruptions in the environment that encompasses both social and technical elements. Each component is important but not sufficient on its own. All three components should work at the same time, and if there are tensions between them, society as a whole should work to harmonize them. Similarly, the Singapore government moved in the same direction and recognized the future AI difficulties that might appear related to discrimination, biased outcomes, decision outcomes, laws, and regulations (Singapore Government, 2020). The Singapore Government (2020) model framework is based on two high-level concepts of AI trust. The first level is about companies using AI for decision-making and how to guarantee clear, explainable, and fair processes. Although explainability, transparency, and fairness are hard to achieve in full, companies should make every effort to ensure these values, contributing to the development of AI. The second level comprises AI solutions that are human-centered. Human interests, including well-being and safety, should be key considerations in the design, development, and deployment of AI as it is used to augment human skills.

A recent report from Harvard University's Berkman Klein Center for Internet & Society highlighted 38 similar efforts from various corporations and groups (Fjeld et al., 2020). There is a clear consensus that RAI represents principles that ensure ethical, transparent, and accountable usage of AI technology in line with user expectations, corporate values, and societal laws and conventions (Flavián & Casaló, 2021). Winfield and Jirotka (2018) describe responsible principles as a collection of processes, procedures, cultures, and beliefs that secure the highest levels of conduct. They emphasize the ethical parts of AI Governance, and they argue that these principles go beyond these principles and promote ethical behaviours inside the organizations. Winfield and Jirotka (2018) state that these are critical components of responsible research and development, which "entails an approach, rather than a mechanism, so it seeks to deal with ethical issues before they arise in a principled manner rather than waiting until a problem surfaces and dealing with it in an ad-hoc way".

Overall, the principles are accountability, diversity non-discrimination and fairness, human agency and oversight, privacy and data governance, transparency, technical robustness and safety, and societal and environmental well-being. It is worth noting that the principles might appear with some variation in the word choices. For instance, "transparency" might appear as "transparency and explainability", while "human agency and oversight" might appear as "human control of technology". Table 7 shows the principles and their sub-principles.

Table 7: Responsible Principles.

Principle	**Sub-dimensions**	**References**
Accountability	**Auditability:** the ability of an AI system to be assessed for its algorithms, data, and design processes. **Responsibility:** the oversight of the various stages and activities involved in AI deployment–and how it should be allocated to appropriate departments.	(de Almeida et al., 2021; European Commission, 2019; Mikalef et al., 2022)
Diversity non-discrimination and fairness	**Accessibility:** the design of systems in such a way as to make them accessible and usable for everyone, regardless of their age, gender, abilities, or characteristics. **No unfair bias:** the rejection of prejudice towards or against people, objects, or positions, as well as inherent biases in datasets, which can lead to undesirable outcomes like unintended discrimination.	(Fjeld et al., 2020; Singapore Government, 2020)
Human agency and oversight	**Human review:** the right of a person to challenge a decision that has been made by an AI. **Human well-being:** the idea that AI must include human well-being as a primary success	(European Commission, 2019; Singapore Government, 2020)

	factor for development.	
Privacy and data governance	**Data quality:** the accuracy of values in a dataset, matching the true characteristics of the entities described by the dataset. **Data privacy:** the development and operation of AI systems in a way that takes data privacy into account throughout the data lifecycle. **Data Access:** the national and international rights laws, during the design of an AI, for data access permissions.	(Matthews, 2020; Singapore Government, 2020)
Technical robustness and safety	**Accuracy:** the ability of an AI system to make correct judgments, such as correctly classifying information into the appropriate categories, or being able to predict, recommend, or make intelligent decisions based on data or models. **Reliability:** the ability of an AI system to work properly within a range of inputs or various situations. **General Safety:** the safety rules and fallback plans that should be established for AI systems in the event of problems. **Resilience:** the AI systems which should be protected against vulnerabilities that can be exploited by adversaries, e.g. hacking.	(European Commission, 2019; Singapore Government, 2020)
Transparency	**Explainability:** the ability to explain both the technical processes of an AI system and the related human decisions (e.g., application areas of a system). **Communication:** the human right to be informed in advance when interacting with an AI agent. **Traceability:** the ability to track data and processes that yield the AI system's decision, including data gathering, data labelling, and algorithms.	(Fjeld et al., 2020; Mikalef et al., 2022; Singapore Government, 2020)

Societal and environmental well-being	**Societal wellbeing:** the ubiquitous exposure to social AI systems in all areas of society, such as work and education, where people do not need to occupy positions that are considered filthy and dangerous. **Environmental well-being:** the promise to tackle some of the most pressing environmental concerns and not damaging the environment.	(European Commission, 2019; Singapore Government, 2020)

2.2.2.1 Accountability

Accountability refers to mechanisms and processes for ensuring responsibility and accountability throughout the development and deployment phases while regulating AI use for auditability and responsibility. Vollmer et al. (2020) recommend that auditability should be evaluated during data acquisition because it is essential to investigate the data concerning its usage and answer questions about data distribution and sample representation. This enables data comparison in different periods. It is difficult, though, to hold AI accountable because of its “black-box” decision-making nature (Caner & Bhatti, 2020). Hence, it is important to have technologies in place to detect flaws and evaluate AI systems in general (Matthews, 2020). One example of how AI systems may be used is to screen CVs as part of a recruiting process. In contrast to humans, an AI system cannot be held personally responsible for this kind of work. The topic of who is truly accountable for a judgement made by an AI system is difficult to resolve (Ayling & Chapman, 2021) because some organizations rely too heavily on AI and some firms have been criticized for their lack of AI support (Schlögl et al., 2019). Nonetheless, despite all this, AI is criticized for being untrustworthy and unreliable when it comes to accountability; thus, it is argued that the employees within the organization should be held accountable rather than the system itself (Ryan, 2020).

2.2.2.2 Diversity, non-discrimination and fairness

Diversity, non-discrimination and fairness refer to the diversity of the datasets and the processes that have been developed in order to involve all people despite their uniqueness. Inclusion is a key aspect of RAI (Korinek, 2019). There are many AI systems that produce discriminating outcomes. Two examples are credit ratings and criminal sentences (Taeihagh, 2021). The European Commission (2019) recommended that AI systems should employ proper mathematical and statistical methods to uncover unintended and develop an approach for removing algorithmic bias (Korinek, 2019). Based on that, systems should be user-centric, allowing all individuals to utilize AI products, regardless of age, gender, and skills. Other types of bias affect language because language is a highly complicated concept and includes features such as word grouping and ordering rules. Using natural-language datasets to train models can lead to a range of biases, and detecting this form of prejudice might be challenging. For example, word groupings of men and women, or the order in which gender appears in a list, adjectives associated with them, and frequency may all encourage bias in the dataset and hence alter the model (Leavy, 2018). Hence, it becomes clear how feeding a biased dataset into a "black-box" system may cause unfair outcomes. The development of technological solutions for data processing is a good starting point, while AI teams should have a specialized bias testing leader to ensure that prejudice is constantly avoided (Shneiderman, 2020).

2.2.2.3 Human agency and oversight

Human agency and oversight refer to the mechanisms that allow cooperation between human and machine in a way that benefits the human. AI systems should guarantee human autonomy and decision-making that promotes a democratic, prosperous, and equitable society (European Commission, 2019). This means that AI has to create a just and welfare state by empowering users and protecting their fundamental rights. Users should have the ability to make autonomous decisions and they should receive the right tools to better understand and interact with AI systems. If possible, users should be able to evaluate or question policies regarding AI and AI systems should provide individuals with information and insights to help them make better choices that align with their goals. However, there is a risk that AI could influence human behaviour at the expense of individual autonomy, which might not be easily detected. To ensure user autonomy, it is important that AI systems prioritize user rights rather than imposing automated decisions (Shneiderman, 2020). One way to do that is by having various governance mechanisms, such as human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) (Caner & Bhatti, 2020). HITL includes human involvement in every decision cycle. The HOTL allows for human intervention during design and ongoing maintenance. HIC refers to the ability to monitor the overall function of an AI system, including its broader impact, and decide when and how to implement it. This includes choosing not to use AI in certain situations or to bypass AI decisions (Matthews, 2020).

2.2.2.4 Privacy and data governance

Privacy and data governance refers to a framework for managing the availability, usability, integrity, and security of data on the basis of internal policies and standards. Ensuring privacy and preventing harm requires data governance that includes data quality, integrity, relevance, accessibility, and privacy protection in data. Processing AI systems must provide privacy and data security guarantees over their lifetime, including not only the initial information provided by users but also the information obtained about them as they interact with the system (Matthews 2019). Digital records of human behaviour can reveal sensitive personal information such as preferences, sexual orientation, age, gender, and religious and political beliefs (European Commission, 2019). It is important to have a trusted collection of information system to ensure that the collected information about individuals is not unlawfully used or they are not unfairly discriminated against. For this purpose, the quality of the data used is very important for the functioning of AI systems since data may have biases, inaccuracies and errors that need to be addressed before training an AI model (Papagiannidis et al., 2021). Cultural variations should also be considered throughout data collection. This is especially true in Western countries, which have a natural separation between the governmental and private worlds because users are the best judge of how to handle their privacy (Ayling & Chapman, 2021). Additionally, data integrity must be protected to prevent undesired outcomes that can alter the behaviour of AI systems, especially those systems that continuously learn from their environment and update their behaviour. Hence, organizations that handle personal data need to have a clear data access policy. These protocols should define who can access the data and under what conditions, while data-access should only be granted to qualified individuals with the necessary knowledge and for a reasonable cause.

2.2.2.5 Technical robustness and safety

Technical robustness and safety refer to the mechanisms that minimize harm and ensure that AI works as intended. The system should be designed to handle potential changes in their operating environment and interactions with other agents, human or artificial, and safeguard the physical and mental well-being of humans (European Commission, 2019). Like any software system, AI systems should be safeguarded against vulnerabilities that could be exploited by hackers, because attackers may target data, models, and hardware (Hamon et al., 2020). If an AI system is attacked, its data and behaviour can be altered, resulting in different decisions or even a system shutdown. Malicious intent can be considered as an attack too, and a good security plan should identify unintended uses or abuses of the AI system and take steps to prevent them. Additionally, AI systems should have safeguards that allow for a fallback plan, such as switching to a rule-based approach or involving a human operator before proceeding (Smuha, 2021). The level of safety measures required depends on the magnitude of the risk posed by the AI system, which is influenced by its capabilities (Hamon et al., 2020). When high risks are in play, proactive measures need to be developed and tested. Therefore, accuracy is crucial for an AI system's ability to make correct judgements, such as proper classification of information or accurate predictions and recommendations. In cases where occasional inaccuracies are unavoidable, the system should provide an indication of the likelihood of such errors and have a review process (Kuziemski & Misuraca, 2020). High accuracy is particularly vital in situations where the AI system directly impacts human lives. The reproducibility and reliability of AI system results are essential too, because a reliable system operates effectively with a variety of inputs and in different situations. Reproducibility refers to whether an AI exhibits the same behaviour when exposed to the same conditions (Chang et al., 2022), enabling developers and policymakers to accurately describe the functioning of AI systems.

2.2.2.6 Transparency

Transparency refers to procedures that boost explicability, allowing see-through on elements relevant to an AI system, i.e., data, processes and business models. AI will affect the lives of millions, but only AI experts understand the techniques used by AI systems (Gasser & Almeida, 2017) and, because of that, trust issues arise on the part of the people who either use an AI system or are affected by it. Hence, a transparent system must be able to trace and document important decision-making processes, including data collection, labelling and algorithms (Larsson et al., 2019). This ability should also be extended to decisions, enabling identification of the cause of any mistakes, which can help prevent harm in the future, and this is known as Explainable-AI (XAI) (Gillath et al., 2021). Explainability has a cost, though. There may be a trade-off between increased precision (potentially reducing accuracy) and increased accuracy (at the cost of explainability) (Reddy et al., 2020). However, when AI has a significant impact on people’s lives, it is crucial to provide clarifications appropriate to the program's decision-making process, policy choices, and the rationale behind its use, ensuring a clear business model and preventing future mistakes (Mezgár, 2021). Traceability also facilitates auditability and interpretation. AI systems should be designed in a way that allows audits by third parties and gives room to interpret the results. Another important aspect of transparency is that AI users should always be aware that they are communicating and interacting with AI, allowing them to switch to a human agent and ensure compliance with fundamental rights (Felzmann et al., 2020). Furthermore, the capabilities and limitations of the AI system must be communicated to AI operators or end-users in a manner appropriate for the use case at hand and this needs to include communication of the accuracy level of the AI provisions and limitations.

2.2.2.7 Societal and environmental well-being

Societal and environmental well-being refers to preserving the prosperity of the broader society and the safety of the environment and having both as stakeholders throughout the entire life cycle of an AI system (European Commission, 2019). Encouraging sustainability and ecological responsibility in AI systems is crucial. Ideally, AI systems should be utilized to benefit all human beings, including future generations, in an environmentally friendly manner. This entails evaluating the system's development, deployment, and use processes, as well as its entire supply chain, with a critical focus on resource usage and energy consumption, choosing less harmful alternatives (Venkataramanan et al., 2019). Careful considerations are needed for social AI systems as they are present in various aspects of our lives, including education, work, care, and entertainment; thus, AI has the potential to influence our perception of social expectations and impact our social relationships and attachments or even completely replace humans at work (Gasser & Almeida, 2017). Hence, AI systems can enhance social skills, but they can also contribute to their deterioration, potentially affecting people's physical and mental well-being. For example, AI could be used to replace people in filthy and dangerous occupations (Zhang et al., 2021), but it may result in "cold care" for the elderly, if machines replace nurses. Therefore, it is crucial to carefully monitor and consider the effects of these systems. In addition to assessing the impact of AI system development, deployment, and use on individuals, it is equally important to evaluate their societal impact, taking into account their effects on institutions, democracy, and society as a whole (Winfield & Jirotka, 2018; Wirtz et al., 2020). As a result, the use of AI systems, especially in situations related to the democratic process, including political decision-making and electoral contexts, should be thoroughly considered and analyzed.

2.3 Business value of Responsible AI Governance

The increasing number of incidents involving AI has highlighted the value of responsible AI (Fuchs, 2018). For instance, Amazon developed software to automate the process of examining resumes with the goal of identifying the top candidates (Tschang & Almirall, 2021). However, in 2015, it was discovered by Amazon's ML experts that their AI recruiting tool exhibited gender-based discrimination against women in technical fields, such as software development. Amazon's ML algorithms were trained on resumes submitted to the corporation over a 10-year span. The AI models used gender-biased data and concluded that women were not well-suited to technical professions. One of the main reasons for that bias was the fact that there was a disproportionate number of resumes from men, which determined the outcome of the model (Dastin, 2022). Furthermore, the value of responsible principles came from challenges unique to AI that require attention, such as the governance of autonomous intelligent systems, responsibility and accountability for algorithms as well as privacy and data security (Wirtz et al., 2020). As a sign of proof, RAI has gained recognition in policymaking, with a few countries outlining what they consider as fundamental principles that describe RAI (Jobin et al., 2019). For instance, the AI readiness index looks at the extent to which countries include AI technologies and now also looks for a new sub-index that measures the adoption of RAI principles and practices that are included when designing AI products.

A great value that RAI practices offer is corporate reputation (Wang et al., 2020). Deviations from social norms can have various consequences. When a company fails to incorporate RAIG practices, its legitimacy is undermined, resulting in a loss of community respect and support (Dai et al., 2018). As a result, the company faces criticism and scrutiny, and the public encourages local or state governments to impose regulations and restrictions on its activities. By ignoring ethical considerations in AI practices, a company risks being caught acting illegally and acting against the broader public interest (Dellmuth & Schlipphak, 2020). If consumers perceive a company’s actions to be morally unacceptable, they may reject its products, further hurting its legitimacy. Hence, the absence of definable AI or accountable principles may intersect with both corporate governance and public opinion, as the corporation has to address community concerns, fears, and concerns affecting the community's well-being (Stupak et al., 2021).

Another reason why RAIG is important and adds value is the challenge of understanding and interpreting AI results. The governance of autonomous intelligent systems, for example, answers the question of how to control autonomous agents but it is not clear how AI makes decisions, especially in some rare scenarios where the AI models are not well trained (Azzutti et al., 2022). This is often referred to as the “black-box” problem, where AI might make unpredictable decisions and cause harm. In a worst-case scenario, where AI is used for military purposes (assuming that NATO and other superpowers continuously increase the use of AI in their systems as they are currently doing with autonomous drones), the AI might be programmed to eliminate all threats and AI might classify even civilians as a potential threat (Johnson, 2019; Mikalef et al., 2022). Some even worry that AI accesses resources, digital or physical, and will eventually pursue its own goals, harming humans in the process (Nath & Levinson, 2014). Consequently, concerns like this gave birth to questions regarding the transparency and accountability of AI systems and although humans are in control of AI systems, the ability of AI systems to learn on their own makes it impossible for operators or developers to predict all actions and outcomes. Therefore, a thorough stakeholder evaluation is required to ensure transparency in AI systems if they want to maximize their value (Helbing, 2019). What is more, the availability of massive amounts of data and the new digitalization opportunities have become a global driver for competitiveness. If we consider AI as the main ingredient for success in such highly competitive environments, where the digitalization process will need AI to maximize its value, then RAIG will be the recipe for its success across all domains and sectors, and RAIG will also assist in global challenges, such as the United Nations Sustainable Development Goals (SDGs) (Jelinek et al., 2021).

Finally, RAIG helps to optimize AI use in order to increase efficiency and reduce operational costs and the risks that come with AI (Abraham et al., 2019; Mikalef et al., 2019). The range of risks may vary from systemic bias, illegal acts, massive financial exposures, political disruptions, and the loss of lives. Businesses try to distance themselves from the association of these risks by using diverse algorithms with data gathered by (and about) governments, enterprises, and individuals (Janssen et al., 2020), hoping that through diverse data collection their outcomes will be trustworthy. However, those who develop RAIG frameworks face two interrelated challenges. First, the emerging AI governance structure is gradually overlapping with the existing cyber regime complex, which is already fragmented, undermining international collaboration. Second, it creates many conflicts in international interest, especially among the major powers, i.e., the United States, China, EU, and Russia, since whoever manages to succeed first will have a tremendous technological advantage over its rivals (Jelinek et al., 2021). Therefore, AI is increasingly influencing the competitiveness of entire nations and regions, making RAIG reliant on collaborative efforts guided by common standards, which could be hindered by structural fragmentation. As for privacy and safety, RAIG can add a lot of value as it can impose safeguards when dealing with human rights and privacy as well as impose procedures for protecting individual data from unlawful external access. Even today many companies use AI technologies to gather data without notifying or receiving direct approval from the users (Remolina Leon & Seah, 2019). For example, when an individual uses a navigation system to find an alternative route home from work, the system needs to access the user's current location (Margetts, 2022). However, this information could potentially be used to create a user profile by the search engine; thus, without explicitly taking the user's consent, these AI applications and services pose a risk to their privacy (Wirtz et al., 2020).

3 Research Methodology

In this chapter, the research philosophy, methodologies, and strategies utilized to address the research questions of the thesis are presented. The research philosophy informs both the methods and the strategies employed in a study as it shapes the researcher's approach to investigating a particular phenomenon or problem. For example, when it comes to methods, a positivist researcher assumes an objective reality that can be measured through empirical methods such as surveys or experiments. In contrast, an interpretive researcher assumes a subjective reality that can be interpreted in multiple ways, and he uses qualitative methods such as interviews and observations.

3.1 Research Philosophy

For the purpose of this thesis, our research incorporates ideas from positivism and interpretivism, providing a more thorough and detailed explanation of our work. Positivism was chosen because it promotes organized and methodical research by emphasizing objective facts, measurable evidence, and empirical observation (Kankam, 2019). Conversely, interpretivism was chosen because it emphasizes the significance of context, meanings, and social structures in forming reality while acknowledging the subjective character of human experiences (Klein & Myers, 1999). By applying both paradigms, we took advantage of positivism's objectivity and rigour and interpretivism's contextual richness and depth of knowledge. To be more specific, the interpretivist approach acknowledges that individuals within a society perceive and comprehend the same "objective" reality through distinct perspectives, driven by personal motivations for their actions (Alharahsheh & Pius, 2020). Studies adopting the interpretivist philosophy commonly focus on the meaning and may adopt a variety of methodologies to encompass diverse aspects of the issue (Al-Ababneh, 2020). The positivist research approach is embraced when the objective entails investigating phenomena using a systematic and empirical approach, emphasizing quantifiable data and scientific principles (Junjie & Yingxin, 2022). This philosophical stance is notably well-suited when researchers aim to identify causal relationships, test hypotheses, and generalize findings to a larger population (Al-Ababneh, 2020). Therefore, by accepting both, we used a variety of techniques, such as qualitative methods to investigate subjective experiences and meanings and quantitative procedures to collect empirical data (Junjie & Yingxin, 2022).

The relevance concerning RAIG and its business value with the research philosophy can be summarized as follows. Positivism's emphasis on objective facts and measurable evidence is crucial for assessing the tangible impact of RAIG on business outcomes. By quantifying RAIG practices, researchers can systematically measure the influence of RAI initiatives on organizational performance, risk management, and stakeholder trust. In addition, by collecting quantitative data, researchers can analyze trends and patterns to assess the effectiveness of RAIG practices. At the same time, interpretivism's focus on context and structures is essential for understanding how RAIG is implemented within different organizational contexts. This perspective acknowledges that the effectiveness of RAIG may vary depending on factors such as organizational culture, regulatory environment, and stakeholder expectations. By considering these differences, researchers can gain insights into the challenges and opportunities associated with implementing RAIG in diverse business settings.

3.2 Research Overview

Different research objectives (ROs) were developed for each issue in order to shape the assessment of the research questions and guide the selection of research methodologies and approaches. The ROs are as follows:

RQ1*: How do organizational factors influence the adoption and implementation of AI technologies?*

- RO1.1: Identify the key organizational factors that influence the adoption and implementation of AI technologies.
- RO1.2: Examine the impact of organizational structure and decision-making processes on the adoption and implementation of AI technologies.
- RO1.3: Examine the ethical and social implications of AI adoption and implementation within organizations and explore ways to address potential ethical concerns.

RQ2*: What are the key drivers and mechanisms for generating value from AI in organizations?*

- RO2.1: Investigate the mechanisms through which AI generates value in organizations, including such factors as improved decision-making, increased operational efficiency, and innovation.
- RO2.2: Explore the overall process for structuring a company’s resources, bundling the existing resources for the creation of new capabilities.

RQ3*: What are the key antecedents and effects for generating value from RAIG in organizations?*

- RO3.1: Identify and examine the key antecedents of generating value from RAIG in organizations, including such factors as organizational culture, leadership and governance structures.
- RO3.2: Investigate the mechanisms through which RAIG generates value in organizations, including such factors as improved decision-making, ethical and responsible AI practices.

RQ4*: How does effective responsible governance of AI initiatives impact organizational outcomes, including competitive advantage and performance gains?*

- RO4.1: Examine the relationship between the responsible governance of AI initiatives and competitive advantage, exploring how responsible practices can contribute to gaining a competitive edge in the market.
- RO4.2: Investigate the influence of responsible governance on stakeholder perceptions, including trust, reputation, and brand image, and how these perceptions contribute to organizational outcomes.

To address the objectives outlined in RO1.1, our initial steps encompassed a comprehensive literature review aimed at examining existing research. This focused exploration concentrated on the foundational factors that propel or hinder the adoption and integration of AI technologies. Through this careful review, we sought to recognize gaps within the current literature landscape while formulating a research roadmap that outlines areas suitable for further examination. Building upon these insights, our pursuit of RO1.2 emerged through the creation of a holistic framework. This detailed framework embraced various facets, effectively encapsulating the multilayered nature of AI integration. Expanding our investigative horizons, RO1.3 motivated a deep dive into the negative dimensions of AI applications. A single case study approach was adopted, involving 14 expert interviews. The primary aim here was to identify the negative aspects posed by AI use and, alongside this, uncover the adaptive strategies implemented by companies to navigate this transformative shift. Transitioning to RO2.1, the insights from the single-case study proved instrumental in achieving this objective. This pivotal understanding laid the foundation for a multi-case analysis for RO2.2, and this centred around the practices and structures companies employ when using AI. Continuing our quest, RO3.1 and RO3.2 prompted a second exploration into the landscape of the literature, with a specific focus on the RAIG. Our objective was to define the key elements that optimize AI's effectiveness while concurrently mitigating potential costs and risks. To validate our research, conceptual models were created, encapsulating our findings. The final goal of these efforts was the validation of our qualitative insights through quantitative means for RO4.1 and RO4.2. Throughout this journey, our methodology holistically embraced both qualitative and quantitative approaches, forming a cohesive plan for understanding the complex interplay between AI, business dynamics, and value generation.

3.3 Research Activities

3.3.1 Literature Review

The review was conducted in distinct phases, based on the established framework of a systematic literature review to ensure the comprehensive incorporation of all pertinent literature up to the present time (Kitchenham, 2004; Okoli, 2015; Templier & Paré, 2015). The initial systematic literature review started by devising a review protocol in alignment with the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2008). In contrast, the second review aimed to offer insights into current research and pinpoint research gaps; thus, we selected a scoping review approach (Paré et al., 2015). A protocol was formulated, detailing the course of primary research and specifying search terms and sources for the literature review (Boell & Cecez-Kecmanovic, 2015). The review unfolded across five sequential steps, which were similar for both papers. Step one involved gathering information based on previously formulated keywords, followed by a selective process. The documents underwent iterative filtering based on relevance, with two rounds of review. The initial round assessed titles, and the second went over abstracts. Moving to the third step, the evaluation of study quality was initiated. The quality assessment stage involved critically evaluating the methodologies employed in the full-text reports. The fourth step encompassed data extraction, involving the systematic retrieval of specific information from each article, which was then organized in a spreadsheet. The fifth and final step involved data synthesis, facilitated through the utilization of a concept matrix. The concept matrix served as a tool to establish connections between distinct research articles. For a comprehensive overview of the process steps see Figure 2, while additional details can be found in MRP1 and MRP2 within part II.

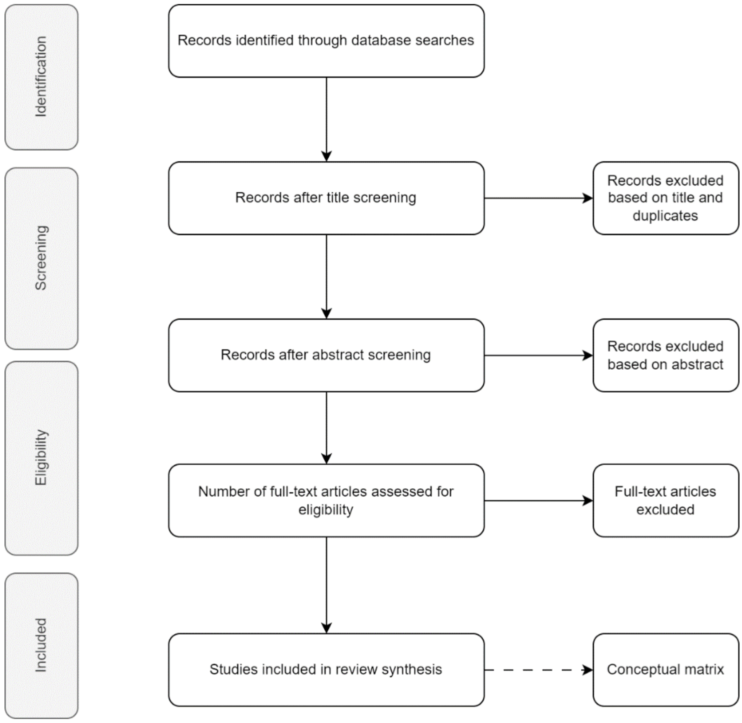

Figure 2: Stages of the study selection process for the literature reviews.

3.3.2 Interview

Interviews serve as a powerful tool for information collection, especially when researchers lack predefined theories or assumptions (Qu & Dumay, 2011), and they are equally effective for refining existing theories and gaining insight into complex phenomena (Tallon et al., 2013). Our research approach involves deconstructing interview data into distinct structural, procedural, and relational practices. These practices form a foundational framework that contributes to the development of effective AI Governance strategies. This research follows a qualitative paradigm, utilizing qualitative data to explore and explain the relevant research questions (Michael, 1997). The chosen methodology involves semi-structured interviews that tap into the experiences, beliefs, and attitudes of key respondents (Wynn Jr & Williams, 2012). Identifying suitable participants was done by reaching out to the human resources department and individuals suitable for handling such communications, i.e., managers. Initial introductions were made via email, supplemented by phone conversations where necessary. Our interviewee criteria encompass individuals who hold key positions, are familiar with AI technologies, and have significantly contributed to AI development, building a comprehensive understanding of AI development over time. Based on these criteria, we ended up with 29 interviews. Participants from both the IT and business/trade departments provided insights, reflecting the need for a diverse approach to understanding RAIG, allowing for a more holistic view of the subject (Pessoa et al., 2019). The interview questions were designed to be open-ended, facilitating engaging and insightful discussions, and this approach enabled interviewers to adapt their questions based on responses and even explore aspects beyond the initial interview guidelines. Prior to each interview, a thorough explanation of the research objectives and expected outcomes was provided to the interviewees. Additionally, interviewees were encouraged to contribute any relevant insights that they felt might enhance the research, thereby fostering a collaborative and comprehensive approach to data collection. For more in-depth information on the interview methodology employed, see MRP3 and MRP4 in part II.

3.3.3 Case Study

Various issues within their real-life contexts can be comprehensively explored from multiple angles by using case studies (Rashid et al., 2019). The analysis of case studies serves as a valuable tool for navigating deeper into the understanding of a particular issue, event, or phenomenon within its natural environment. Case studies have historically been perceived as lacking in rigour and objectivity compared to other forms of social research; thus, case studies as a research strategy have to be designed and implemented with careful justification (Gibbert et al., 2008). Despite initial reservations about the validity of case studies, they are widely embraced due to their potential to offer insights that might otherwise remain difficult to catch (Yin, 1981). Furthermore, case studies are frequently employed to develop more structured research instruments, thereby enabling interviews, surveys, or experiments during the exploratory phases of a project. They provide valuable data, especially in contemporary scenarios where direct manipulation of relevant behaviours is not feasible. Typically, case study research draws evidence from diverse sources such as documents, artifacts, interviews, and observations. For the purposes of this thesis, both a single case study and a multi-case study approach were employed, and we investigated four companies. The selection criteria for these cases were based on shared characteristics, including industry, utilization of AI systems, team size, and the cultural environment. All selected firms operate within the same industry and possess comparable abilities in collecting, analyzing, and interpreting data for strategic decision-making. A common thread among these firms is their belief in the imminent necessity of developing, expanding, and adopting AI systems to maintain or gain a competitive edge over rivals and emerging players. Moreover, the nature of AI projects undertaken by these firms indicates shared challenges, necessitating similar solutions. For more in-depth information on the case studies, see MRP3 and MRP4 in part II.

3.3.4 Questionnaire - Survey

Questionnaire-based surveys, as a way to do quantitative research, offer many advantages as they help in gathering a great deal of information from many diverse participants, which is important for understanding the research topic and analyzing the data. Questionnaire-based surveys provide an organized structure for a framework, ensuring consistency when collecting data, even if the data is about people's thoughts, feelings, and actions. Moreover, it covers a wide range of topics and can be done in a relatively short time (Straub et al., 2004) and it works well for finding connections between different groups of people and patterns that might not be easy to see with other methods. Because questionnaires follow a structure, the information collected is consistent and easy to understand and the data analysis is easy to perform. At the same time, questionnaire-based surveys let people share their opinions without feeling uncomfortable. In this thesis, we follow the guidelines from Recker and Rosemann (2010) and after we created our trial questionnaire, we shared it with a panel of experts for careful assessment and refinement of indicators, questions, and wording. Then, we sent two internet questionnaire-based surveys to Nordic businesses, where we used a 7-point Likert scale, where a value of one means disagree entirely, and seven means agree entirely.

The initial questionnaire was conducted over a period of four weeks in April and May of 2021. We collected data for the questionnaire using two methods. The first approach involved contacting key IT leaders through email invitations. After sending the initial email, we followed up with a reminder a week later. Each participant received a personalized report comparing their responses to the survey averages. Additionally, we collaborated with a panel service company to expand data collection across Nordic countries, targeting senior IT executives. At the conclusion of the data collection phase, we obtained two separate datasets. The first dataset resulted from the email-based approach, generating 24 complete responses. The second dataset was provided by the panel service company and included 120 complete responses. Combining these two datasets created a comprehensive dataset with 144 complete responses for further analysis.

Likewise, we followed a similar process for the second questionnaire. This was conducted from November 2023 to December 2023. We used the same approach as described above and we gathered 329 complete responses. We explored relationships between RAIG, legitimacy, communication use, and firm performance. The investigation stated that RAIG significantly impacts both internal and external legitimacy positively. It aimed to explore how the implementation and adherence to robust RAIG principles within organizations contribute to enhancing their perceived legitimacy, both within the organization (internal) and among external stakeholders (external). In addition, we investigated the role of RAI communication use as a moderator, suggesting that effective communication strategies regarding RAI practices can strengthen the relationship between RAIG and legitimacy (both internal and external). We also examine how organizations communicate their AI governance initiatives and how these initiatives influence the acceptance of their legitimacy by stakeholders. The end goal is to focus on the relationship between the outcomes of legitimacy and firm performance. We propose that both internal and external legitimacy positively influence a firm's performance. This exploration aims to understand how the perceived legitimacy, either within the organization or among external parties, impacts the overall performance metrics of a company. By exploring these relationships, we uncovered the interplay between RAIG, legitimacy, communication strategies, and firm performance. Hence, we provide valuable insights into how Responsible AI practices, coupled with effective communication, affect the perceived legitimacy of organizations and, subsequently, their performance outcomes.

For more in-depth information on the questionnaires, see MRP5 and MRP6 in part II.

3.3.5 Mapping Methodologies and Research Questions

To address the research questions, we employed diverse methodologies tailored to the distinct nature of each research question. The initial phase heavily relied on an extensive literature review, serving as the cornerstone that laid the foundation for our entire research. The literature reviews not only equipped us with a robust understanding of the field but also played a pivotal role in shaping and formulating the subsequent research questions (RQ1, RQ2, RQ3, and RQ4) that drove this thesis forward. To delve into RQ1 and RQ2, we utilized interviews and case studies. These qualitative methodologies provided a rich, contextual understanding and insights into the intricate dynamics of the subject matter. Through interviews, we gained firsthand perspectives and narratives, while case studies allowed us to dissect real-world scenarios, enriching our comprehension and the analysis of the phenomena under investigation. To address RQ4, we used a questionnaire-survey. This quantitative approach enabled us to gather structured data from a broader sample, facilitating a comprehensive analysis and empirical validation of certain aspects of our research questions. The use of these methodologies was deliberate, aligning with the distinct demands of each research question. This holistic approach ensured a deep exploration of the topic, combining qualitative depth with quantitative breadth, thereby enriching the overall depth and credibility of our findings. To check the outline of the methodologies applied for each respective research question, see Table 8.

Table 8: Mapping of research methodologies and research questions.

	RQ1	RQ2	RQ3	RQ4
Literature review	•	•	•	•
Interview	•	•		
Case study	•	•		
Questionnaire - Survey				•

3.4 Research Validity

3.4.1 Construct Validity

Construct validity involves assessing the appropriateness of measurements for the researched theoretical concepts (Peräkylä, 2004). We enhance construct validity by employing various data sources while safeguarding against research bias. Our data collection encompassed both quantitative and qualitative approaches, i.e., questionnaires, structured interviews, observations, video, and audio recordings. The large number of participants contributed to a substantial data sample, allowing for statistical analysis. However, we sought confirmation within qualitative data before formalizing the results, considering the complexity of the subject (Conway & Lance, 2010). By doing so, we ensured robustness and improved data interpretation. To reinforce construct validity, a great deal of effort was put into minimizing subjectivity during the research design and data collection phases (Jordan & Troth, 2020). Another action we took to reinforce validity was to share our work with researchers who were not directly engaged in this thesis but who were part of the research group, and they provided us with feedback based on their expertise.

Based on the above, we formulated and tested our hypotheses over a span of three years. We initiated a pilot study, employing a limited sample to evaluate our hypotheses and questionnaires and then we analyzed the results employing appropriate statistical techniques, such as PLS-SEM and correlation analyses, to evaluate the relationships between the variables. To be more precise, our analysis involved an examination of various factors. We assessed the Composite Reliability (CR) and Cronbach Alpha (CA) values at the construct level, in addition to the AVE values. To ensure the establishment of discriminant validity, we employed two distinct methods. Initially, we applied the Fornell–Larcker criterion, which mandated that the square root of each construct's AVE should exceed the highest correlation with any other construct (MacKenzie et al., 2011). Subsequently, we investigated whether the outer loading of each indicator surpassed its cross-loadings with other constructs. The outcomes of our analysis affirmed the validity of the reflective measures, underscoring the efficacy of all the items as indicators for their respective constructs. Moving on to the validation of the structural model, we utilized several metrics. These included coefficients of determination (R2), predictive relevance (Stone-Geisser Q2), and the effect size of the path coefficients. To ascertain the statistical significance of our estimates, we adopted a bootstrap approach involving 10,000 resamples. This approach was underpinned by t-statistics, which provided a robust foundation for our conclusions.

3.4.2 Internal Validity

Internal validity refers to the accuracy of the results when it comes to reflecting the truth within the study, which is critical in research methodology (Pannucci & Wilkins, 2010). In the context of this study, our approach involved interviews and supplementary data sources, including reports and internal documents, i.e., we used triangulation across various sources. Given the limited empirical data available on the operational mechanisms applied by firms, we adopted an exploratory, comparative case study approach using NVivo and axial coding to group the comments and observations, which allowed for better interpretations (Charmaz, 2014). This approach boosted the generalizability of our findings for cross-case analyses (Ramesh et al., 2017), while at the same time, we drew from established guidelines (Baskarada, 2014; Eisenhardt, 1989; Stewart, 2012) to conduct our multiple case study. To ensure the credibility and trustworthiness of our research process and findings, we used the dimensions of credibility, dependability, transferability, and confirmability, as suggested by Korstjens and Moser (2018) and Sikolia et al. (2013). To ensure a robust quantitative analysis, we leveraged a structural equation model (PLS-SEM) to evaluate the hierarchical research models (Ringle et al., 2015). PLS-SEM's versatility in assessing the relationships between constructs, reflective or formative, aligns with the complexity of our study, analyzing direct and indirect effects, thus enriching our evaluation of the interrelationships between our constructs (Hair Jr et al., 2017). By employing this approach, we aimed to secure our dataset's credibility and ensure that our findings were not compromised by methodological errors.

3.4.3 External Validity

External validity involves assessing how applicable the study findings are to diverse settings. This is important because the confirmation of external validity indicates that the conclusions can be broadly extended to similar individuals or populations. To ensure external validity we sampled Scandinavian/Nordic countries (see MRP4 and MRP5), which, according to the Global Economic Forum's 2019 Global Competitiveness Report, exhibit extensive ICT adoption and proficient digital skills, making them well-prepared for digital transformation (Schwab, 2019). The case selection process was based on shared characteristics, such as industry sectors, the utilization of AI systems, development team sizes, and cultural contexts. The chosen firms operate within comparable sectors and have similar capabilities in data collection, analysis, and interpretation for decision-making. Another common factor for these firms is the necessity for AI development, expansion, and adoption in the coming years in order to maintain their competitive edge. We also sampled companies from Western Europe and the USA (see MRP6) for similar reasons. Furthermore, the similarity in challenges faced by these firms in AI projects necessitates similar solutions. Therefore, the rationale for selecting these selected companies is threefold: (1) their geographical proximity, (2) the comparable size and experience of their AI teams, despite variations in company sizes, and

(3) their limited cultural disparities, or at least not fundamental differences when it comes to MRP6. By selecting firms from similar industries, we can compare the cases for commonalities and key differences and spot how AI Governance has been implemented. This approach enhances the overall external validity of the study by offering insights that could be broadly applicable and beneficial beyond the specific cases examined.

3.4.4 Reliability

To ensure the reliability of our studies, we took several steps in our literature reviews by following a systematic approach. We applied inclusion and exclusion criteria to set boundaries for the reviews. Included were studies focusing on AI/RAIG's business value or its adoption and use in organizations. Technical aspects like infrastructure or model benchmarking were excluded and we only considered publications written in English from 2010 onward, as significant organizational AI usage emerged in the past decade. Our reviews covered journal articles and conference proceedings, leaving out books, dissertations, reports, and non-peer-reviewed publications. We continued by constructing search strings. Two keyword sets were formed: one related to AI (for MRP1) and RAIG (for MRP2), the other to the business perspective (for MRP1) and organizational perspective (for MRP2). These were combined using wildcard symbols to streamline the searches. We used Google Scholar, Scopus, Business Source Complete and more to ensure comprehensive coverage. This one-month collection process per review was supplemented by a targeted search in AIS journal baskets using the same strings. After eligibility checks, two co-authors independently assessed paper quality based on scientific rigour, credibility, and relevance. Rigour involves appropriate research methods, credibility, evaluation, and presentation of findings, while relevance refers to the significance for the academic community. The remaining papers were used for data extraction and synthesis. To check the whole process, see Figure 2. After that, a concept matrix categorized and synthesized the study findings. Papers were analyzed, and their information was organized in a spreadsheet for easier cross-study comparisons and the interpretation of higher-order insights. The extracted data was checked for research methods, key definitions, analysis levels, findings, theories, and other important concepts. Data extraction was conducted by two co-authors based on the matrix, with all co-authors contributing to the categories and additional dimensions. Based on all the above, the synthesis for each literature review took place.

3.4.5 Ethical Considerations

In this thesis, it is important to note that we did not collect any sensitive data, such as health records. However, we took stringent measures to ensure the security of the information we gathered. For instance, all recorded interviews were securely stored on the NTNU server. Additionally, we obtain informed consent from participants, ensuring they understand the study's purpose and the procedures involved. To preserve the confidentiality of the participants, we implemented strategies to prevent survey data from being linked back to individual respondents. It is worth noting that we had applied to the Norwegian Centre for Research Data (NSD) (known now as SIKT) and received approval for our data collection. In our pursuit of a comprehensive sample size, we aimed to foster diversity and inclusivity by engaging companies across various countries for our survey and involving interviewees from diverse departments. However, we acknowledge certain limitations within our dataset. Specifically, our interview data originates solely from the energy sector, while our survey respondents predominantly represent regions within Western Europe and the United States. Recognizing these constraints is crucial as they may potentially influence the breadth and generalizability of our findings. These precautions show our commitment to both ethical research practices and personal privacy.

4 Results

In this chapter, the primary discoveries of the thesis are presented. To facilitate our analysis, the outcomes are arranged according to each specific research question and its corresponding research objectives.

4.1 Organizational factors of AI in Business – RQ1

4.1.1 Organizational framework of AI in Business – RO1.1 (MRP1)

To evaluate AI's impact on business value, we have categorized AI into three interconnected levels, as illustrated in Figure 3 (see MRP1). Within this organizational framework, we demonstrate that several critical factors, encompassing technological readiness, organizational considerations, and environmental factors, significantly influence an organization's capacity to implement and leverage AI. As a result, we establish two overarching categories for the utilization of AI in organizations and provide a summary of the current knowledge pertaining to applications within these categories. Additionally, we distinguish between the effects of AI, categorizing them as first-order effects and second-order effects, which manifest themselves at the process and organizational levels, respectively. Consequently, we posit that it is crucial initially to examine second-order effects by tracing them back to their underlying first-order effects. For more information, refer to MRP1.

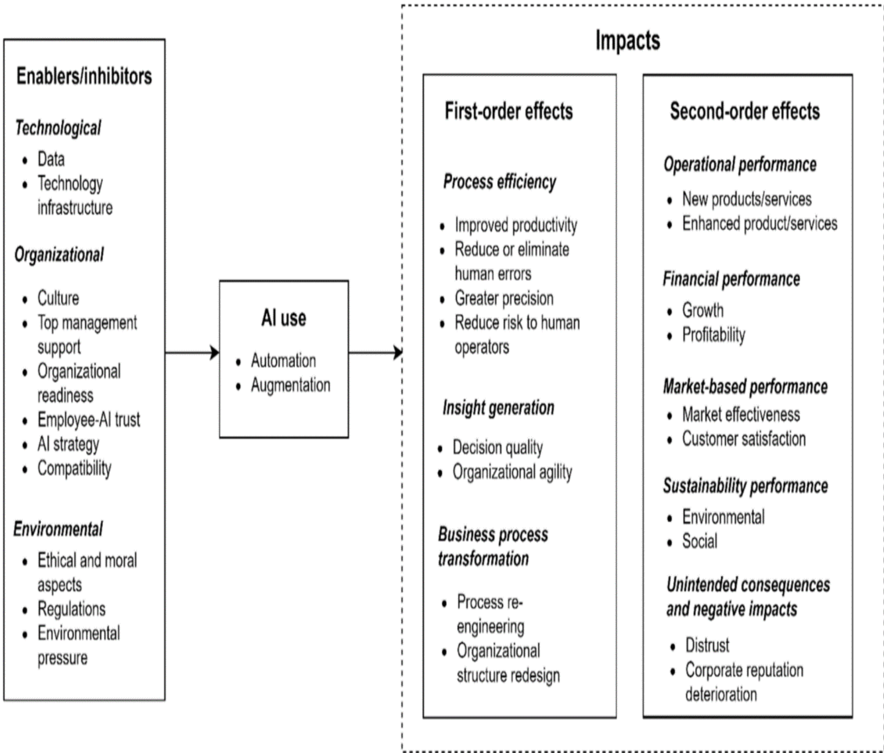

Figure 3: Organizational framework of AI and business value.

The framework sees data as the bedrock of AI development, where AI systems learn from large datasets and make decisions based on these datasets rather than explicit rules set by human experts (Pumplun, 2019; Schmidt et al., 2020). Access to data and the quality of data are critical enablers of AI adoption because large volumes of training data are essential for high-quality AI applications (Afiouni, 2019). However, a common challenge is the scarcity of adequate training data (Baier et al., 2019). Data should exhibit the three Vs: volume, velocity, and variety, a concept commonly referred to as "big data" (Mariani & Wamba, 2020); thus, data quality is paramount for reliable predictions (Alsheibani et al., 2020; Demlehner & Laumer, 2020b). The principle of "garbage-in, garbage-out" underscores the importance of high-quality training data (Lee et al., 2019), where incomplete data, incorrect entries, and noisy features are common quality issues.

Recognizing and addressing these problems requires collaboration between data scientists and domain experts (Baier et al., 2019). To exploit AI's potential, organizations have to possess the appropriate technological infrastructure. This includes computing power infrastructure, AI algorithms, and rich datasets (Wamba-Taguimdje et al., 2020). AI algorithms build models based on data, often requiring substantial computing power (Baier et al., 2019). Many organizations lack the resources, leading to the emergence of cloud-based solutions provided by large companies like Google, Amazon, and Microsoft (Borges et al., 2021; Schmidt et al., 2020). The organizational context plays a pivotal role in AI adoption, with culture as a critical factor (Mikalef & Gupta, 2021; Pumplun, 2019). A culture that fosters innovation and a willingness to embrace new technologies is more likely to succeed in AI adoption (Lee et al., 2019). Such cultures support employees' ability to adapt and use AI applications. Top management support is a strong determinant of AI adoption (Alsheibani et al., 2020) because top managers play a pivotal role in establishing the organization's culture and can allocate resources for AI adoption.

Organizational readiness involves financial resources, skills, and expertise. AI adoption usually requires a significant budget and skilled employees proficient in AI technologies alongside domain expertise. Employee-AI trust is essential as AI systems may change employee roles and responsibilities. Building trust between employees and AI is complex but crucial (Makarius et al., 2020), and AI strategies should align AI adoption with organizational goals. These strategies outline the processes, plans, and timeframes for AI implementation, necessitating structural modifications and data governance (Mikalef & Gupta, 2021), where compatibility between AI and business processes and business cases is crucial (Pumplun, 2019). Business cases should articulate the problems AI aims to solve and how algorithms will improve processes. Business processes need to adapt to AI's requirements since AI adoption is also influenced by the external environment.

Ethical and moral aspects are central considerations (Baier et al., 2019) and are part of the framework. AI should be developed based on ethical principles and without embedding biases (Alsheibani et al., 2020). Transparency, bias, and discrimination are challenges in AI development that need attention. Regulations such as the GDPR have implications for data usage and anonymization, making AI adoption more complex (Baier et al., 2019; Pumplun, 2019) but these regulations are not enough by themselves because they do not address the overall challenge of AI adoption but only part of it. Intellectual property and industry-specific requirements can also impact AI adoption (Dwivedi, 2021).

The adoption of AI is significantly driven by environmental pressures, which compel organizations to adapt and innovate swiftly in response to their rivals (Demlehner & Laumer, 2020a). To maintain or gain a competitive edge, organizations need to actively reconfigure and adapt to a rapidly changing environment. The fear of losing their competitive advantage serves as a strong motivator for organizations to embrace IT innovations, including AI (Alsheibani et al., 2020). Additionally, there is a substantial demand from the customer side, as customers increasingly seek personalized services and products (Pumplun, 2019), exemplified by Amazon’s recommendation engine.

4.1.2 Impacts of AI in Business – RO1.2 (MRP1)

Business executives are keenly interested in how AI can enhance competitive performance. To address this, we have to examine AI's effects at both the process and firm levels. First-order impacts pertain to changes in organizational processes, which are typically measured by key performance indicators (KPIs) like efficiency, effectiveness, productivity, and quality. AI's influence on processes is categorized into three effects: **(1) Process Efficiency.** AI automation and the augmentation of tasks improve productivity and quality by reducing human error, increasing speed, and freeing up employees to focus on higher-value activities. For instance, AI can automate visual recognition tasks in manufacturing, enhancing efficiency and reducing errors (Demlehner & Laumer, 2020). **(2) Insight Generation.** AI reveals hidden patterns in vast data sets, aiding better-informed decision-making and faster responses to market dynamics. It empowers organizations to make more precise decisions by uncovering valuable insights (Lichtenthaler, 2019). **(3) Business Process Transformation.** AI enables organizations to innovate and overhaul their processes, often leading to reengineering and organizational restructuring. It introduces new skills, changes job roles, and allows resource reallocation (Makarius et al., 2020).

Second-order impacts concern firm-level effects and can be categorized into four areas: **(1) Operational Performance.** AI can introduce new products and services, identify market opportunities, and enhance the quality of existing offerings, contributing to business growth (Mishra & Pani, 2020; Davenport & Ronanki, 2018). **(2) Financial Performance.** Companies implementing AI have experienced increased revenue and cost reduction, though further research is needed to explore other financial metrics (Mikalef & Gupta, 2021). **(3) Market-Based Performance.** AI enhances marketing effectiveness by enabling precise customer segmentation and personalized marketing strategies, as well as improving customer satisfaction by preventing negative experiences (Afiouni, 2019). **(4) Sustainability Performance.** AI plays a crucial role in environmental sustainability by minimizing energy consumption and reducing pollution. It also supports circular economy strategies, promoting recycling and reduced emissions (Toniolo et al., 2020).

In the social domain, AI presents challenges related to privacy and discrimination. Organizations need to ensure data privacy and mitigate discriminatory outcomes while taking advantage of AI's ability to reduce human bias in processes like recruitment and customer segmentation. Additionally, AI can enhance employee safety and working conditions by automating hazardous tasks and allowing employees to focus on more meaningful and creative work (Toniolo et al., 2020). For more information, refer to MRP1.

4.1.3 Ethical and Social Implications of AI in Organizations – RO1.3 (MRP4)

The interviewees provided insights into the impact of AI in the context of trading. They expressed a mix of optimism and concerns regarding AI's role in their daily work and its future implications for their careers. Expectations for AI technologies were high, with the hope of substantial benefits. However, in practice, the transition to AI trading was not without its challenges. Problems arose during production, often due to missing or incorrect data, leading to system underperformance and unmet expectations.

Not all AI projects proved to be profitable, and some tests even yielded negative results, causing frustration among employees. The organization's decision to develop its AI systems had both advantages and disadvantages. On the one hand, it allowed for tailored AI solutions, but on the other hand, the lack of adaptation in unexpected scenarios led to costly mistakes. Intraday trading increasingly relied on high-speed AI agents, reducing the need for manual trading, which presented challenges related to deskilling. As traders performed less manual work, their expertise and competence were at risk. To adapt to this changing landscape, the company sought to use AI for heavy lifting tasks while keeping manual trading for reassurance. The future of traders' roles remained uncertain, with potential outcomes ranging from transitioning to other positions within the company to potential job loss in a technology-dominated market. The frequency of trade interactions in the energy market was expected to change significantly, with transactions happening every 15 minutes rather than hourly.

Respondents expressed the need to automate trading processes as much as possible, raising concerns about job security. While some believed that human involvement remained necessary, others anticipated a shift towards monitoring and control tasks. Responsibility for AI outcomes was a point of contention. It was challenging to determine whether AI developers, who built the software, or traders, who mainly monitored AI, should be held accountable. This ambiguity was exacerbated by the lack of explainability tools, making it difficult to address concerns regarding transparency and bias. Legal complications and violations related to AI-generated trading patterns further complicated the issue. The organization's lack of clear roles for AI-related tasks raised questions about responsibility in the case of issues. The importance of addressing these concerns led to the promotion of robustness and reliability through standardized processes and infrastructure. Managers had varying views on the allocation of responsibility, further highlighting the blurred boundaries in this regard. AI trading outcomes raised questions about trust, especially when compared to traditional methods. Inconsistent results and the need to provide predictions even with outdated data sometimes led to difficulties in explaining AI's decision-making.

To build trust in AI, the company developed tools that allowed human controllers to evaluate and decide on AI-produced outcomes. Communication between departments, especially the exchange of domain-specific terms between traders and AI developers, was a minor but potentially significant challenge. Conflict and differing perceptions about AI's effectiveness in trading were apparent among traders. Some resisted accepting AI outputs, believing that the models could not consider all aspects of the market. Perceptions of AI varied from being seen as a magical problem-solving tool to concerns about unrealistic expectations. Lastly, potential conflicts between traders and managers emerged, particularly in the balance between robustness and reliability provided by AI and traders' desire to maintain manual trading. Consumption and energy prices played a significant role, in influencing market behaviour. While AI offered the benefit of minimizing high-risk decisions, it was vital for the organization to ensure a smooth transition to AI adoption without losing employees' trust and loyalty. Unrealistic expectations from both traders and developers posed challenges in aligning AI capabilities with the organization's needs. In conclusion, AI's introduction into trading has brought a mix of benefits and challenges, impacting the nature of work, responsibility, and inter-departmental relationships. As the energy market evolves, addressing these challenges and building trust in AI's capabilities will be crucial for the organization's long-term success and for the employees facing changing roles and uncertainties in their careers. For more information, refer to MRP4.

Table 9: Themes, observations, and nodes for the dark side of AI trading.

Themes	Observations	Nodes
Nature of work	Deskilling	Individual
	AI false expectations	Organizational
	Unemployment	Social & environmental
	Mobilizing human capital	Organizational
	Losing interest in work	Individual
	Hacking attempts	Organizational / Social & environmental
Responsibility	Lack of AI decision explainability	Organizational
	Absence of AI accountability	Organizational / Interpersonal
	Manipulating the market	Social & environmental
Conflicts and effects	Portfolio risks	Individual
	“Enforcing” patterns / overconsumption	Social & environmental
	Conflict of interest between managers and traders	Interpersonal

	Selling overseas / lack of energy	Social & environmental
	Conflicts among AI developers and traders	Interpersonal
	Conflicts among AI traders and non-AI traders	Interpersonal

4.2 Mechanisms and Capabilities of AI in Business – RQ2

4.2.1 Mechanisms of AI use – RO2.1 (MRP1)

The applications of AI encompass a wide range of areas, including marketing, production management, enterprise management, and customer service (Alsheibani et al., 2020; Jelonek, 2020). AI's impact spans an organization's value chain, promising transformative changes in various aspects of daily life (Wamba-Taguimdje et al., 2020). AI applications can be broadly categorized into two main types: AI for automation and AI for augmentation. Automation involves the use of AI to replace human tasks, while augmentation enhances human intelligence by providing valuable insights for decision-making. For more information, refer to MRP1.

Automation, driven by recent advances in AI, enables machines to perform complex cognitive tasks, such as learning and problem-solving, often referred to as "Intelligent Automation" (Welling, 2019). AI technologies are capable of automating tasks that were previously considered too challenging, including knowledge and service work (Coombs et al., 2020). Examples include using virtual robots to process emails in industries like manufacturing and construction (Wamba-Taguimdje et al., 2020) and employing chatbots in the credit card insurance industry to address customer inquiries, process claims, and sell products (Nuruzzaman & Hussain, 2018). AI also contributes to creating new products and services, automating tasks for customers. For instance, conversational intelligent agents like Siri and Alexa automate tasks such as sending messages, making calls, and controlling smart home devices through voice commands (Castillo et al., 2021). Facial recognition in smartphones automates user authentication, exemplifying the diverse applications of AI in automation.

Augmentation involves using AI to enhance human decision-making by processing vast amounts of data beyond human cognitive capabilities (Schmidt et al., 2020). Predictive analytics, for instance, can help managers gain insights from data to make informed decisions, such as identifying new management control indicators and recommending corrective actions (Bytniewski et al., 2020). AI is invaluable in analyzing opinions, attitudes, and emotions related to products or services, providing detailed insights into how customers perceive offerings (Bytniewski et al., 2020; Davenport et al., 2020; Jelonek, 2020). In healthcare, machine vision aids physicians by processing MRI images to detect tiny hemorrhages, detect cancer patterns, or assist in complex surgeries (Jarrahi, 2018; Makarius et al., 2020). AI is used in public relations to monitor social media and predict media trends, as well as in marketing for customer segmentation and lifestyle-based classification (Galloway & Swiatek, 2018; Mishra & Pani, 2020). It also predicts customer habits, anticipates future trends, and optimizes recommendation systems in the fashion industry (Wamba-Taguimdje et al., 2020). Additionally, AI enhances customer intelligence by offering personalized recommendations for products and services. For example, Netflix's recommendation engine uses various customer data parameters to provide tailored content recommendations, increasing the likelihood of customers choosing the content they will enjoy.

4.2.2 Creation of new capabilities to add both Business Value and Customer Value – RO2.2 (SRP1)

Resource orchestration is a crucial process for organizations, encompassing the structuring, bundling, and leveraging of resources, with the aim of creating new capabilities and adding value to both the business and customers. In an uncertain environment, such as one with high competition and regulatory changes, resource orchestration plays a pivotal role in maintaining and enhancing a firm's competitive advantage. For AI-related resources, this process involves acquiring, accumulating, divesting, stabilizing, enriching, pioneering, mobilizing, coordinating, and deploying. For more information, refer to SRP1.

In acquiring subprocesses, firms purchase resources from the market, such as high-computation infrastructure or skilled AI personnel. The acquisition should align with the organization's strategic goals, but in a highly uncertain environment, it can be costly and necessitate cautious decision-making. In accumulating subprocesses, firms should internally develop and accumulate knowledge and the capabilities related to AI, especially in environments characterized by high uncertainty. This tacit knowledge and expertise become essential when dealing with rapidly changing markets and unforeseen challenges. In divesting subprocesses, to optimize AI resources, companies need to identify and release underutilized or outdated resources, whether they are data sets, technology, or personnel. In uncertain conditions, careful consideration of the importance and long-term relevance of these resources is vital. Bundling resources involves the combination of various resources to create capabilities, and it encompasses stabilizing, enriching, and pioneering processes. Stabilizing subprocesses involves making minor improvements to existing capabilities maintaining a competitive edge over time. However, in highly uncertain environments, where significant changes are required, stabilizing may not be a feasible option. The enriching subprocess extends existing capabilities through resource addition, and skill development is critical, especially when acquiring new resources is costly or challenging. Collaborations and alliances with external partners can also enhance AI initiatives. The pioneering process is exploratory in nature, integrating radically different resources to develop new value propositions. This is particularly important in uncertain environments to gain a competitive edge. Leveraging capabilities aims to mobilize, coordinate, and deploy processes for value creation. High-level managers play a crucial role in these processes, aligning AI initiatives with business goals, coordinating resources efficiently, and deploying strategies in line with the environmental context. Mobilizing a subprocess identifies the capabilities required for AI development, which is crucial, but not sufficient, for maintaining a competitive advantage, particularly in uncertain environments. Coordinating a subprocess integrates AI capabilities smoothly with other organizational capabilities, which is essential because high-level managers are responsible for facilitating this integration through effective communication and collaboration. Deploying capability configurations depends on the leveraging strategy chosen and the environmental context. The success of deployment can be challenging, especially when external partners are involved.

4.3 Antecedents and effects of AIG in Organizations – RQ3

4.3.1 Challenges, barriers, and solutions for adopting AIG – RO3.1(MRP3)

A challenge for AI governance is the unification of technologies and infrastructure, which is essential for compatibility among diverse AI tools. The need for increased speed and scalability has been driven by growing data volumes, leading to efficiency gains and automation. Additionally, fostering an AI culture within the organization is crucial for employees to embrace and trust AI, as a lack of understanding can lead to resistance during the digital transformation process. Other inhibitors include insufficient domain knowledge, data limitations, and legal regulations that restrict certain AI applications. The outcomes across various firms largely align with their shared goals. Key priorities include reducing maintenance costs and forecasting energy consumption, both contributing significantly to business value. Flexibility and robustness in AI systems have become crucial, allowing adaptation to market trends while ensuring non-costly and reliable customer experiences. Table 10 shows the challenges, recommended actions, and desired outcomes.

Table 10: Challenges, recommended actions, and desired outcomes.

	Challenges	Recommended Actions	Outcomes
Development	AI cloud is challenging to build	Offline recommendation system Develop intelligence on top of external platforms	Boost flexibility
	AI development does not necessarily follow traditional software development	Standardize executable components Unify technological tools Create shared libraries	Robustness Reduce amount of workload
	Prediction techniques vary based on sector	Allow human interaction in high uncertainty to prevent high AI bias.	Robustness
	Lack of data	Choose AI algorithms based on data volume and data types Generate data from existing data Read data from different sources Buy data from vendors using APIs	Boost flexibility Robustness
	Lack of domain knowledge by AI developers	Allow domain experts to lead	Save money and time Robustness
Employees	Misunderstanding of AI capabilities	AI training to understand what the models can do and what they cannot do	Better communication between departments Easier adoption of AI
	Employees do not adopt AI	AI training to understand how to use the new technologies	Better communication between departments Easier adoption of AI
	Employees fear losing their position because of AI	AI training to explain why their expertise cannot be replaced	Better communication between departments Easier adoption of AI
	Different vocabulary for different departments	AI training to be familiar with different	Better communication between departments

		terms and processes. Create different dashboards for different concepts	Easier adoption of AI Measure performance
Value	Classical optimization tools are still better than AI models	Automate operations: 1. that take place 24-7 2. where there is a 1-1 correlation between workload and number of employees 3. which are repetitive and boring	Save money and time Scaling up becomes easier Reduce amount of workload
	Hard to predict effort and costs	Avoid nice to have features as they will delay the whole process considerable Use KPIs to quantify performance	Save money and time Scaling up becomes easier
External Environment	Giving out knowledge to external partners	Develop intelligence on top of external platforms instead of using external solutions	Maintain competitive advantage
	Distance from third parties can affect development	Develop internal AI team to speed up processes considerably	AI Development is focused on your specific problem not on a generic solution maintain competitive advantage
	Legal constraints and GDPR	Create clear data management roles	Security

Figure 4 (see MRP3) presents a model that encompasses structural, procedural, and relational components, demonstrating the techniques firms have employed in recent years. Enablers include the presence of an AI culture and suitable architecture within the company, while inhibitors primarily consist of legal constraints, domain-specific challenges, high development costs, and AI-phobia. Companies looking to integrate AI are advised to address these challenges proactively to prevent potential failures and resource wastage. The ultimate outcomes sought by firms include a competitive edge, cost reduction, and the establishment of dependable AI systems, which are vital for success in competitive markets. For more information, refer to MRP3.

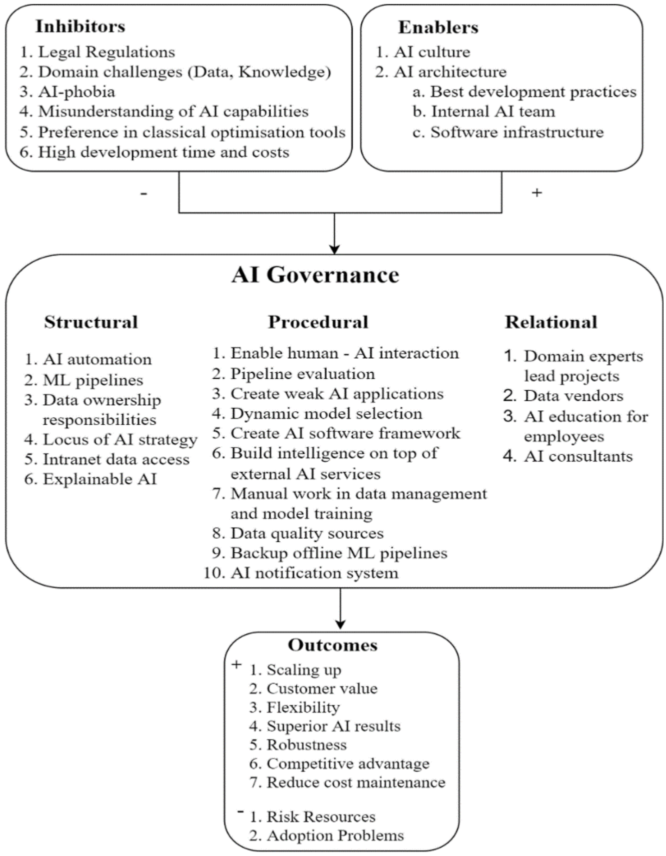

Figure 4: Model for AI Governance.

4.3.2 AI structures and implementation – RO3.2 (MRP3)

In terms of procedural practices, all the companies from our multi-case analysis (see MRP3 for more details) pursued the development of new capabilities through the use of external software. This involved the internal AI teams creating algorithms, trading strategies, and machine learning pipelines. They leveraged platforms from third-party partners while retaining their domain knowledge in-house. Their approach aimed to maintain control and prevent third parties from repurposing the same software for various applications. For every project, common elements included a strong focus on data governance, data quality, and data security. Instead of extensive data cleaning, these firms preferred to rectify data source issues through data collection corrections and the integration of APIs to ensure data accuracy. The continuous evaluation of machine learning pipelines was integral to ensuring their robustness and quality. The resulting AI products were categorized as "weak AI," primarily designed for executing specific tasks or offering decision-making suggestions, often in the form of AI assistants. However, end-users were required to follow AI suggestions intuitively and complement them with their domain knowledge. Moreover, these intelligent systems incorporated features like notification systems, error detection, and decision-making tools to assess system credibility and performance using key performance indicators (KPIs).

Regarding their practices, AI strategy played a central role for top managers, who were responsible for designing products tailored to specific needs while adding business value. They needed to allocate resources precisely and plan rigorously, recognizing that AI projects diverged from the conventional software development timeline and cost expectations. Managers faced the challenge of distinguishing essential features from nice-to-have requests, as the latter could significantly delay projects and inflate development costs. It is essential to note that AI development was generally more costly than traditional software development. Managers could estimate the effort required to build a pipeline based on project specifications. The practice of reusing components from one pipeline for another was common in AI projects, substantially reducing development time while instilling confidence in the final product's quality, maintainability, and extensibility for new features.

In terms of data management practices, a shift was observed towards securing data, using secure databases, and creating distinct roles for data access. Typically, these roles included developers who had full data access and end-users who had limited data access. In all cases, domain experts played a pivotal role in all development phases. Their domain knowledge was critical to project success, and they often took on project management roles. With the collaboration of AI developers, they designed notification systems to determine which notifications should be delivered via email or displayed in dashboards. External AI consultants were primarily engaged at the project's outset, specifically when the development team faced uncertainties regarding particular aspects of the project, such as cloud services or ML optimizations. Lastly, establishing an AI culture within the company through comprehensive training proved challenging, particularly in the initial stages. Employees exhibited mistrust, sometimes seeing AI recommendations as naive. Many employees viewed AI and automation as threats to their positions, leading to the need for workshops and internal meetings to address these concerns and dispel fears. The communication focused on emphasizing the point that AI aimed to assist employees rather than replace them, helping build trust and understanding.

4.4 Relationship between RAIG and Competitive Performance – RQ4 (MRP5)

Our literature reviews have illuminated the diverse array of approaches available for investigating the impact of RAIG. In our research, we have intentionally focused on a specific pathway that serves to examine two key dimensions. First, we explored how RAIG, particularly when channeled through KMC and strategic alignment, can significantly influence competitive performance within organizations. Second, we explored the ways in which RAIG, with an emphasis on legitimacy and effective communication strategies, can boost overall firm performance to new heights. This chosen research path allows us to dissect the relationship between RAIG, strategic management practices, and organizational performance, shedding light on the mechanisms that underpin these relationships.

4.4.1 Relationship between RAIG and competitive performance through KMC – RO4.1

We provide empirical evidence supporting the idea that RAIG influences KMC and, in turn, indirectly impacts a company's competitive performance. The research, based on a large sample of Scandinavian companies, emphasizes the need for businesses to consider how RAIG can affect their performance outcomes. In more specific terms, RAIG is found to directly enhance a company's KMC by expanding its knowledge assets and operational capacities, thereby improving its competence and capabilities. This, in turn, contributes to better competitive performance. Interestingly, the study did not find evidence to support the assumption that strategic alignment has a significant impact on KMC. This discrepancy may be because responsible AI implementation often begins with technical teams at the operational level, potentially conflicting with top-down management processes. To successfully integrate responsible AI practices, managers should first comprehend the necessary steps and requirements. Establishing an AI governance framework is crucial to restructure organizational processes and support responsible AI initiatives, which can be a substantial undertaking without proper planning and change management efforts. For more information, see MRP5.

4.4.2 Relationship between RAIG and competitive performance through Legitimacy – RO4.2 (MRP6)

This study looks at how RAIG practices, communication about using AI responsibly, legitimacy, and company performance are connected. Our model is illustrated in Figure 5 (see MRP6). We collected data from 329 employees in companies across Western Europe and the USA and used a method called PLS structural equation modeling to analyze it. We used an online survey to look at how RAIG practices affect a company's performance and legitimacy. The results show that RAIG practices directly improve performance and legitimacy, as we expected. The results also indicate that companies can improve their performance by gaining trust from stakeholders, who are more likely to engage with the company's products, services, and practices. The model also shows evidence that communicating about RAIG practices is important for gaining legitimacy, both inside and outside the company. However, we did not find any significant effect of communication on the relationship between RAIG practices and legitimacy, which was unexpected. In short, the findings show that RAIG practices are crucial for getting support from both internal and external environment and for improving performance. This study is the first, to the best of our knowledge, to show how RAIG, legitimacy, and performance are linked, highlighting the importance of using RAIG strategies and good communication to keep a company's legitimacy intact and achieve positive results in business.

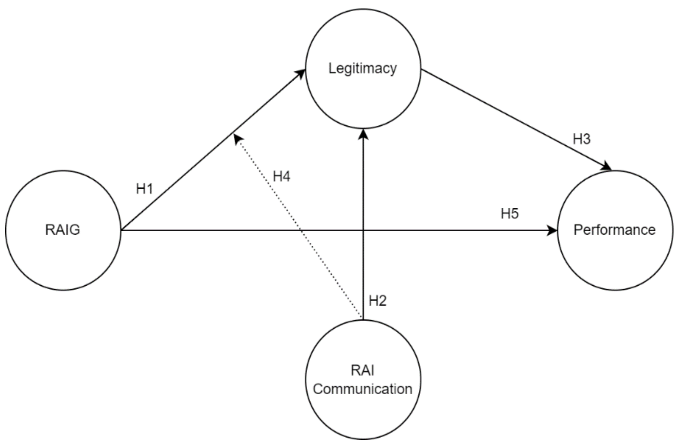

Figure 5: RAIG and competitive performance through legitimacy.

5 Discussion

In this chapter, the research results presented in the previous chapter are compiled, examining their impact on the research, practical applications, and policy considerations. Furthermore, it addresses the constraints and outlines potential directions for future research.

5.1 Theoretical Implications

The current thesis offers novelty regarding the role of RAIG practices in combination with legitimacy and the impact on firm performance. This inquiry makes a significant contribution to expanding the conceptualization of RAIG beyond operational and ethical dimensions, highlighting its role in bestowing legitimacy upon organizations with RAIG practices. It emphasizes the importance of organizational legitimacy in AI governance and underscores how RAIG practices can increase perceived legitimacy, thereby impacting organizational performance (Tseng & Lee, 2014). In doing so, we extend the work of previous researchers who have investigated the impact of corporate social responsibility (CSR) activities on firm performance and legitimacy (Colleoni, 2013; Khuong et al., 2021; Koh et al., 2023; Min et al., 2023) by arguing that firm performance and legitimacy can be increased through RAIG practices too. However, the study also points out a lack of understanding regarding how RAI communications interact with broader AI governance practices. Although the study shows some evidence in order to understand the mechanisms through which RAIG practices contribute to organizational legitimacy and how legitimacy influences performance outcomes, it did not find moderating effects of contextual factors on this relationship which could provide insights into AI governance practices across different organizational contexts (Zhang et al., 2023). Additionally, the study did not take into consideration the effectiveness of different RAI communication strategies and how organizational commitment to RAI practices impacts stakeholder perceptions of legitimacy; thus, extending legitimacy theory to other related disciplines could also increase our understanding of factors influencing AI governance practices and their implications for performance (Jan et al., 2021).

In addition, we uncover the dark side effects of AI and how to mitigate them. Previous researchers focused on bias and fairness when it comes to hiring and criminal justice or ethical considerations when it comes to AI-generated content through deepfakes (Widder et al., 2022; Završnik, 2020). Our research focused on the dark side effects of AI in B2B marketing by examining employee adaptation to new AI technologies and by providing insight into managers' actions. What is more, current B2B state research addresses accountability problems related to algorithmic misbehaviour (Rana et al., 2021), numerous ethical and legal concerns (Boyd & Wilson, 2017). We added to the literature by finding challenges in models learning from new data and stressed the importance of an effective organizational structure for successful AI integration. Also, we found that despite the need for high-quality data, extracting insights remains difficult due to relevance and management issues, while determining responsibility for AI decisions poses a significant challenge (Papagiannidis Emmanouil et al., 2023). Beyond that, we argue that domain experts (traders in our case), though no longer directly involved, remained an important pillar for AI success. The companies we investigated prioritized explainable AI over higher-margin options and emphasized compliance in unregulated areas. Our study also explored the impact of reduced relationship bonds on B2B interactions (Gligor et al., 2021), showing socio-economic effects on client behavior, particularly in energy consumption.

Researchers have investigated the harm inflicted upon the organization itself and the harm inflicted upon others (Sun et al., 2022). We expand the literature by going through organizational aspects, such as procedural changes and managing human capital, which are affected by AI adoption. We argue that managers have to find solutions to address challenges while preserving the firm's public image. We investigated the negative implications of AI in trading on individuals and society, including concerns about AI fear, deskilling, and unemployment (Paschen et al., 2019). We continued by arguing that while adopting AI is crucial, establishing procedures and mechanisms for aligning AI applications with business objectives is equally critical (Puntoni et al., 2021). AI's dynamic nature necessitates recognizing its negative aspects to ensure businesses function as intended. Furthermore, organizations have to consider the dark sides of AI when planning, designing, and building AI strategies and products. AI can lead to value co-creation but may also introduce challenges, ranging from job loss and privacy concerns to machine ethics, security issues, and the development of superintelligence. Effective AI governance is essential for bridging the gap between accountability and ethics in technological advancement (Davenport et al., 2020). This governance should address difficulties encountered during the AI deployment process and provide options and probabilities for addressing complex management tasks.

However, it is important to note that not all companies have successfully developed AI solutions that result in significant organizational impact and added business value. The central argument is that while adopting AI is crucial, it is equally vital to establish the necessary processes and mechanisms to develop and align AI applications with the demands of the business environment (Fadler & Legner, 2021). A key challenge identified in our studies is the dynamic nature of AI governance, which necessitates continuous adaptation and modification in response to evolving conditions, including how employees perceive AI (Min et al., 2023). This dynamic aspect places increased importance on establishing effective processes, mechanisms, and structures to ensure that AI functions as intended and aligns with the organization's goals. Additionally, the research highlights the various approaches that companies take toward AI governance, such as creating ML pipelines and interactive dashboards.

Notably, not all companies prioritize explainability in the early stages, focusing on what they perceive as more urgent priorities. In contrast to articles that primarily focus on the technical aspects of AI workflow implementation, this research emphasizes the development challenges and practical solutions firms can employ to build AI through effective organizational practices. The proposed model (see more at MRP5) underscores how, despite the presence of inhibitors and barriers and the diverse approaches to AI governance, following best practices can yield positive outcomes. We identified the specific procedural, structural, and relational components necessary for achieving this. Therefore, our work sparks a discussion on the composition of AI governance and its potential dimensions. It delves into the connection between governance practices and the challenges they help overcome, involving various actors and practices. This is particularly significant for understanding how AI-based applications generate value by shedding light on how different resources are leveraged in the pursuit of business value. Additionally, we provide insights into the process view of AI deployments, opening up a dialogue about the unique challenges within each deployment phase.

One major challenge that researchers need to highlight is the ever-changing nature of AI governance, which requires ongoing adjustments to ensure AI aligns with the organization's goals (Wu et al., 2015). We identify various approaches that firms take in AI governance, such as creating ML pipelines and interactive dashboards. However, not all of them prioritize explainability, as they are often focused on more immediate concerns. Unlike other articles that primarily focus on the technical aspects of AI implementation, this research emphasizes the practical challenges and solutions involved in building effective AI through organizational practices (Papagiannidis et al., 2021). AI governance, in this context, is viewed not as a single process but as a collection of critical components that have to be considered when designing and deploying AI practices to overcome challenges and achieve successful outcomes. This study presents a model (see MRP3) that suggests that, despite barriers and diverse approaches to AI governance, positive results can be achieved if best practices are followed. It identifies the specific procedural, structural, and relational elements crucial for success. This exploratory work sparks discussions about the composition of AI governance and its dimensions. It also explores the connections between the challenges AI governance helps address and the actors and practices involved in the process. This research is valuable for understanding how AI-based applications generate value and how resources are leveraged. It also provides insights into the various phases of AI deployment and the unique challenges encountered in each phase.

Finally, there are many other paths for research when it comes to AI. We also establish a foundation for guiding prospective research ventures by examining existing assumptions and pinpointing domains with noticeable knowledge gaps. The research framework does not intend to comprehensively enumerate all possible research directions but rather underscores pivotal areas where our comprehension of how AI influences organizational conduct and competition is deficient. Consequently, we outline five research streams, each of which introduces various research pathways to enhance our insight. Figure 6 (see MRP1) visually depicts this research framework, illustrating the themes within enumerated circles.

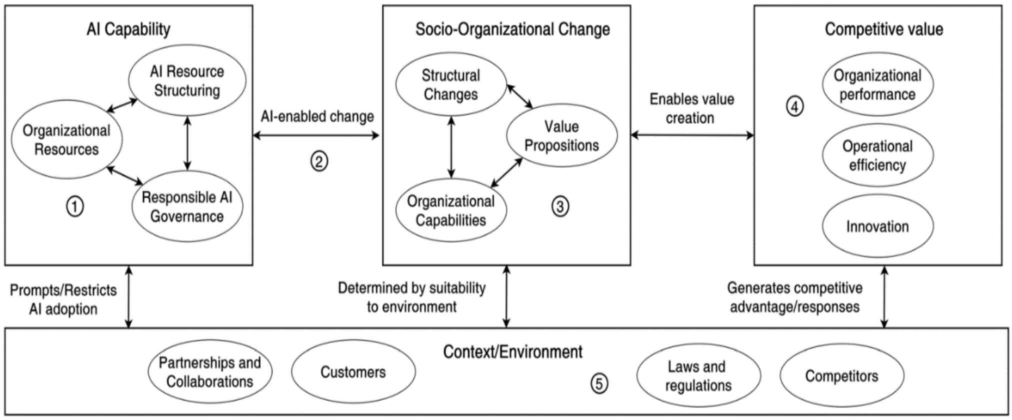

Figure 6: AI and business value framework.

Theme 1: AI Adoption and Diffusion

Difficulties in the process of adopting and deploying AI. Many organizations face substantial challenges when attempting to adopt AI (Hammer & Karmakar, 2021). Technological readiness, organizational preparedness, and external factors such as government regulations play critical roles in the adoption of AI. Additionally, factors like infrastructure costs, talent acquisition, and partnerships can influence the adoption dynamics (Alsheibani et al., 2018). Furthermore, internal conflicts between shareholders and managers regarding automation and augmentation can delay actual AI deployment, and AI adoption may challenge cultural norms, creating barriers for both managers and customers (Dwivedi, 2021).

Governance of AI projects. To achieve the full potential of AI in core operations, organizations need effective governance, resource management, and project oversight, spanning the entire project life cycle (Tallon & Pinsonneault, 2011). AI applications evolve through multiple maturation phases and may require explainability and accountability principles, particularly in areas like the public sector. Understanding the key activities that underpin AI governance is a significant research area, with the potential to optimize performance and alignment between business and technology functions.

Theme 2: AI and Socio-organizational Change

How does AI change organizational culture? The adoption of AI has been associated with fostering innovative organizational cultures. Research is needed to explore how AI may influence culture more broadly, affecting learning, collaboration, communication patterns, and the organization's overall openness to innovation. The "dark side" of AI, including the potential for distrust among employees due to opaque AI decision-making and privacy concerns, should also be investigated in the context of organizational culture.

What is the role of AI-driven automation in decision-making? The automation of decision-making by AI systems can reduce employees' workload but may also introduce challenges (Acemoglu & Restrepo, 2018). These include addressing potential biases in AI decision models, especially when making decisions related to gender, ethnicity, and personal data (Cirillo et al., 2020). AI automation can also impact human interactions, potentially leading to "AI anxiety" (Li & Huang, 2020). Research should explore whether AI automation genuinely benefits decision-makers or challenges their values.

How does AI change the organization's structure? Both the adoption of AI and the organization's structure can influence each other. While organizational structure may affect the ability to adopt AI, AI adoption can result in the reconfiguration of the organizational chart, changes in roles, and shifts in data and information flows. Quantitative research is needed to examine the relationship between organizational structure and AI adoption, as well as how AI impacts organizational structure.

Theme 3: AI-driven Value Propositions

How does the orientation of AI impact value propositions? AI can be leveraged for both internal-oriented functions (improving internal processes) and external-oriented functions (enhancing customer-facing products and services) (Davenport et al., 2020). The research should focus on understanding how the placement of AI in the value chain affects business performance and how organizations should organize themselves to realize value from AI applications.

What is the role of complexity in AI application inimitability and value? Complex AI systems may be harder for competitors to imitate, potentially providing a competitive advantage (Monostori, 2003). Research is required to explore the relationship between the complexity of AI systems and their value creation for businesses. Understanding when complexity leads to value creation is vital for organizations.

Theme 4: Competitive Value of AI

What are the effects of AI on financial performance? Organizations often expect that AI will improve financial performance by increasing revenue, growth, and reducing costs (Alsheibani et al., 2020; Eriksson, 2020). However, there is a need for research to investigate the long-term financial consequences of AI adoption, particularly for small and medium-sized enterprises. Understanding when and how AI applications generate positive financial returns is crucial.

What are appropriate key performance indicators (KPIs) to measure AI success? Measuring the success of AI projects is challenging due to the unique outcomes of AI applications. Research should focus on identifying the appropriate KPIs to assess AI outcomes, especially after AI applications have been deployed and used in practice (Ehret & Wirtz, 2017). These KPIs should be quantifiable and provide insights into the impact of AI on the business.

How can AI drive innovation? While AI technology is behind innovative products and services, the socio-technical dynamics that lead to innovation need to be better understood. The interaction between AI technology, managers, knowledge workers, and their collaborations must be explored in more detail to facilitate technology-driven innovation.

Theme 5: AI and the Extended Organization

Extended organizational boundaries and partnerships. Organizations often engage in various forms of relationships with external partners, such as mergers, acquisitions, joint ventures, and alliances. Understanding how these relationships influence the types of AI applications developed and the nature of organizational engagements is essential. Research should investigate the governance schemes and conflicts of interest in such AI-specific partnerships and explore the optimal ways of organizing boundaries and partnerships (Yang et al., 2018).

What is the role of AI in shaping the reputation of the organization? AI's introduction can affect trust and reputation within organizations. Research should delve into how AI impacts trust and, in turn, organizational reputation. The findings can guide organizations in making informed decisions about AI adoption (Cohen et al., 2020).

5.2 Practical Implications

Our work carries practical implications across different sectors. For managers, our findings emphasize the importance of adopting practices to ensure responsible use of AI that will lead to successful AI products. For AI developers, our findings emphasize the integration process where they need to prioritize explainability tools for their systems in order to enhance accountability and make end users feel more secure when using AI for decision-making. This approach also helps managers ensure employee safety, comply with legal regulations, and create transparency in the decision-making process. Additionally, organizations have to establish the necessary infrastructure to centralize their advanced systems, with a particular emphasis on AI. This should be done in a way that reduces inequality, promotes social empowerment, preserves individual autonomy, and ensures equitable benefits for all stakeholders. The explainability of AI plays a crucial role in building public trust and facilitating a better understanding of the technology, which, in turn, simplifies AI monitoring and allows for the reallocation of employees displaced by AI, thereby protecting their jobs. Furthermore, firms should develop tools for testing AI decisions to uncover new patterns in data, leading to deeper insights. This process encourages employees to generate innovative ideas that can drive productivity and innovation. Back-testing, a common practice for validating results, allows appropriate domain experts totest their hypotheses and refine AI algorithms through their expertise, ultimately enhancing the company's competitiveness. Managers should prioritize these processes to keep their domain experts actively engaged in improving AI systems.

Our research provides empirical evidence supporting the idea that RAIG directly influences KMC, which, in turn, indirectly affects a company's competitive performance. We suggest that pushing for changes in structures and processes may conflict with the organization's existing goals and priorities. Therefore, it is crucial for managers aiming to integrate responsible AI into their operations to first understand the requirements and then take the necessary steps to develop a responsible AI system. In the absence of well-defined AI governance practices, the process of redesigning organizational structures, accommodating responsible AI work, and implementing new management practices can be a substantial undertaking.

That is why structural, procedural, and relational practices are needed. **(1) Structural practices.** Quality data is crucial for successful AI outputs, and preventing data poisoning is essential to avoid manipulation of AI results. Data poisoning can occur in two ways: injecting incorrect information into the system and creating a backdoor for exploitation. These issues highlight the need for RAIG, emphasizing data curation, continuous monitoring, and human oversight to prevent unexpected AI behaviour. Companies should update AI infrastructure based on their specific needs to avoid issues like hardware obsolescence and security breaches. **(2) Procedural Practices.** AI safety research should consider different AI paradigms and their safety implications, as well as anticipate the requirements of future, more powerful AI systems. Designing AI with respect to human autonomy, equality, and social empowerment is essential. While transparency in AI decision-making is valuable, it may also pose security risks, and its necessity depends on the context. Research should focus on defining the degree of transparency needed, the expandability of AI models, and the practicality of transparent algorithms in different applications. **(3) Relational Practices.** Businesses are increasingly recognizing the importance of AI ethics, with cases involving AI-related concerns affecting organizational decision-makers. However, there is a need to overcome obstacles related to the awareness and understanding of AI ethics among managers and non-IT personnel. To foster organizational awareness, research could explore methods to reduce algorithmic aversion, leverage AI for efficient messaging, and align AI ethics with company goals. Managing user expectations and trust in AI is also critical, requiring checks and balances, peer reviews, risk assessments, and alignment with agile working methods. Additionally, promoting inclusion and diversity strategies within AI research can enhance innovation and collaboration, improving overall employee and customer experiences.

The findings emphasize the need for firms to adopt new procedures when integrating AI into their operations to gain a competitive edge and enhance efficiency. It is crucial to establish a unified system for building AI pipelines, aligning it with the tools used by developers. This approach makes the system more robust, easier to maintain, and allows for improvements in various components. Additionally, managers should implement clear procedures that employees can easily understand and follow. Without well-defined guidelines, there is a risk of wasting time and resources that could be better invested in other projects delivering greater business value. Firms should leverage AI for automating repetitive tasks, a change that employees generally appreciate since it frees them from monotonous work. However, it is equally important for managers to engage in extended conversations with employees from various departments. These discussions should reassure them that AI is not a threat to their job security but rather an opportunity for education and upskilling. Fostering this understanding is essential for maintaining internal stability, preserving trust in leadership, retaining talent, and ensuring the smooth adoption of new technologies. Companies can effectively use dashboards as a means of facilitating communication between humans and AI because dashboards serve as valuable tools for managing information, tracking key performance indicators (KPIs), metrics, and other critical data points. This approach helps overcome the black-box nature of AI models by presenting data in a visual format that simplifies complex datasets. It empowers end-users to evaluate results, identify outliers or anomalies in processed data, and promote transparency. Therefore, dashboards enhance the ability to review and adjust data analysis models directly.

In practical terms, AI developers should not underestimate the importance of incorporating explainability tools for AI decision-making. This is vital for ensuring that accountable individuals, such as domain experts, have a clear understanding of decision processes, which fosters a sense of responsibility and confidence in using AI (Paschen et al., 2019). This approach helps managers ensure the well-being of their employees and align their processes with potential legal requirements that mandate transparency in decision-making. Moreover, organizations should establish suitable infrastructure for centralizing their advanced systems, with a particular focus on AI (Al-Surmi et al., 2022). AI, in particular, should be designed to reduce inequalities, promote social empowerment, maintain individual autonomy, and distribute benefits equitably (Puntoni et al., 2021). Ensuring AI's explainability is crucial as it plays a major role in building public trust and fostering a better understanding of the technology (Keegan et al., 2022). This not only simplifies AI monitoring but also allows for the redeployment of employees from roles AI takes over. Lastly, organizations should invest in tools for testing AI decisions to uncover new patterns and gain a deeper insight into data and information (Davenport et al., 2020). This approach can inspire employees to generate innovative ideas, leading to new strategies and tactics that enhance productivity and drive innovation. Much like how domain experts perform back-testing to validate hypotheses, managers should prioritize such processes, enabling domain experts to actively improve AI products through their expertise (Mikalef & Gupta, 2021). This contribution can result in more competent algorithms and, consequently, a more competitive company.

Organizational leaders are encouraged to prioritize investments in RAIG frameworks to drive performance improvements, maintain competitiveness, and strategically manage legitimacy (Abioye et al., 2021). Prioritizing transparency and accountability in AI governance practices can increase perceived legitimacy, positively influencing performance metrics (Du & Xie, 2021). Effective RAI communications are essential for shaping stakeholder perceptions of legitimacy, building trust, and credibility. Organizational leaders should align communication strategies with RAIG practices and organizational goals to enhance legitimacy and performance Soomro et al., 2021). Considering contextual factors and industry-specific dynamics is crucial for building effective strategies related to AI governance (Akram et al., 2018).

5.3 Policy Implications

Policymakers play a pivotal role in shaping the ethical and safe deployment of AI. The processes they are involved in include making regulations, applying standards, and providing guidelines to ensure that AI technologies are developed, used, and managed in a manner that upholds human rights, fairness, transparency, and accountability. Policymakers should provide a framework for addressing challenges related to bias, privacy, and the societal impact of AI, helping to strike a balance between innovation and responsible use. Effective governance and policy making in the AI domain are essential for building public trust, fostering innovation, and safeguarding against the potential harms associated with AI technologies. There are three key factors that affect RAIG. First, societal expectations and norms, which encompass unwritten codes of conduct shared by society and adopted by organizations, play a crucial role in shaping AI perspectives and ethical considerations. Organizations often adjust their operational approaches to align with these norms, aiming to maintain a positive public image. For instance, Google introduced a framework for RAIG, and entities like the European Commission and the Singapore Government have developed guidelines to address the ethical and legal concerns related to AI. Second, from an organizational standpoint, a belief that AI should serve people's best interests rather than the other way around is vital. Concerns about data privacy and security, particularly in AI recommendation systems using big data, are of great significance due to the challenges in explaining AI decisions and the potential impact on individuals' lives. Privacy legislation has to consider societal and political implications, and organizations aiming to establish trustworthiness need to address issues like false positives and negatives, overfitting, and the responsible use of personal information. Third, organizations need the capability to swiftly capture data influenced by evolving norms and adapt their responsible AI principles accordingly. Influencers, celebrities, and social movements can significantly affect public perspectives, requiring companies to stay aware of emerging technologies and societal shifts. Adjusting policies and strategies in response to such changes is essential to maintain a positive public image. Figure 7 (see MRP2) shows the antecedents, structure, and effects of RAIG.

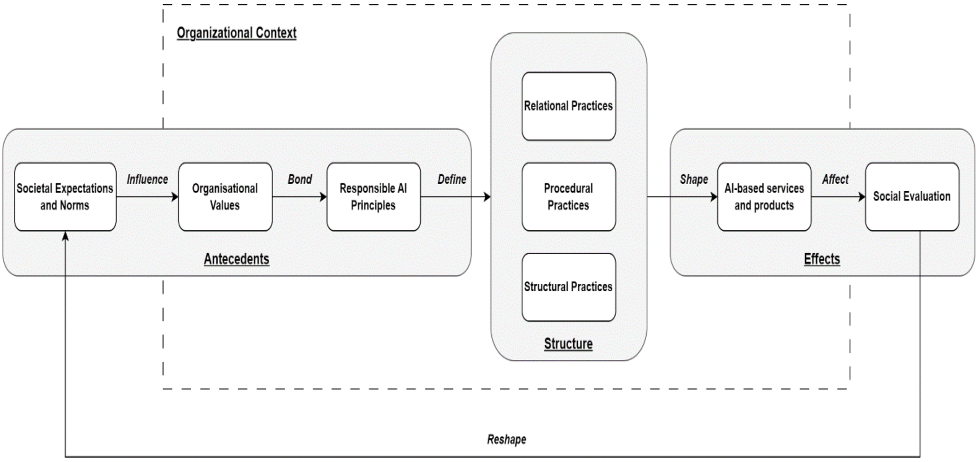

Figure 7: Antecedents, structure, and effects of RAIG.

As far as AI applications and services are concerned, they offer significant opportunities for policy making, business-management, and decision making through data analytics. This is particularly evident in industries like energy, where AI helps optimize resource usage during potential energy crises. Policymaking should take into consideration the fact that responsible AI practices attract investors, but there are also potential negative effects, such as automated trading affecting stock markets and concerns about AI weaponization. These factors could shape future norms regarding AI. Hence, a lack of RAIG can lead to AI anxiety on a business level. It may result in the automation of work activities, job role redefinition, and potential job performance issues, impacting employee satisfaction and commitment. Research is needed to understand how societal norms influence organizational values and how these values affect responsible AI principles. Achieving a balance between automation and augmentation is crucial to avoid job displacement and overreliance on AI. Incorporating end-user involvement in AI development is vital, and cost-effective techniques should be explored to capture user needs. Giving users control over their data in semi-autonomous systems is essential for respecting their autonomy. AI norms vary across countries, so researchers should provide clear and practical definitions of “AI Ethics” based on cultural norms. Dedicated committees and oversight bodies may be established to evaluate the ethical implications of AI projects. Ethical AI applications can promote brands, but addressing bias in ML algorithms is a necessity. Detecting and eliminating the bias associated with social expectations is a complex challenge. Agent-based models and ML-based inference models can help improve sequential decision-making by studying behavioural patterns. Creating bias detection tools and fairness metrics, like AI Fairness 360 (AIF360), is essential for addressing bias in AI systems.

In this work, we focus on examining how AI impacts behaviour and operations within B2B contexts. Several studies (Davenport et al., 2020; Farrokhi et al., 2020; Grewal et al., 2004;

Kushwaha et al., 2021; Troisi et al., 2020) have touched on various aspects related to AI, but none have thoroughly explored the direct implications of AI within the B2B domain. These studies provide valuable insights into AI in broader contexts, laying the groundwork for more focused research in the B2B sector. In response to this research gap, we propose several avenues for future investigation into the dark side effects of AI in B2B settings (see MRP4). Researchers could explore both the positive and negative impacts of AI on B2B operations, considering how AI may influence a company's reputation and market positioning. It is crucial to gain a comprehensive understanding of the potential consequences of AI and how to effectively manage and mitigate any negative effects and, based on that, new policies should be created. This extends beyond the business field and encompasses societal implications. However, we need to acknowledge that merely implementing an AI system is not a guarantee of success. Firms need to integrate AI into their organizational culture and provide adequate training to their employees, as new technologies, including AI, often introduce vulnerabilities. Policies that go through security breaches or AI hacks are required in order to build trust in AI systems. Additionally, policies centralizing a business ecosystem around AI require specialized considerations, as business environments significantly differ from each other.

5.4 Limitations

While the study demonstrates robustness, it is important to acknowledge several limitations. In our literature reviews (MRP1 and MRP2), we attempted to explore themes within the realm of IT-business value, although our approach was not exhaustive in documenting and presenting these themes in the paper. While we systematically examined and evaluated the articles' contents, we did not adhere to a specific protocol for data recording and reporting. Additionally, there is a lack of comparisons and integrations with other related studies, meaning that this thesis does not furnish a framework or methodology for implementing RAIG practices. Instead, the thesis aims to establish a foundation for synthesizing knowledge from various research strands in this field and propose directions for future research, primarily centered on the business aspects of AI. Another limitation arises from the diverse origins of the papers included, leading to fragmentation. Consequently, integrating them into a cohesive framework within the context of RAIG presents a notable challenge. Furthermore, the sources predominantly came from Western publications, which may introduce bias, as perspectives from regions like Asia, including China, India, and Japan, are conspicuously absent. Finally, further research is essential to expand the current frontier of knowledge in AI by incorporating principles and philosophies from traditional disciplines into existing AI frameworks, which could potentially serve as the basis for RAIG frameworks. Different research avenues can be pursued in response to the questions we posed. For instance, investigating how the automation and augmentation of jobs are affected by automating tasks or identifying obstacles to digital transformation based on RAIG.

Our qualitative studies have limitations too (see MRP3 and MRP4). First, our data collection primarily relies on interviews with companies that do not extensively deal with sensitive data. Consequently, our data may carry a degree of bias and offer an incomplete perspective on the challenges associated with these practices. Second, although we conducted numerous interviews with key personnel within these organizations, our data collection represents a snapshot in time and may not fully capture the entire spectrum of practices. Third, the organizations' relatively short experience with AI deployment and their cautious, gradual approach to mitigating risks might have positively influenced their experiences, possibly affecting the results. Last, all the cases we examined originate from the same sector, potentially raising concerns about the generalizability of our findings.

As for our quantitative studies (see MRP5 and MRP6), the participating companies are located in specific geographical areas where there is a reputation for upholding high standards of responsible and ethical practices. Consequently, it would be intriguing to explore how regions in different geographical areas, such as North America or Asia, address similar challenges. Another limitation of our survey is that it provides a mere snapshot of the activities of these companies. Given our limited knowledge of how they develop and enhance their AI products over time, we cannot discern the evolution of their practices and the mechanisms they employ. It is worth noting that we did not gauge diverse performance metrics, including those related to social responsibility, reputation, or trust for both studies. These metrics have the potential to exert both positive and negative influences on a company's standing in the market, as they can encapsulate the value an organization may accrue in the medium or long term. Another limitation could be response bias, where respondents alter their true opinions to match the norms. Alongside that limitation are selection bias, non-response bias, and the depth of our surveys, which are common limitations that apply to surveys.

6 Conclusion

One of the main obstacles to achieving RAIG is the lack of well-defined guidelines and regulations specific to AI. With AI technologies advancing at a rapid pace, organizations, and policymakers face challenges in navigating the regulatory environment and determining suitable mechanisms to govern AI systems. This lack of universal standards adds another layer of complexity to AI governance, resulting in fragmented regulations that slow down innovation. There is a notable disconnect between the use of AI and the implementation of responsible AI practices. Despite some initial efforts, the low adoption rate shows that it is vital to identify effective practices for AI governance, explore any undesired effects associated with them, and investigate the relationship between RAIG and organizational performance.

This doctoral thesis has examined the ways in which AI provides a competitive advantage over the competition, with a particular focus on how responsible AI use can promote a company's credibility and through that enhance its overall firm performance. The thesis offers valuable insight into various aspects, including the procedural, structural, and relational practices of AI implementation, while exploring the potential dark side effects of AI adoption. Furthermore, the thesis lists the enablers and inhibitors of AI use that facilitate the deployment of AI, providing a complete framework and model for AI governance that can be beneficial for enterprises seeking to adopt or reorganize their AI strategy.

In addressing the research gap in RAIG, we employed a diverse set of research techniques. We initiated our inquiry with an exploratory approach aimed at uncovering key concepts and their interconnections. Subsequently, we conducted confirmatory studies to confirm our assumptions. Our methodological framework included several components, such as a systematic literature review, a single-case study and a multi-case study, in-depth interviews, a questionnaire, and a survey. In sum, our research involved the active participation of 473 firms, alongside interviews with 29 experts from the energy sector in Norway. Our findings discuss the ways in which RAIG boosts the capacity to acquire and distribute knowledge when there is strategic alignment with the company's objectives. Other findings are about the classification of typologies of AI use, with primary and secondary consequences of AI deployment. Additionally, in our findings, we introduced a set of hypotheses about the activities that drive the core processes related to the orchestration of AI resources. Based on these, we ended up with four main contributions:

C1: *Improve understanding of AI in business, its overall business value, and how to gain its competitive advantage.*

C2: *Identify enablers, inhibitors, and antecedents of AI.*

C3: *New knowledge on RAIG and which RAIG practices and principles are considered essential.*

C4: *Explore the relationship between RAIG and firm performance.*

Most importantly, this thesis is an initial effort to provide empirical insights into the relationship between RAIG, legitimacy and firm performance, with its impact on both business and society.

6.1 Avenues for Future Research

In our journal papers, we have highlighted different directions for future research. Some of these include exploring AI adoption and diffusion, addressing overarching challenges in the AI landscape, and delving into AI-powered value propositions. Future researchers may investigate and develop the following:

- Develop ethical frameworks and guidelines specific to AI technologies to create robust decision-making processes.
- Investigate methods for assessing the societal and environmental impacts of AI systems, focusing on the ecosystems that the AI system will affect.
- Study different governance models and structures for overseeing AI development and deployment, including both public and private players.
- Explore the role of AI in addressing global challenges like healthcare, sustainability, and automation while ensuring ethical and responsible use.
- Investigate governance challenges posed by AI technologies, such as virtual agents.

6.2 Final Remarks

In summary, the adoption of responsible AI practices represents a strategic necessity for organizations seeking to outperform their competitors. By placing a priority on RAIG implementation, organizations can unlock a plethora of advantages that can be seen throughout their operations. Foremost among these, prioritizing RAIG bestows organizations with credibility, trustworthiness, and acceptance within their stakeholder ecosystem. Creating an environment marked by ethical principles and transparency boosts confidence among stakeholders, making stronger and more enduring relationships with partners and customers. Beyond this, RAIG serves as a preventive mechanism against risks, addressing ethical concerns and issues. By proactively engaging with these matters, organizations substantially reduce the probability of enduring reputational damage or causing harm to users, physical or not. This proactive stance ultimately safeguards the long-term sustainability of the organization. Moreover, the alignment of responsible practices with organizational goals serves as a catalyst for improved firm performance. Organizations that embrace RAIG are better positioned to develop innovative solutions that not only fulfil societal needs but also adhere to ever-evolving regulatory requisites. Additionally, decision-making within such organizations stands to benefit significantly from the insights garnered through RAIG frameworks. These insights translate into more informed and ethical decision-making processes, thereby catalyzing positive business outcomes. Perhaps one of the most salient advantages of RAIG is its inherent adaptability. As organizations embrace these practices, they become more agile, enabling them to pivot swiftly in response to shifting norms and regulatory landscapes. In essence, RAIG is far from a mere luxury for organizations; it is an indispensable framework. It serves as the linchpin ensuring that organizations not only meet their ethical and legal obligations but also flourish in a fiercely competitive environment where trust, risk management, and sound decision-making are paramount. To neglect RAIG is to court severe consequences that could imperil the very existence of the organization. Therefore, it should be regarded as a strategic imperative for any forward-thinking entity aiming to thrive in the AI-driven landscape of today and tomorrow.

References

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*, 103299.

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424-438.

Accenture. (2019). *The sphere of control: Accenture 2019 Global Risk Management Study Financial Services Report*. Accenture. https://www.accenture.com/us-en/insights/financial-services/global-risk-study

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda* (pp. 197-236). University of Chicago Press.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adam Cutler, A. W., Amit Paka. (2020). *Staying ahead of the curve: The business case for responsible AI*. T. Economist. https://www.eiu.com/n/staying-ahead-of-the-curve-the-business-case-for-responsible-ai/

Afiouni, R. (2019, December). *Organizational learning in the rise of machine learning* International Conference on Information Systems, Munich, Germany.

Akram, M. S., Goraya, M., Malik, A., & Aljarallah, A. M. (2018). Organizational performance and sustainability: exploring the roles of IT capabilities and knowledge management capabilities. *Sustainability*, *10*(10), 3816.

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, *60*, 102387.

Al-Asadi, M. A., & Tasdemir, S. (2022). Using artificial intelligence against the phenomenon of fake news: a systematic literature review. *Combating Fake News with Computational Intelligence Techniques*, *1001*, 39-54. https://doi.org/10.1007/978-3-030-90087-8_2

Al-Surmi, A., Bashiri, M., & Koliousis, I. (2022). AI based decision making: combining strategies to improve operational performance. *International Journal of Production Research*, *60*(14), 4464-4486.

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial Intelligence Adoption: AI-readiness at Firm-Level. *Artificial Intelligence*, *6*, 26-2018.

Alsheibani, S. A., Cheung, Y., Messom, C., & Alhosni, M. (2020). *Winning AI Strategy: Six-Steps to Create Value from Artificial Intelligence* AMCIS, 11, Online. https://aisel.aisnet.org/amcis2020/adv_info_systems_research/adv_info_systems_research/1

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019, May). *Software engineering for machine learning: A case study* 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), (pp. 291-300), Montreal, QC, Canada.

Aqeel, M., Ali, F., Iqbal, M. W., Rana, T. A., Arif, M., & Auwul, R. (2022). A Review of Security and Privacy Concerns in the Internet of Things (IoT). *Journal of Sensors*, *2022*, 1-20. https://doi.org/10.1155/2022/5724168

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

Ayling, J., & Chapman, A. (2021). Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics, 2(3)*, 405-429.

Azzutti, A., Ringe, W.-G., & Stiehl, H. S. (2022). Machine Learning, Market Manipulation, and Collusion on Capital Markets: Why the "Black Box" Matters. *European Banking Institute Working Paper Series 2021 - no. 84, University of Pennsylvania Journal of International Law*, *43*(1), 79. https://doi.org/10.2139/ssrn.3788872

Baby, C. J., Khan, F. A., & Swathi, J. (2017). *Home automation using IoT and a chatbot using natural language processing* 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), 1-6.

Baier, L., Jöhren, F., & Seebacher, S. (2019, June). Challenges in the Deployment and Operation of Machine Learning in Practice. ECIS, Stockholm and Uppsala, Sweden.

Baskarada, S. (2014). Qualitative Case Study Guidelines. *The Qualitative Report*, *19(40)*, 1-25. https://ssrn.com/abstract=2559424

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., & Mojsilović, A. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, *63*(4/5), 4: 1-4: 15.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. *Information Systems Frontiers*, *22*(2), 299-313.

Boell, S. K., & Cecez-Kecmanovic, D. (2015). On being 'systematic'in literature reviews. *Formulating research methods for information systems, 30*, 161-173. https://doi.org/10.1057/9781137509888_3

Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, *57*(17), 102225. https://doi.org/10.1016/j.ijinfomgt.2020.102225

Bourgon, J. (2007). Responsive, responsible and respected government: towards a New Public Administration theory. *International review of administrative sciences*, *73*(1), 7-26.

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: how might the New Zealand government respond? *Policy Quarterly*, *13*(4). Available at: https://apo.org.au/sites/default/files/resource-files/2017-11/apo-nid120146.pdf

Brackett, M., & Earley, P. S. (2017). *DAMA-DMBOK: data management body of knowledge*. Technics Publications, LLC.

Brighton, H., & Gigerenzer, G. (2015). The bias bias. *Journal of Business Research*, *68*(8), 1772-1784.

Butcher, J., & Beridze, I. (2019). What is the State of Artificial Intelligence Governance Globally? *The RUSI Journal*, *164*(5-6), 88-96.

Bytniewski, A., Matouk, K., Chojnacka-Komorowska, A., Hernes, M., Zawadzki, A., & Kozina, A. (2020). *The functionalities of cognitive technology in management control system* Asian Conference on Intelligent Information and Database Systems, 230-240, Springer International Publishing, https://doi.org/10.1007/978-3-030-42058-1_19.

Caner, S., & Bhatti, F. (2020). A Conceptual Framework on Defining Businesses Strategy for Artificial Intelligence. *Contemporary Management Research*, *16*(3), 175-206.

Castillo, D., Canhoto, A. I., & Said, E. (2021). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, *41*(13-14), 900-925.

Chang, H., Nguyen, T. D., Murakonda, S. K., Kazemi, E., & Shokri, R. (2022). *On adversarial bias and the robustness of fair machine learning* International Conference on Learning Representations, https://arxiv.org/pdf/2006.08669.pdf

Charmaz, K. (2014). *Constructing grounded theory*. Sage Publications.

Chatila, R., Dignum, V., Fisher, M., Giannotti, F., Morik, K., Russell, S., & Yeung, K. (2021). Reflections on Artificial Intelligence for Humanity. In (1 ed., pp. 13-39). Springer Cham. https://doi.org/10.1007/978-3-030-69128-8

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., Gigante, A., Valencia, A., Rementeria, M. J., & Chadha, A. S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine*, *3*(1), 1-11.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, *35*(4), 410-422.

Cohen, M., Vellambi, B., & Hutter, M. (2020). Asymptotically unambitious artificial general intelligence. Proceedings of the AAAI Conference on Artificial Intelligence, 34(3), 2467-2476, New York, USA.

Colleoni, E. (2013). CSR communication strategies for organizational legitimacy in social media. *Corporate Communications: an international journal*, *18*(2), 228-248.

Conboy, K. (2019). Being promethean. *28*(2), 119-125. https://doi.org/10.1080/0960085X.2019.1586189

Conboy, K., Dennehy, D., & O'Connor, M. (2020). 'Big time': An examination of temporal complexity and business value in analytics. *Information & Management*, *57*(1), 103077.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of business and psychology*, *25*, 325-334.

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems*, *29*(4), 101600.

Dai, N. T., Du, F., Young, S. M., & Tang, G. (2018). Seeking legitimacy through CSR reporting: Evidence from China. *Journal of Management Accounting Research*, *30*(1), 1-29.

Damper, R. I., Marchand, Y., Adamson, M. J., & Gustafson, K. (1999). Evaluating the pronunciation component of text-to-speech systems for English: A performance comparison of different approaches. *Computer Speech & Language*, *13*(2), 155-176.

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of Data and Analytics* (1 ed., pp. 296-299). Auerbach Publications. https://doi.org/10.1201/9781003278290

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24-42.

de Almeida, P. G. R., dos Santos, C. D., & Farias, J. S. (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, *23*(3), 505-525. https://doi.org/10.1007/s10676-021-09593-z

de Laat, P. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? *Philosophy & technology*, *34*(4), 1135-1193. https://doi.org/10.1007/s13347-021-00474-3

Dellmuth, L., & Schlipphak, B. (2020). Legitimacy beliefs towards global governance institutions: a research agenda. *Journal of European Public Policy*, *27*(6), 931-943. https://doi.org/10.1080/13501763.2019.1604788

Demlehner, Q., & Laumer, S. (2020a). Shall We Use It or Not? Explaining the Adoption of Artificial Intelligence for Car Manufacturing Purposes.

Demlehner, Q., & Laumer, S. (2020b, June). *Shall we use it or not? Explaining the adoption of artificial intelligence for car manufacturing purposes* 28th European Conference on Information Systems - Liberty, Equality, and Fraternity in a Digitizing World, Marrakech Morocco. https://aisel.aisnet.org/ecis2020_rp/177

Dignam, A. (2020). Artificial intelligence, tech corporate governance and the public interest regulatory response. *Cambridge Journal of Regions, Economy and Society*, *13*(1), 37-54. https://doi.org/10.1093/cjres/rsaa002

Dignum, V. (2017). *Responsible autonomy* Proceedings of the 26th International Joint Conference on Artificial Intelligence, 4698–4704, Melbourne, Australia. https://doi.org/10.5555/3171837.3171945

Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (1 ed., Vol. 2156). Springer Nature. https://doi.org/10.1007/978-3-030-30371-6

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, *129*, 961-974.

Dwivedi, Y. K., Hughes, Laurie, Ismagilova, Elvira. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Ehret, M., & Wirtz, J. (2017). Unlocking value from machines: business models and the industrial internet of things. *Journal of Marketing Management*, *33*(1-2), 111-130.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550.

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. *AI and Ethics*, *1*(1), 73-80.

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: a Literature Review. *Information Systems Frontiers*, *24*(5), 1-26. https://doi.org/10.1007/s10796-021-10186-w.

Eriksson, T., Bigi, A. and Bonera, M. (2020). Think with me, or think for me? On the future role of artificial intelligence in marketing strategy formulation. *The TQM Journal*, *32*, 795-814. https://doi.org/10.1108/TQM-12-2019-0303

European Commission. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Fadler, M., & Legner, C. (2021). *Toward big data and analytics governance: redefining structural governance mechanisms* Proceedings of the 54th Hawaii International Conference on System Sciences, Hawaii. https://doi.org/10.24251/hicss.2021.691

Farrokhi, A., Shirazi, F., Hajli, N., & Tajvidi, M. (2020). Using artificial intelligence to detect crisis related to events: Decision making in B2B by artificial intelligence. *Industrial Marketing Management*, *91*, 257-273. https://doi.org/10.1016/j.indmarman.2020.09.015

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, *26*(6), 3333-3361. https://doi.org/10.1007/s11948-020-00276-4

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & 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). https://doi.org/10.2139/ssrn.3518482

Flavián, C., & Casaló, L. V. (2021). Artificial intelligence in services: current trends, benefits and challenges. *The Service Industries Journal*, *41*(13-14), 853-859. https://doi.org/10.1080/02642069.2021.1989177

Floridi, L., & Cowls, J. (2021). A unified framework of five principles for AI in society. In *Ethics, governance, and policies in artificial intelligence* (pp. 5-17). Springer. https://doi.org/10.2139/ssrn.3831321

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., & Moro, E. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*(14), 6531-6539.

Fuchs, D. J. (2018). The dangers of human-like bias in machine-learning algorithms. *Missouri S&T's Peer to Peer*, *2*(1), 1. https://scholarsmine.mst.edu/peer2peer/vol2/iss1/1

Future of Life Institute. (2017, Retrieved November 24, 2022). *Beneficial AI 2017*. Retrieved 2022, November 16 from https://futureoflife.org/event/bai-2017

Galloway, C., & Swiatek, L. (2018). Public relations and artificial intelligence: It's not (just) about robots. *Public relations review*, *44*(5), 734-740. https://doi.org/10.1016/j.pubrev.2018.10.008

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58-62.

Ghadage, Y. H., Shelke, S. D. (2016). *Speech to text conversion for multilingual languages* International Conference on Communication and Signal Processing (ICCSP), 0236-0240, IEEE, Melmaruvathur, India.

Ghallab, M. (2019). Responsible AI: requirements and challenges. *AI Perspectives*, *1*(1), 1-7. https://doi.org/10.1186/s42467-019-0003-z

Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of responsible AI: From ethical guidelines to cooperative policies. *Frontiers in Computer Science*, *4*, 873437.

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, *29*(13), 1465-1474. https://doi.org/10.1002/smj.722

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Gligor, D. M., Pillai, K. G., & Golgeci, I. (2021). Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain. *Journal of Business Research*, *133*, 79-88. https://doi.org/10.1016/j.jbusres.2021.04.043

Google. (2019). *Perspectives on Issues in AI Governance*. Retrieved 16 November 2021 from https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2020). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*(3), 534-551. https://doi.org/10.5465/amr.2019.0178

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing science*, *23*(4), 519-529.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99-120.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Wendell 40 Dopazo Joaquin 41 Furlanello Cesare 42, M. A. Q. C. S. B. o. D. S. T. K. R. S. S.-A. T. W. W. R. D. M. C. E. J., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., & Kundaje, A. (2020). Transparency and reproducibility in artificial intelligence. *Nature*, *586*(7829), E14-E16.

Hair Jr, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, *1*(2), 107-123.

Hammer, A., & Karmakar, S. (2021). Automation, AI and the future of work in India. *Employee Relations: The International Journal*, *43*(6), 1327-1341.

Hamon, R., Junklewitz, H., & Sanchez, I. (2020). *Robustness and explainability of Artificial Intelligence : from technical to policy solutions* (Publications Office of the European Union, Issue. https://data.europa.eu/doi/10.2760/57493

Harfouche, A., Quinio, B., Skandrani, S., & Marciniak, R. (2017). A framework for artificial knowledge creation in organizations. ICIS Seoul, South Korea.

Helbing, D. (2019). Machine Intelligence: Blessing or Curse? It Depends on Us! In *Towards Digital Enlightenment* (pp. 25-39). Springer.

Hickok, M., & Maslej, N. (2023). A policy primer and roadmap on AI worker surveillance and productivity scoring tools. *AI and Ethics*, *3*, 673–687. https://doi.org/10.1007/s43681-023-00275-8

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration. https://www.cochrane-handbook.org

Jan, A. A., Lai, F.-W., Draz, M. U., Tahir, M., Ali, S. E. A., Zahid, M., & Shad, M. K. (2021). Integrating sustainability practices into islamic corporate governance for sustainable firm performance: From the lens of agency and stakeholder theories. *Quality & Quantity*, 1-24.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586.

Jelinek, T., Wallach, W., & Kerimi, D. (2021). Policy brief: the creation of a G20 coordinating committee for the governance of artificial intelligence. *AI and Ethics*, *1*(2), 141-150. https://doi.org/10.1007/s43681-020-00019-y

Jelonek, D., Mesjasz-Lech, A., Stępniak, C., Turek, T., Ziora, L. (2020, March). *The Artificial Intelligence Application in the Management of Contemporary Organization: Theoretical Assumptions, Current Practices and Research Review. In: Arai, K., Bhatia, R. (eds) Advances in Information and Communication* FICC, San Francisco, USA.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Johnson, J. (2019). Artificial intelligence & future warfare: implications for international security. *Defense & Security Analysis*, *35*(2), 147-169.

Jordan, P. J., & Troth, A. C. (2020). Common method bias in applied settings: The dilemma of researching in organizations. *Australian Journal of Management*, *45*(1), 3-14.

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM Computing Surveys (CSUR)*, *55*(2), 1-38. https://doi.org/10.1145/3491209

Keegan, B. J., Dennehy, D., & Naudé, P. (2022). Implementing Artificial Intelligence in Traditional B2B Marketing Practices: An Activity Theory Perspective. *Information Systems Frontiers*, 1-15.

Khuong, M. N., Truong An, N. K., & Thanh Hang, T. T. (2021). Stakeholders and Corporate Social Responsibility (CSR) programme as key sustainable development strategies to promote corporate reputation—evidence from vietnam. *Cogent Business & Management*, *8*(1), 1917333.

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*. https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Koh, K., Li, H., & Tong, Y. H. (2023). Corporate social responsibility (CSR) performance and stakeholder engagement: Evidence from the quantity and quality of CSR disclosures. *Corporate social responsibility and environmental management*, *30*(2), 504-517.

Kolbjørnsrud, V., Amico, R. and Thomas, R.J. (2017). Partnering with AI: how organizations can win over skeptical manager. *Strategy & Leadership*, *45*(1), 37-43. https://doi.org/10.1108/SL-12-2016-0085

Korinek, A. (2019). *Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence* (Vol. w26130). National Bureau of Economic Research.

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120-124.

Kushwaha, A. K., Kumar, P., & Kar, A. K. (2021). What impacts customer experience for B2B enterprises on using AI-enabled chatbots? Insights from Big data analytics. *Industrial Marketing Management*, *98*, 207-221.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications policy*, *44*(6), 101976.

Larsson, S., Anneroth, M., Felländer, A., Felländer-Tsai, L., Heintz, F., & Ångström, R. C. (2019). Sustainable AI: An inventory of the state of knowledge of ethical, social, and legal challenges related to artificial intelligence. *AI Sustainability Center*, *68*, 101926. https://doi.org/10.1016/j.techsoc.2022.101926

Leavy, S. (2018). *Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning* IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 14-16, Gothenburg, Sweden.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: the case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity*, *5*(3), 44.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2021). Trustworthy AI: From Principles to Practices. *Association for Computing Machinery*, *55*(9). https://doi.org/10.1145/3555803

Li, J., & Huang, J.-S. (2020). Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society*, *63*, 101410. https://doi.org/10.1016/j.techsoc.2020.101410

Lichtenthaler, U. (2019). An Intelligence-Based View of Firm Performance: Profiting from Artificial Intelligence. *Journal of Innovation Management*, *7*(1), 7-20.

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A., & Tang, J. (2022). Trustworthy ai: A computational perspective. *ACM Transactions on Intelligent Systems and Technology*, *14*(1), 1-59.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS quarterly*, 293-334.

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262-273.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*(4), 603-609.

Margetts, H. (2022). Rethinking AI for Good Governance. *Daedalus*, *151*(2), 360-371.

Mariani, M. M., & Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. *Journal of Business Research*, *121*, 338-352.

Mark Van Rijmenam, & Schweitzer, J. (2018). How to build responsible AI? Lessons for governance from a conversation with Tay. AOM Specialized Conference: Big Data and Managing in a Digital Economy, Surry, UK.

Matthews, J. (2020). Patterns and Anti-Patterns, Principles and Pitfalls: Accountability and Transparency in AI. *AI Magazine*, *41*(1), 82-89.

Mezgár, I. (2021). From Ethics to Standards; an Overview of AI Ethics in CPPS. *IFAC-PapersOnLine*, *54*(1), 723-728.

Michael, D. M. (1997). Qualitative Research in Information Systems. *MIS Quarterly Executive*, *21*(2), 241-242.

Mikalef, P., Conboy, K., & Krogstie, J. (2021). Artificial intelligence as an enabler of B2B marketing: A dynamic capabilities micro-foundations approach. *Industrial Marketing Management*, *98*, 80-92.

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems*, *31*(3), 257-268.

Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019). Developing an artificial intelligence capability: A theoretical framework for business value. International conference on business information systems, 409-416, Seville, Spain.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), 103434.

Min, J., Kim, J., & Yang, K. (2023). CSR attributions and the moderating effect of perceived CSR fit on consumer trust, identification, and loyalty. *Journal of Retailing and Consumer Services*, *72*, 103274.

Mishra, A. N., & Pani, A. K. (2020). Business value appropriation roadmap for artificial intelligence. *VINE Journal of Information and Knowledge Management Systems*, *51*(3), 353-368. https://doi.org/10.1108/vjikms-07-2019-0107

Mökander, J., & Floridi, L. (2023). Operationalising AI governance through ethics-based auditing: an industry case study. *AI and Ethics*, *3*, 451–468. https://doi.org/10.1007/s43681-022-00171-7

Monostori, L. (2003). AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. *Engineering applications of artificial intelligence*, *16*(4), 277-291.

Mora-Cantallops, M., Sánchez-Alonso, S., García-Barriocanal, E., & Sicilia, M.-A. (2021). Traceability for trustworthy ai: A review of models and tools. *Big data and cognitive computing*, *5*(2), 20. https://doi.org/10.3390/bdcc5020020

Nath, V., & Levinson, S. E. (2014). *Autonomous military robotics*. Springer. https://doi.org/10.1007/978-3-319-05606-7_5

Nuruzzaman, M., & Hussain, O. K. (2018). *A survey on chatbot implementation in customer service industry through deep neural networks* 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), 54-61, Xi'an, China.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*(1), 43. https://doi.org/10.17705/1CAIS.03743

Pagallo, U., Aurucci, P., Casanovas, P., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Schafer, B., & Valcke, P. (2019). AI4People-On Good AI Governance: 14 Priority Actions, a SMART Model of Governance, and a Regulatory Toolbox. https://ssrn.com/abstract=3486508

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and avoiding bias in research. *Plastic and reconstructive surgery*, *126*(2), 619.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2021). *Deploying AI Governance practices: A revelatory case study* 20th IFIP Conference e-Business, e-Services, and e-Society, Galway, Ireland.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2022). Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. *Information Systems Frontiers*, *25*(1), 123–114. https://doi.org/10.1007/s10796-022-10251-y

Papagiannidis, E., Mikalef, P., Conboy, K., & Van de Wetering, R. (2023). Uncovering the dark side of AI-based decision-making: A case study in a B2B context. *Industrial Marketing Management*, *115*, 253-265.

Papagiannidis Emmanouil, Mikalef, P., & Conboy, K. (2023). Responsible AI governance: a systematic literature review. *Manuscript in preparation*.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183-199.

Partnership on AI. (2018). *About us*. Retrieved January 25, 2019 from https://www.partnershiponai.org/about/

Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business & Industrial Marketing*, *34*(7), 1410–1419. https://doi.org/10.1108/JBIM-10-2018-0295

Peräkylä, A. (2004). Reliability and validity in research based on tapes and transcripts'. *Social research methods: a reader. New York: Routledge*, 325-330.

Pessoa, A. S. G., Harper, E., Santos, I. S., & Gracino, M. C. d. S. (2019). Using reflexive interviewing to foster deep understanding of research participants' perspectives. *International Journal of Qualitative Methods*, *18*. https://doi.org/10.1177/1609406918825026

Plastino, E., & Purdy, M. (2018). Game changing value from Artificial Intelligence: eight strategies. *Strategy & Leadership*, *46*(1), 16-22.
Pumplun, L., Tauchert, C., & Heidt, M. (2019, June). *A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors* n Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden. https://aisel.aisnet.org/ecis2019_rp/106
Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, *85*(1), 131-151.
Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative research in accounting & management*, *8*(3), 238-264. https://doi.org/10.1108/11766091111162070
Ramadoss, T. S., Alam, H., & Seeram, R. (2018). Artificial intelligence and Internet of Things enabled circular economy. *The International Journal of Engineering and Science*, *7*(9), 55-63.
Ramesh, B., Cao, L., Kim, J., Mohan, K., & James, T. L. (2017). Conflicts and complements between eastern cultures and agile methods: an empirical investigation. *European Journal of Information Systems*, *26*(2), 206-235.
Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2021). Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems*, *31*(3), 364-387. https://doi.org/10.1080/0960085X.2021.1955628
Rashid, Y., Rashid, A., Warraich, M. A., Sabir, S. S., & Waseem, A. (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods*, *18*. https://doi.org/10.1177/1609406919862424
Recker, J., & Rosemann, M. (2010). A measurement instrument for process modeling research: development, test and procedural model. *Scandinavian journal of information systems*, *22*(2), 1.
Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, *27*(3), 491-497.
Remolina Leon, N. a., & Seah, J. (2019). How to Address the AI Governance Discussion? What Can We Learn From Singapore's AI Strategy? *Centre for AI & Data Governance*, 1-18. https://ink.library.smu.edu.sg/caidg/1
Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. SmartPLS GmbH, Boenningstedt. *Journal of Service Science and Management*, *10*(3).
Ryan, M. (2020). In AI we trust: ethics, artificial intelligence, and reliability. *Science and Engineering Ethics*, *26*(5), 2749-2767.
Schlögl, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). *Artificial intelligence tool penetration in business: Adoption, challenges and fears* International Conference on Knowledge Management in Organizations, 259-270, Zamora, Spain.
Schmidt, R., Zimmermann, A., Möhring, M., & Keller, B. (2020, August). Value creation in connectionist artificial intelligence–a research agenda. AMCIS 2020 proceedings-Advancings in information systems research, Online.
Schneider, J., Abraham, R., & Meske, C. (2020). AI Governance for Businesses. *arXiv preprint arXiv:2011.10672*.
Schwab, K. (2019). *The global competitiveness report 2019*. https://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf
Sharma, P., Kingshott, R., Leung, T. Y., & Malik, A. (2022). Dark side of business-to-business (B2B) relationships. *Journal of Business Research*, *144*, 1186-1195.
Sharma, R. (2022). Corporate Social Responsibility and Customer Satisfaction: Role of Artificial Intelligence. *Acta Universitatis Bohemiae Meridionalis, 25*(2), 162–174. https://doi.org/doi:10.32725/acta.2022.016

Shneiderman, B. (2020). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1-31.

Sikolia, D., Biros, D., Mason, M., & Weiser, M. (2013, May). *Trustworthiness of grounded theory methodology research in information systems* Proceedings of the Eighth Midwest Association for Information Systems Conference, Illinois. http://aisel.aisnet.org/mwais2013

Singapore Government. (2020). *Model Artificial Intelligence Governance Framework* (Second Edition: Singapore Digital (SG:D), Issue. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International*, *20*(4), 97-106.

Smuha, N. A. (2021). Beyond a human rights-based approach to AI governance: promise, pitfalls, plea. *Philosophy & technology*, *34*(1), 91-104.

Stewart, J. (2012). Multiple-case study methods in governance-related research. *Public Management Review*, *14*(1), 67-82.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, *13*(1), 24.

Stupak, I., Mansoor, M., & Smith, C. T. (2021). Conceptual framework for increasing legitimacy and trust of sustainability governance. *Energy, sustainability and society*, *11*(1), 1-57.

Sun, Y., Li, S., & Yu, L. (2022). The dark sides of AI personal assistant: effects of service failure on user continuance intention. *Electronic Markets*, *32*(1), 17-39.

Taeihagh, A. (2021). Governance of artificial intelligence. In: Policy and Society. *Taylor & Francis, pp. 1–21*.

Tallon, P. P., & Pinsonneault, A. (2011). Competing perspectives on the link between strategic information technology alignment and organizational agility: insights from a mediation model. *MIS quarterly*, 463-486.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: toward a theory of information governance. *Journal of Management Information Systems*, *30*(3), 141-178.

Tarafdar, M., Pullins, E. B., & Ragu-Nathan, T. (2015). Technostress: negative effect on performance and possible mitigations. *Information Systems Journal*, *25*(2), 103-132.

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems*, *37*(1), 6.

Thelisson, E., Morin, J.-H., & Rochel, J. (2019). AI governance: Digital responsibility as a building block: Towards an index of digital responsibility. *Delphi*, *2*, 167.

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, *2*(1), 10-12.

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447-464.

Troisi, O., Maione, G., Grimaldi, M., & Loia, F. (2020). Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management*, *90*, 538-557.

Tschang, F. T., & Almirall, E. (2021). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*, *35*(4), 642-659. https://ink.library.smu.edu.sg/lkcsb_research/6669

Tseng, S.-M., & Lee, P.-S. (2014). The effect of knowledge management capability and dynamic capability on organizational performance. *Journal of Enterprise Information Management*.

Turel, O., & Ferguson, C. (2020). Excessive use of technology: can tech providers be the culprits? *Communications of the ACM*, *64*(1), 42-44.

Venkataramanan, V., Packman, A. I., Peters, D. R., Lopez, D., McCuskey, D. J., McDonald, R. I., Miller, W. M., & Young, S. L. (2019). A systematic review of the human health and social well-being outcomes of green infrastructure for stormwater and flood management. *Journal of environmental management*, *246*, 868-880.

Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K. S., & Myles, P. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *bmj*, *368*(l6927). https://doi.org/10.1136/bmj.l6927

Wamba-Taguimdje, S.-L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, *26*(7), 1893-1924. https://doi.org/10.1108/bpmj-10-2019-0411

Wang, H., Huang, J., & Zhang, Z. (2019). *The Impact of Deep Learning on Organizational Agility* International Conference on Interaction Sciences, Munich, Germany. https://aisel.aisnet.org/icis2019/governance_is/governance_is/26

Wang, Y., Xiong, M., & Olya, H. (2020). *Toward an understanding of responsible artificial intelligence practices* Proceedings of the 53rd hawaii international conference on system sciences, 4962-4971, Hawaii International Conference on System Sciences (HICSS).

Welling, M. (2019). Artificial intelligence versus intelligence engineering. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.364ce476

Widder, D. G., Nafus, D., Dabbish, L., & Herbsleb, J. (2022). *Limits and possibilities for "Ethical AI" in open source: A study of deepfakes* Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, https://doi.org/10.1145/3531146.3533779

Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133). https://doi.org/10.1098/rsta.2018.0085

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, *43*(9), 818-829.

Wu, S. P.-J., Straub, D. W., & Liang, T.-P. (2015). How information technology governance mechanisms and strategic alignment influence organizational performance. *MIS quarterly*, *39*(2), 497-518.

Wu, W., Huang, T., & Gong, K. (2020). Ethical principles and governance technology development of AI in China. *Engineering*, *6*(3), 302-309.

Wynn Jr, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS quarterly*, *36*(3), 787-810. https://doi.org/10.2307/41703481

Yang, X., Yu, X., & Liu, X. (2018). Obtaining a sustainable competitive advantage from patent information: A patent analysis of the graphene industry. *Sustainability*, *10*(12), 4800.

Yin, R. K. (1981). The case study crisis: Some answers. *Administrative science quarterly*, *26*(1), 58-65. https://doi.org/10.2307/2392599

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *In ERA forum*, *20*(4), 567-583.

Zhang, D., Pee, L., & Cui, L. (2021). Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. *International Journal of Information Management*, *57*, 102304. https://doi.org/10.1016/j.ijinfomgt.2020.102304

Zhang, L., Yang, X., Zhu, S., & Xia, Z. (2023). Business Model Innovation and Performance of Startups: The Moderating Role of External Legitimacy. *Sustainability*, *15*(6), 5351.

Zhang, Y., & Yang, F. (2021). Corporate social responsibility disclosure: Responding to investors' criticism on social media. *International Journal of Environmental Research and Public Health*, *18*(14), 7396.

Zicari, R. V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., Kararigas, G., Kringen, P., McCullough, M., & Möslein, F. (2021). Z-Inspection®: a process to assess trustworthy AI. *IEEE Transactions on Technology and Society*, *2*(2), 83-97.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair. *559*(7714), 324–326. https://doi.org/0.1038/d41586-018-05707-8

Part II

The research papers that we produced during the period 2020-2024.

Main research papers added in full length:

MP1: Enholm, Ida Merete; Papagiannidis, Emmanouil; Mikalef, Patrick; Krogstie, John. (2021) **Artificial intelligence and business value: a literature review.** Information Systems Frontiers.

MRP2: Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran. (2023) **Responsible AI governance: a systematic literature review.** The Journal of Strategic Information Systems.

MRP3: Papagiannidis, Emmanouil; Enholm, Ida Merete; Dremel, Christian; Mikalef, Patrick; Krogstie, John. (2022) **Toward AI governance: identifying best practices and potential barriers and outcomes.** Information Systems Frontiers.

MRP4: Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran; Rogier Van de Wetering. (2023) **Uncovering the dark side of AI-based decision-making: A case study in a B2B context.** Industrial Marketing Management.

MRP5: Papagiannidis, Emmanouil; Mikalef, Patrick; Krogstie, John; Conboy, Kieran. (2022) **From responsible AI governance to competitive performance: the mediating role of knowledge management capabilities.** Lecture Notes in Computer Science (LNCS).

MRP6: Papagiannidis, Emmanouil; Mikalef, Patrick. (2023) **Exploring the link between responsible AI governance, legitimacy, and firm performance.**

Secondary research papers added with abstract.

SRP1: Papagiannidis, Emmanouil; Enholm, Ida Merete; Mikalef, Patrick; Krogstie, John. (June 2021) **Structuring AI resources to build an AI capability: a conceptual framework**. Proceedings of the European Conference on Information Systems (ECIS) 2021.

SRP2: Papagiannidis, Emmanouil; Enholm, Ida Merete; Dremel, Christian; Mikalef, Patrick; Krogstie, John. (June 2021) **Deploying AI governance practices: a revelatory case study**. Proceeding of 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society (I3E 2021).

SRP3: Papagiannidis, Emmanouil; Mikalef, Patrick; Conboy, Kieran; Rogier van de Wetering. (September 2022) **The dark side of AI-based decision-making: a study of B2B trading.** Proceedings of the Conference: 21st IFIP Conference I3E2022 e-Business, e-Services, and e-Society.

PAPER 1

Artificial Intelligence and Business Value: a Literature Review

Ida Merete Enholm, Emmanouil Papagiannidis, Patrick Mikalef, John Krogstie

(Information Systems Frontiers)

Information Systems
Frontiers
https://doi.org/10.1007/s10796-021-10186-w

Artificial Intelligence and Business Value: a Literature Review

Ida Merete Enholm[1] • Emmanouil Papagiannidis[1] • Patrick Mikalef[1] • John Krogstie[1]

Accepted: 7 August 2021

Abstract

Artificial Intelligence (AI) are a wide-ranging set of technologies that promise several advantages for organizations in terms off added business value. Over the past few years, organizations are increasingly turning to AI in order to gain business value following a deluge of data and a strong increase in computational capacity. Nevertheless, organizations are still struggling to adopt and leverage AI in their operations. The lack of a coherent understanding of how AI technologies create business value, and what type of business value is expected, therefore necessitates a holistic understanding. This study provides a systematic literature review that attempts to explain how organizations can leverage AI technologies in their operations and elucidate the value-generating mechanisms. Our analysis synthesizes the current literature and highlights: (1) the key enablers and inhibitors of AI adoption and use; (2) the typologies of AI use in the organizational setting; and (3) the first- and second-order effects of AI. The paper concludes with an identification of the gaps in the literature and develops a research agenda that identifies areas that need to be addressed by future studies.

Keywords Artificial intelligence · Systematic literature review · Research agenda · Artificial intelligence capabilities

1 Introduction

While Artificial Intelligence (AI) is not something new, it has gained much attention in recent years (Ransbotham et al., 2018). AI has been argued to be a force of disruption for businesses worldwide and in a wide range of sectors (Davenport & Ronanki, 2018). Organizations implementing AI applications are expected to attain gains in terms of added business value, such as increased revenue, cost reduction, and improved business efficiency (AlSheibani et al., 2020). A recent study by MIT Sloan Management Review found that more than 80% of organizations see AI as a strategic

Ida Merete Enholm
idamen@ntnu.no

Emmanouil Papagiannidis
emmanouil.papagiannidis@ntnu.no

Patrick Mikalef
patrick.mikalef@ntnu.no

John Krogstie john.krogstie@ntnu.no

[1] Department of Computer Science, Norwegian University of Science and Technology, Sem Saelandsvei 9, 7491 Trondheim, Norway

opportunity, and almost 85% see AI as a way to achieve com- petitive advantage (Ransbotham et al., 2017). In the search for competitive advantage, many organizations are thus investing in AI technologies. However, despite the growing interest in AI, many companies struggle to realize value from AI (Fountaine et al., 2019). The expected benefits of AI may be absent even though companies invest time, effort, and re- sources into the adoption process (Makarius et al., 2020).

The introduction of AI in organizational operations signals a new set of barriers and challenges (Duan et al., 2019). Some of these include bridging cross-domain knowledge to develop models that are accurate and meaningful (Duan et al., 2019), identifying, integrating and cleansing diverse sources of data (Mikalef & Gupta, 2021), and integrating AI applications with existing processes and systems (Davenport & Ronanki, 2018). To capture the potential value from AI, organizations need to understand how to overcome these challenges as well as the value-adding potential of these technologies. Yet, recent research on AI is more focused on a technological understanding of AI adoption than identifying the organizational challenges associated with its implementation (Alsheibani et al., 2020). While some studies have identified research gaps (Dwivedi et al., 2019), and looked at important aspects in being able to leverage AI technologies (Mikalef & Gupta, 2021), there is still a lack of a holistic understanding of how AI is adopted and used in organizations, and what are the main value-generating mechanisms.

Published online: 25 August 2021

In this paper we attempt to address this gap by providing a synthesis of the current body of knowledge and developing an agenda that can help advance our knowledge. We therefore perform a systematic collection of the extant literature, and put forward a narrative review by summarizing the existing body of literature and providing a comprehensive report which guides future studies (Templier & Paré, 2015). The objective of this paper is to identify in which ways organizations can deploy AI, and what value-generating mechanisms AI can enable. The first step in our study is collecting studies that examine organizational adoption and use of AI from 2010 onwards. After assessing the papers' relevance and quality, the remaining studies are analyzed and synthesized which lead to a framework form understanding AI business value. Based on the synthesis, a research agenda is created, identifying areas that need to be addressed by future research.

2 Research Methodology

The review was conducted in six distinct stages, following the established method of a systematic literature review in order to ensure that all relevant literature to date was included in our analysis (Kitchenham, 2004). First, the review protocol was developed which outlined the choice and structure of keywords and phrases. Second, the inclusion and exclusion criteria for relevant publications were identified in order to filter those publications that were of interest towards our review. Third, the search for papers was performed based on the pre-defined phrases as combinations of the keywords. The articles found in the search were critically assessed before performing data extraction and synthesizing the findings. The previously mentioned stages (Fig. 1) are described in further detail in the next subsections.

2.1 Protocol Development

The systematic literature review started by developing a review protocol following the method of the *Cochrane Handbook for Systematic Reviews of Intervention* (Higgins, 2008). In this protocol, the main research questions were established together with the search strategy, inclusion, exclusion, and quality criteria. The method of synthesis was also established in the protocol. The following research questions motivated the review process: *What aspects enable or inhibit AI use in the organization? What are the types of AI uses in organizations? Through what mechanisms is AI value realized?* These research questions formed the basis for deciding how to proceed in the next steps a what sets of keywords and data sources to utilize.

2.2 Inclusion and Exclusion Criteria

A number of inclusion and exclusion criteria were applied to set boundaries for the systematic literature review. Studies were included if they were focused on how AI can provide business value or how AI is adopted and used in an organizational context. This meant that studies that focused on solely technical aspects of AI, such as technical infrastructure or benchmarking of difference models were not in the scope of papers that were selected. Only publications from 2010 onwards were selected since the majority of organizational uses of AI, with novel methods, have been in the last decade. Studies not written in English were excluded from this review. In addition, the systematic literature review included journal articles and conference proceedings. Book series, dissertations, reports, and webpages were excluded, as were also other publications that were not peer-reviewed.

2.3 Data Sources and Search Strategy

The first step in the search strategy was to form search strings. Two sets of keywords were created (Appendix Table 8): the first set containing keywords related to AI and associated technologies, and the second set regarding the organizational perspective. Keywords from the two sets were combined to form the search string using wildcard symbols in order to reduce the number of search strings. The search terms were then applied in the search engine Google Scholar, as well as several other electronic databases such as Scopus, Business Source Complete, Emerald, Taylor & Francis, Springer, Web of Knowledge, ABI/inform Complete, IEEE Xplore, and the Association of Information Systems (AIS) library. This was done to ensure that all relevant articles had been indexed. The collection procedure started on September 14, 2020 and was concluded on September 30, 2020. To further ensure that the most important articles had been identified, we performed a separate search in the AIS basket of eight journals using the same sets of strings.

2.4 Quality Assessment

Two of the co-authors went through the papers independently after the eligibility check and assessed their quality in terms of several criteria. Studies were examined in terms of scientific *rigor*, *credibility*, and *relevance*. Scientific rigor meaning that the appropriate research method has been applied. Credibility refers to if the research is believable and the findings are well presented. Relevance refers to if the findings are relevant to the academic community and organizations engaging in AI projects. Together these quality criteria ensure that the papers remaining after this stage are likely to make a valuable contribution to the review. After this stage, 43 papers were left for data extraction and synthesis.

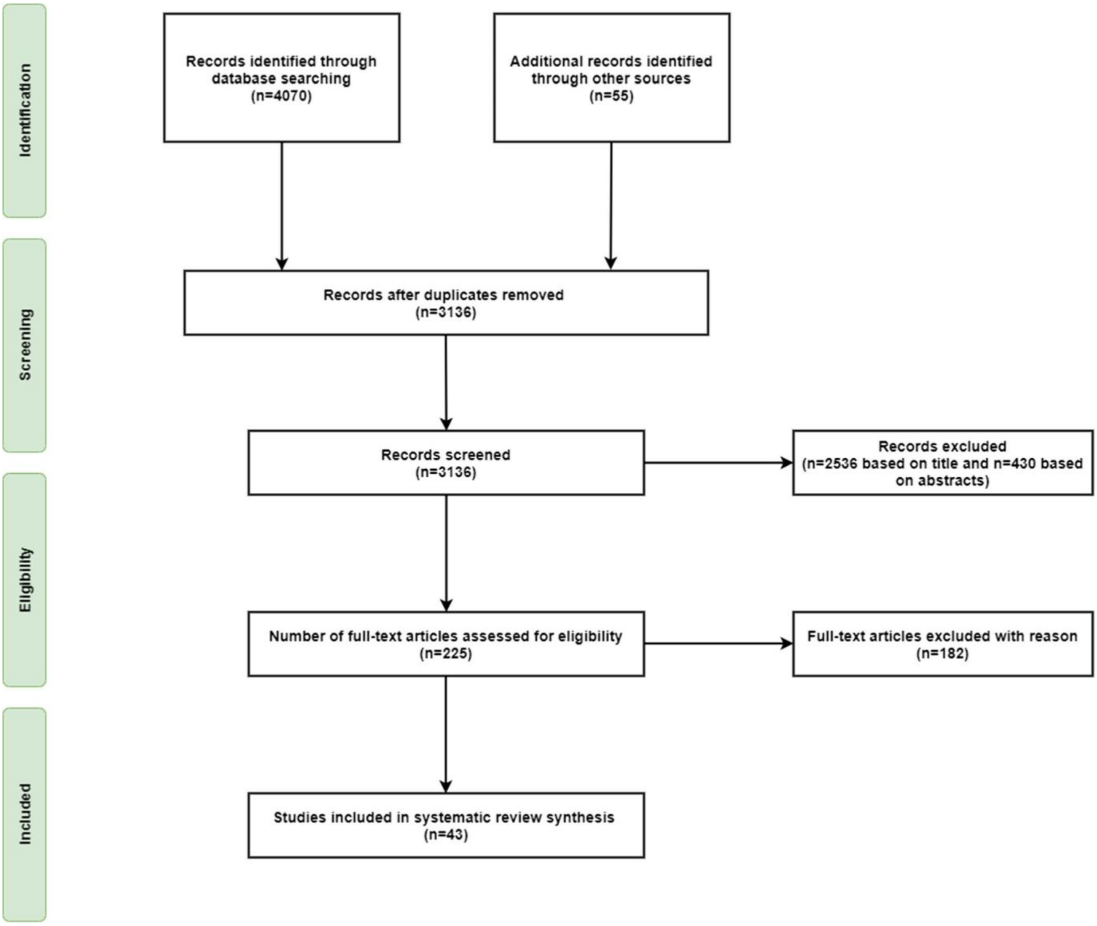

Fig. 1 Stages of the study selection process.

2.5 Data Extraction and Synthesis of Findings

A concept matrix was created in order to categorize the studies and synthesize findings. This was done by analyzing the papers and organizing information from the studies in a spreadsheet. Organizing the studies in this way makes it easier to make comparisons across studies and translate the findings into higher-order interpretations. The studies were analyzed based on the following areas of focus: organizational performance outcomes of AI, adoption, and use of AI in an organizational context, and organizational change caused by the adoption of AI. The information recorded included the research methodology, important definitions, level of analysis, key findings, theories used, context of investigation, and other important concepts from the paper. Two of the co-authors performed the data extraction based on the developed matrix, and then through an iterative process all co-authors reached a consensus about the context included in each category, and about adding additional dimensions to capture all relevant data. The remaining 43 papers were all analyzed and added to the concept matrix before the findings were synthesized.

3 Definitions

While AI has gained much attention in the last years due to the recent advancements in computer hardware, computer network speeds, the vast amount of available data, and processing algorithms (Alsheibani et al., 2020), there is considerable ambiguity about what the notion means and what it entails. The development of AI consists of several sub-disciplines based on fundamentally different approaches (Schmidt et al., 2020), and their meaning is often used interchangeably to encompass a broad set of technologies and applications

(Dwivedi et al., 2019). Therefore, it is essential to draw a clear distinction between these core concepts and provide comprehensive definitions. We draw a distinction between three key areas of focuses: *AI as a scientific discipline*, *technologies used to realize AI*, and *AI capabilities*. These three levels provide a distinction between the discipline and its objective, the tools and technologies used to attain the goal, and the organizational capacity to use a set of diverse tools and technologies that support AI. In the sub-sections below, we present the definitions used in past research and provide a synthesis of the current body of knowledge.

3.1 Artificial Intelligence

Several definitions of AI have been published in an attempt to distinguish it from other conventional information technologies (Table 1). To understand the concept of AI, it is necessary to first understand the notions of "*artificial*" and "*intelligence*" separately. "*Intelligence*" can be described as involving mental activities, such as learning, reasoning, and understanding (Lichtenthaler, 2019). "*Artificial*", on the other hand, refers to something that is made by humans, rather than occurring naturally (Mikalef & Gupta, 2021). By combining these two together, Artificial Intelligence can be understood as making machines capable of simulating intelligence (Wamba-Taguimdje et al., 2020).

From the definitions in Table 1, it is evident that there is a consensus that AI refers to giving the computer human-like capabilities, meaning that computers are able to perform tasks that normally require human intelligence. This includes activities such as understanding, reasoning, and problem-solving (Mikalef & Gupta, 2021). AI emulates human performance by acting as an intelligent agent, which performs actions based on a specific understanding of input from the environment (Eriksson et al., 2020). In other words, the aim of AI is to try to reproduce human cognition by emulating how humans learn and process information. Cognitive technology is a term often used when referring to this capability. Cognitive technologies resemble the action of the human mind (Bytniewski et al., 2020),meaning that it provides the computer the function to think and act like a human.

In their definition, some scholars focus on the idea that AI should not need to be explicitly programmed to perform an intelligent task (Demlehner & Laumer, 2020). It should be able to sense, interpret, learn, plan, comprehend, and act on its own (Demlehner & Laumer, 2020; Kolbjørnsrud et al., 2017; Wang et al., 2019), meaning that AI should be able to correctly interpret external data, learn from this data, and use this learning to achieve specific goals and tasks through flexible adaption (Makarius et al., 2020). Doing so should be achieved without following predetermined rules or action sequences throughout the whole process (Demlehner & Laumer, 2020).

It is also identifiable that there are two main ways of defining AI. The first of these defines AI as a tool that solves a specific task that could be impossible or very time-consuming for a human to complete (Demlehner & Laumer, 2020; Makarius et al., 2020). The second group of definitions regards AI as a system that mimics human intelligence and cognitive processes, such as, interpreting, making inferences, and learning (Mikalef & Gupta, 2021). Both categories of definitions share some similarities but also present some important differences. A common notion in both categories is

Table 1 Sample definitions of artificial intelligence

Author(s) and date	Definition
Kolbjørnsrud et al. (2017)	AI is defined as computers and applications that sense, comprehend, act, and learn.
Afiouni (2019)	AI is the general concept for computer systems able to perform tasks that usually need natural human intelligence, whether rule-based or not
Lee et al. (2019)	Artificial Intelligence: Intelligent systems created to use data, analysis, and observations to perform certain tasks without needing to be programmed to do so
Wang et al. (2019)	AI is a broad concept that captures the intelligent behavior of the machine
Makarius et al. (2020)	Artificial Intelligence: a system's capability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaption
Schmidt et al. (2020)	Artificial Intelligence: The endeavor to mimic cognitive and human capabilities on computers
Demlehner and Laumer (2020)	Artificial Intelligence: a computer system having the ability to percept, learn, judge, or plan without being explicitly programmed to follow predetermined rules or action sequences throughout the whole process.
Wamba-Taguimdje et al. (2020)	Artificial Intelligence: defined as a set of "theories and techniques used to create machines capable of simulating intelligence. AI is a general term that involves the use of computer to model intelligent behavior with minimal human intervention"
Mikalef and Gupta (2021b)	AI is the ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals.

that AI does not necessarily replace humans, but instead, AI operates as an augmentation agent for performing difficult and time-consuming tasks (Mikalef & Gupta, 2021). Yet, both categories of definitions have some diverging points.

While one category of definitions assumes that AI is perfectly capable of imitating human behavior (Kolbjørnsrud et al., 2017; Wang et al., 2019), the second category of definitions regards AI as a tool, assuming it cannot exactly replicate human capabilities (Wamba-Taguimdje et al., 2020). Another noticeable difference is that some definitions refer to AI as a discipline of scientific inquiry (Schmidt et al., 2020), while others perceive the notion as an applied capacity of a system or machine (Afiouni, 2019; Lee et al., 2019). These definitions show that there are noticeable underlying assumptions, and some important differences about what AI is and what it encompasses. For the purpose of this article, we adopt the stance that AI is an applied discipline that aims to enable systems to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals.

3.2 AI Technologies

Moving from the broad definition of what AI encompasses, the next level of definitions attempts to capture the techniques used to realize the objectives set in the previous definitions. Our analysis of the extant literature points out to the fact that this can be achieved through several different ways, with the largest proportion of studies focusing on cases where machine learning, and deep learning were being used. This section provides an overview of how some of the main types of AI technologies are defined in the literature, highlighting some key aspects of them, and outlining some important differences in terms of their application areas.

3.2.1 Machine Learning and Deep Learning

Machine learning is a subset of AI techniques, and one of the most widely used methods over the last few years. Machine learning has gained a lot of interest over the past few years, particularly due to the increase in data availability coupled with advances in computational power (Afiouni, 2019). Several definitions of machine learning exist in the literature, some of them shown in Table 2 as identified in our sample of papers. The objective of machine learning is to train a machine to be able to learn from data and make inferences, predictions, and identify associations, which can guide decisions (Afiouni, 2019; Wang et al., 2019). Machine learning techniques accomplish this by parsing data, learning for data, and making informed decisions based on what has been learned (Wang et al., 2019). This is an inductive approach in which decision rules are identified based on the collected data using statistical methods (Schmidt et al., 2020).

Machine learning algorithms can be further sub-divided into four categories: *supervised*, *semi-supervised*, *unsupervised*, and *reinforcement learning* (Wang et al., 2019). In supervised learning, the training data include the target value (Schmidt et al., 2020). The system then identifies patterns from the training data and infer its own rules from the labeled data (Afiouni, 2019). For unsupervised learning approaches, however, the target value is not included in the training set. The system has to analyze the structure of the training data and its statistical properties to solve the problem (Afiouni, 2019). Unsupervised learning is often used to discover hidden patterns in the data set with prominent applications being automatic clustering, anomaly detection, and association mining (Schmidt et al., 2020). In semi-supervised learning, both labeled and unlabeled data are used (Quinio et al., 2017). In contrast, reinforcement learning does not learn from past data (Afiouni, 2019). Rather, it enables learning from feedback received through interactions with an external environment (Quinio et al., 2017). The core idea is that the system has an objective set by a human agent and receives rewards based on how well the objective is met, which involves finding the best strategy or combination of actions (Afiouni, 2019).

Machine learning can be either *shallow* or *deep*. All four training categories apply to both shallow and deep machine learning. Shallow-structured learning architectures are the most traditional, where it learns from data described by predefined features (LeCun et al., 2015). In contrast, deep machine learning, usually referred to as deep learning, can derive structure from data in a multi-layered manner (Wang et al., 2019). What differentiates deep learning from the more traditional machine learning is the use of an artificial neural network architecture (Afiouni, 2019; Wamba-Taguimdje et al., 2020) Neural network solutions refer to the human brain's functionality (Jelonek et al., 2019) by imitating human neurons (Schmidt et al., 2020). Deep learning is based on creating deep neural networks with several hidden layers, where the layer closest to the data vectors learns simple features, while the higher layers learn higher-level features (Quinio et al., 2017). It represents the world through a hierarchy of concepts, in which each concept can be divided into more straightforward concepts (Borges et al., 2020). In recent years, deep learning has become an area with considerable attention due to its many use cases and its ability to produce remarkably accurate results in various domains (Wang et al., 2019).

3.2.2 Other AI Technologies

While machine learning applications appear to be dominating the research interest in the Information Systems (IS) domain, there are also several other key AI technologies that have been examined in empirical studies and are presented in Table 3. Today, most of these technologies are used in combination with machine learning or deep learning, to provide solutions

Table 2 Sample Definitions of Machine Learning

Author(s) and date	Definition
Wang et al. (2019)	Machine learning empowers the machine to "learn" without explicit programming. This learning process is accomplished by machine itself through collecting data, analyzing data and making predictions.
Wang et al. (2019)	The principle of machine learning incorporates training algorithms to enable machines to learn how to make accurate predictions. There are four training categories of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement.
Afiouni (2019)	Machine learning is that subset of AI that is capable of "learning from data and making predictions and/or decisions" without human dictated rules.
Schmidt et al. (2020)	Machine learning uses an inductive approach in which decision rules are identified based on collected data using statistical methods.
Wamba-Taguimdje et al. (2020)	Machine Learning - automatic learning: machine 'learn' from the datasets offered to them

that to evolve and learn. For instance, in the case of chatbots both natural language processing (NLP) and machine learning are applied (Baby et al., 2017). The functionality enabled through NLP allows chatbots to understand and communicate using the human language. On the other hand, the machine learning algorithms facilitate chatbots to learn and evolve as they get access to more data (Castillo et al., 2020). Other notable types of AI technologies studies in IS empirical works are presented in Table 3.

3.3 AI Capabilities

While the previous definitions concern the broader quest of what AI aims to achieve, as well as the methods and technologies used to actualize these objectives, the notion of an AI capability is revolved around the organizational capacity to deploy such applications in support of operations (Mikalef & Gupta, 2021). With AI becoming an increasingly important asset for organizations, there is a growing body of research examining how such technologies and techniques can be leveraged towards the attainment of organizational goals (Bytniewski et al., 2020; Schmidt et al., 2020; Wang et al., 2019). The notion of an AI capability has thus been introduced to explain how this value is achieved, and how organizations should be organized in order to realize value from AI investments.

While there are still few studies adopting an analysis of AI from the focus point of an organizational capability, there is growing body of research building on this concept as presented in Table 4. The definitions differ slightly but all encompass what an organization should be able to do with AI

Table 3 Definition of other AI Technologies

Technology	Definition	Reference(s)
Natural language processing (NLP)	NLP: the process through which machines can understand and analyze language as used by humans.	Jarrahi (2018)
Computer vision	Computer vision: Algorithmic inspection and analysis of images.	Jarrahi (2018)
Expert system	Expert systems are directed at imitating human decision-making by capturing and representing the expertise of experts for other organizational members to use, serving as a knowledge base.	Afiouni (2019); Lichtenthaler (2019)
Planning and scheduling	The development of action strategies and sequences for subsequent execution	Lichtenthaler (2019)
Speech synthesis systems	Includes text-to-speech and speech-to-text solutions.舄Text-to-speech: the production of speech by machines, by automatic conversion of text to a phonemic specification of pronunciation of the sentences to utter. Speech-to-text systems takes a human speech utterance as an input and requires a string of words as output	Lichtenthaler (2019) Damper et al. (1999) Ghadage and Shelke (2016)

Table 4 Sample definitions of AI capability

Author(s) and date	Definition
Schmidt et al. (2020)	AI capability: The ability of organizations to use data, methods, processes and people in a way that creates new possibilities for automation, decision making, collaboration, etc. That would not be possible by conventional means.
Schmidt et al. (2020)	AI-capabilities are digital capabilities that integrate AI-specific assets, for instance, AI-algorithms, training data, etc. in order to enable value creation.
Wamba-Taguimdje et al. (2020)	AI capabilities could be defined as the firm's ability to create a bundle of organizational, personnel, and AI resources for business value creation and capture.
Mikalef and Gupta (2021)	An AI capability is the ability of a firm to select, orchestrate, and leverage its AI-specific resources.

investments, while some also include the desired outcomes of deploying an AI capability. The definition of Schmidt et al. (2020) for instance belongs to the latter category, as they define AI capabilities as "*the ability of organizations to use data, methods, processes and people in a way that creates new possibilities for automation, decision making, collaboration, etc. that would not be possible by conventional means*". This definition includes not only data and methods, but also the people and processes required to orchestrate and leverage AI into action. Similarly, other definitions include complementary resources that are required to reap the benefits provided by AI technologies (Wamba-Taguimdje et al., 2020). All definitions through converge in that they have an underlying notion that an AI capability is about how an organization uses its AI-specific resources in order to enable value creation (Schmidt et al., 2020; Wamba-Taguimdje et al., 2020). These AI-specific resources can be both technological, e.g. training data and AI-algorithms (Schmidt et al., 2020), and non-technical, e.g. employee skills (Wamba-Taguimdje et al., 2020). Hence, the notion of AI capability extends the view of AI to not only focus on the technical resources, but also include all related organizational resources that are important in order to exploit the full strategic potential of AI.

4 Synthesis of Literature Review

This section presents the findings from the systematic literature review, structured according to the thematic codes that emerged during the analysis of past studies. The findings were obtained through an analysis process following the research methodology. To be able to assess the body of knowledge on AI and business value, we differentiated between three interdependent levels, which are depicted in Fig. 2. In this organizational framework we show that there are several important factors relating to the technological readiness, organizational aspects, and environmental factors that have an important impact on the ability of organizations to deploy and utilize AI. In turn, we develop two broad categories of AI use in organizations and summarize the current knowledge regarding the applications within these categories. Next, we differentiate the impacts of AI into first-order effects and second-order effects. These represent impacts that materialize at the process and firm levels respectively. We therefore argue that second-order effects need to be examine first through the first-order effects they stem from. The section is structured in accordance with the organizing framework, concluding with an overview of theories that have been used in the study of AI and business value.

4.1 Enablers and Inhibitors of AI Use

Based on the clustering of the context of papers, we find that enablers and inhibitors can be subdivided into three main categories: *technological*, *organizational*, and *environmental*. Based on this categorization we discuss what the current body of research and what we know so far about aspects that either accelerate AI deployments or generate obstacles for use. The findings are summarized in Table 5 and discussed below.

4.1.1 Technological

Data At the core of AI is data. Large data sets are used to train the AI (Pumplun et al., 2019; Schmidt et al., 2020). AI learns to make decisions based on these data sets, rather than based on an explicitly defined set of rules defined by expert knowledge (Pumplun et al., 2019; Schmidt et al., 2020). Therefore, an essential enabler of AI adoption in organizations is the data they produce, e.g., sensor data (Demlehner & Laumer, 2020), or have access to (Mikalef & Gupta, 2021). The term big data is often used to refer to these large data sets. According to Beyer and Laney (2012), big data is "*high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization*" (Mikalef et al.,

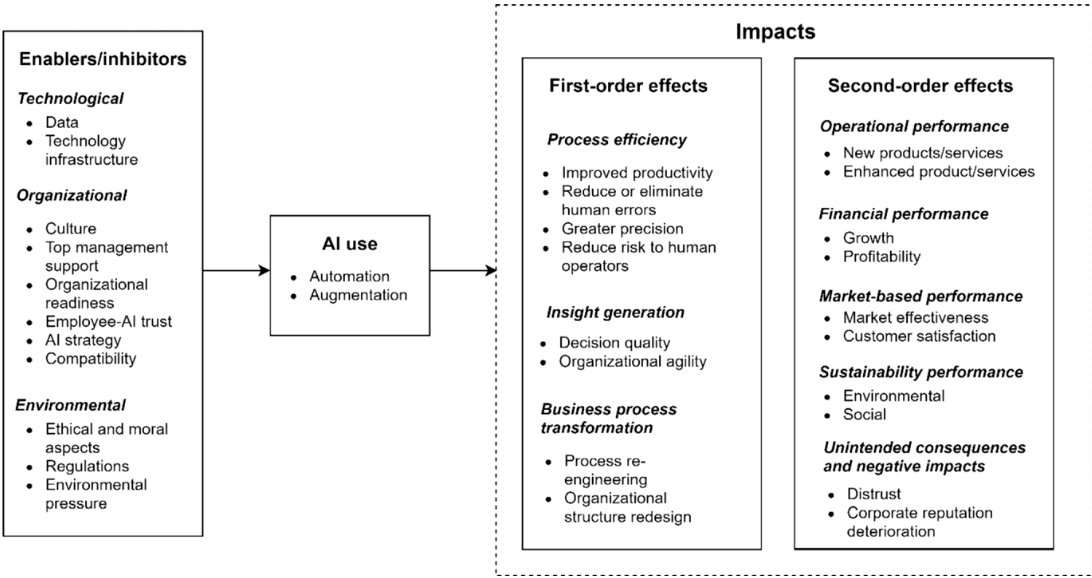

Fig. 2 Organizational framework of AI and business value

2018). This definition captures big data's main characteristics, namely the "three Vs": *volume*, *velocity*, and *variety*. To develop high-quality AI applications, large volumes of training data have to be available (Afiouni, 2019; Keding, 2020; Pumplun et al., 2019; Schmidt et al., 2020). A common challenge when using AI is the lack of enough training data (Baier et al., 2019). Velocity, or timeliness, refers to the speed at which the data are collected and updated (Gregory et al., 2020; Mikalef et al., 2018). Timeliness affects AI systems that heavily rely on the freshness of data, e.g., time-series forecasting. In addition, having a wider range of variety in the training data broadens the model's ability to make predictions, thus increasing its accuracy (Wang et al., 2019).

Another critical aspect of the data used to train the AI is the quality of the data (Alsheiabni et al., 2018; Baier et al., 2019; Demlehner & Laumer, 2020; Lee et al., 2019). Data quality is crucial for providing reliable predictions (Alsheiabni et al., 2018). "Garbage-in, garbage-out" is a fundamental principle for AI (Lee et al., 2019), meaning that if training data has low quality, the insights generated by the AI will also be of low quality and not useful in the organizational context. Common problems regarding the data's quality include incomplete data, incorrect entries, and noisy features (Baier et al., 2019). Recognizing these quality problems can be quite challenging. Thus, data scientists and domain experts need to collaborate closely to identify data quality problems (Baier et al., 2019). An important aspect of quality also relates to using data that are free from bias and follow responsible and trustworthy principles. Bias can be introduced in the used data at different points, such as during the generation or collection, or even during the processing. Ntoutsi et al. (2020) propose a number of methods in their work in order to understand, mitigate, and account for bias in order to reduce negative consequences. Such bias stems not only during the generation or collection, but is also a result of annotation, when data is assigned semantic meaning (Geva et al., 2019). Hence, we see that from the body of empirical studies that data characteristics are multifaceted, and are a core requirements in order to be able to actualize AI applications (Afiouni, 2019; Mikalef & Gupta, 2021).

Technology infrastructure A complementary and equally important aspects for organizations is having the right technology infrastructure for adopting AI AlSheibani et al., 2020). To successfully deploy AI in an organization, three things are needed: *computing power infrastructure*, *algorithms*, and *rich data* sets (Wamba-Taguimdje et al., 2020). AI algorithms build models based on the data. These algorithms can be complex, and the data sets can be enormous. Therefore, the infrastructure could require massive amounts of computing power (Baier et al., 2019; Wamba-Taguimdje et al., 2020). In other words, having high speed and being 'infinitely' scalable (Wamba-Taguimdje et al., 2020). It is not feasible for many companies to have these resources on-site (Schmidt et al., 2020). Large companies, such as Google, Amazon, and Microsoft, have thus started to provide infrastructure for machine learning in the cloud (Borges et al., 2020), e.g., Google Cloud AI. These solutions give other organizations online access to the infrastructure necessary for adopting AI

Table 5 Enablers and inhibitors of AI use

Category	Aspects		References
Technological	Data	• Availability	Afiouni (2019); Alsheiabni et al. (2018); Baier et al. (2019); Demlehner and Laumer (2020); Keding (2020); Pumplun et al. (2019); Schmidt et al. (2020)
		• Volume	Afiouni (2019); Keding (2020); Pumplun et al. (2019); Schmidt et al. (2020)
		• Velocity	Gregory et al. (2020)
		• Variety/diversity	Wang et al. (2019)
		• Quality	Alsheiabni et al. (2018); Demlehner & Laumer, (2020), Baier et al. (2019)
	Technology infrastructure	• Computing power	Baier et al. (2019), Wamba-Taguimdje et al. (2020)
		• Cloud infrastructure	Wang et al. (2019), Schmidt et al. (2020), Borges et al. (2020)
		• Algorithms	Wamba-Taguimdje et al. (2020)
Organizational	Culture		Pumplun et al. (2019), Lee et al. (2019)
	Top management support		Alsheiabni et al. (2018), Demlehner and Laumer (2020), AlSheibani et al., (2020)
	Organizational readiness	• Financial resources	Pumplun et al. (2019)
		• Employee skills	Alsheibani et al. (2020), Pumplun et al. (20192019)
	Employee-AI trust		Makarius et al. (2020), Keding (2020)
	AI strategy		Keding (2020), Finch et al. (2017)
	Relative advantage		AlSheibani et al., (2020), Alsheibani et al. (2018)
	Compatibility		Pumplun et al. (2019), Alsheiabni et al. (2018), AlSheibani et al., (2020)
Environmental	Ethical and moral aspects		Coombs et al. (2020), Alsheibani et al. (2020), Baier et al. (2019)
	Regulations	• Governmental policies and regulations	Baier et al. (2019), Pumplun et al. (2019)
		• Industry requirements	Coombs et al. (2020), Pumplun et al. (2019)
	Environmental pressure	• Competitive pressure	AlSheibani et al., (2020), Demlehner and Laumer (2020), Pumplun et al. (2019)
		• Customer pressure/ readiness	Pumplun et al. (2019)

(Borges et al., 2020; Schmidt et al., 2020; Wang et al., 2019). Therefore, to adopt AI, companies either need access to a cloud-based solution or have the right computational hardware to facilitate the use of AI on their own.

4.1.2 Organizational

Organizational enablers and inhibitors of AI are concerned with how the organizational context, such as strategic orientation and organizational structure, affects the organization's ability to adopt AI successfully.

Culture The culture in the company is argued in research to be a strong force in the decision to adopt AI (Mikalef & Gupta, 2021; Pumplun et al., 2019). AI can be seen as an innovative technology, possibly changing the company's business model and systems (Lee et al., 2019). Thus, the organization must be able to respond to this change. This includes having employees willing to use the new technology in the long run (Pumplun et al., 2019). Innovative cultures have a passion for and willingness to exploit new, opportunistic ideas, and are therefore more likely to embrace AI technologies (Mikalef & Gupta, 2021). Having employees who are continuously willing to learn and innovate will support the deployment and use of AI applications (Lee et al., 2019). This is because employees with an innovative mindset are more open to using a new technology, as well as being able to identify and seize new opportunities for applications of AI. Therefore, organizations with an innovative culture are posited to be better positioned to integrate AI in their work line (Mikalef & Gupta, 2021).

Top Management Support One of the strongest determinants of AI adoption, and a recurrently noted aspect is top management support (Alsheiabni et al., 2018; AlSheibani et al., 2020; Alsheibani et al., 2020; Demlehner & Laumer, 2020). Adopting AI is a complicated process where many challenges must be faced, organizational as well as technological. Top managers and business owners should thus take part in exploring AI technologies and not leave this solely to the technologists (Alsheibani et al., 2020). For example, a company's culture has shown to influence AI adoption, as discussed above, and top managers play a crucial role in establishing this culture (Lee et al., 2019). Also, top-level management can support the adoption of AI by allocating resources and providing capital funds (AlSheibani et al., 2020). The dedication and engagement of top-level management is thus suggested to be a strong contributor towards organizational AI deployment

Organizational Readiness Organizational readiness refers to the availability of the complementary organizational resources needed for AI adoption (Alsheiabni et al., 2018; AlSheibani et al., 2020). As with other innovations, the adoption of AI requires financial resources through a dedicated budget (Pumplun et al., 2019). A high budget, with no obligations to meet specific performance targets, is suggested to enable the adoption of AI, as employees have the ability to learn while working with the development of AI solutions (Pumplun et al., 2019). Additionally, the implementation of AI is heavily dependent upon the skills of the organization's human resources. Adopting new technology may lead to new skill requirements. Organizations adopting AI need employees with technical skills to create and deploy AI systems, e.g., they should be able to utilize technical AI libraries such as TensorFlow, PyTorch, or Keras (Pumplun et al., 2019). They also need domain experts who understand the tasks, workflows, and logic of the existing business processes and have the ability to consider how AI systems can improve them (Alsheibani et al., 2020; Pumplun et al., 2019). An evaluation of the internal availability of expertise is thus required in order to ensure that technical employees, as well as managerial staff know not only how to utilize such novel tools and technologies, but also for what business functions they should be targeted towards (Mikalef & Gupta, 2021).

Employee-AI Trust AI systems have been shown to be able to perform tasks that replicate human cognition or automate previously manual tasks (Zheng et al., 2017). In many of these cases, humans were the ones responsible for carrying out such tasks, and the implementation of AI can consequently change the roles of the organization's employees. Roles may need to be redesigned, and new roles can emerge. Thus, the employees need to understand the purpose of AI, what role it will play, and how it will change the employee's role and responsibilities within the organization (Makarius et al., 2020). Employees possibly have to co-work or base their decisions based on AI systems. This means that they have to trust the AI system, and have an understanding about how they operate and reach conclusions (Makarius et al., 2020). The interaction between humans and AI is a complex process and building trust between humans and machines can be difficult. AI does not experience emotions the same way as a human does, and neither does it have the same empathy capabilities (Makarius et al., 2020). Employee-AI trust can thus be an inhibitor of AI use, with employees causing strong inertial forces to change. The problem of trust however, also applies to managers since they need to know that AI will operate according to the design directives. A manager's willingness to trust an AI system is related to the degree to which there is an understanding of the technology (Keding, 2020).

AI Strategy To reap the benefits of AI, organizations should develop an AI strategy (Finch et al., 2017a; Keding, 2020). The strategy should describe how the organization will adopt and implement AI in order to utilize its benefits. The actions described should align with the company's existing goals

(Keding, 2020). AI strategies are not merely stating what the organization would like to achieve with the implementation of AI, but also provide specific processes, plans, and timeframes for actualizing these objectives. In addition, an AI strategy is likely to require considerable modifications to how the organization is structured, the level of collaboration between departments, as well as how data is governed throughout the organization (Mikalef & Gupta, 2021). Thus, it is essential first to define the relative advantage and compatibility of the AI solution to organizational goals and strategy.

Compatibility Compatibility refers to the fit between the desired application and technology (Pumplun et al., 2019). A stronger fit between the technology and the task will lead to higher levels of adoption and use (Mishra & Pani, 2020). The compatibility concept can be divided into two subcategories: business processes and business case (Pumplun et al., 2019). A concrete, solid business case has to be formulated and aligned with existing strategies (Alsheiabni et al., 2018; AlSheibani et al., 2020; Pumplun et al., 2019). This means defining an exact problem that the adoption of AI is intended to solve (Pumplun et al., 2019). A solid business case should describe what the AI technology will do and demonstrate how its algorithms will enhance business processes' execution and outcomes (Alsheiabni et al., 2018). When adopting AI, new requirements will arise. The company's business processes must be adapted to these new requirements for the adaption to be successful (Pumplun et al., 2019).

4.1.3 Environmental

Organizations operate in dynamic and constantly changing environments, consisting of actors such as competitors and government, that have an influence on how the organization can and should conduct business. This, in turn, exerts different types of pressure on the organization's ability and propensity to adopt AI. This section presents environmental enablers and inhibitors of AI use.

Ethical and Moral Aspects Ethical and moral aspects are essential aspects when adopting AI. AI systems have human-like abilities, which means that the boundaries between humans and machines become less transparent. Thus, the organization must ensure that AI applications have been developed based on ethical principles and do not embed within them unknown biases (S. A. Alsheibani et al., 2020; Baier et al., 2019; Coombs et al., 2020). AI ethics have been defined as "*... a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies*" (Alsheibani et al., 2020). AI ethics can help organizations make sure that their use of technology aligns with their values. Transparency, bias, and discrimination are some of the challenges that emerge when developing AI systems (Alsheibani et al., 2020; Baier et al., 2019). AI is data-driven, thus it can lead to potentially biased and discriminatory outcomes if the underlying data set is imbalanced or discriminatory (Baier et al., 2019). It can also replicate the biases and preconceptions of the system designer. In fact, there have been several reports on prominent companies such as Apple and Amazon, on misuse of AI which resulted in discrimination and bias (Dastin, 2018; Vigdor, 2019).

In taking a more holistic perspective on ethical and moral aspects surrounding AI, several public and private bodies have initiated working groups with the aim of defining key principles that should underlie AI use (European Commission, 2019a). A recent report published by the European Commission, highlights seven key dimensions that organizations should consider when deploying AI applications (European Commission, 2019b). These go beyond aspects related to bias, and include dimensions such as transparency of AI applications, accountability, safety and security, societal and environmental well-being, design for universal access, and human agency and oversight. The purpose of reports such as the above and other empirical works is to minimize the potential risks faced by organizations (Arrieta et al., 2020), and to ensure that AI applications enact behaviors that are more ethically correct than humans (Coombs et al., 2020). Building on such principles is also argued to help organizations balance between black-box and white-box AI applications, or in other words, finding the right equilibrium between accuracy and interpretability (Loyola-Gonzalez, 2019; Wanner et al., 2020).

Regulations Government policies and regulations manifest the social attitudes on ethical and moral issues, and provide directives that shape how AI applications are developed. In May 2018, the General Data Protection Regulation (GDPR) was enforced in the European Union (EU) and the European Economic Area (EEA). GDPR regulates activities such as the processing of personal data. This new law has caused some issues for organizations employing AI solutions as they struggle to provide personal data to use in the training of their intelligent machines (Pumplun et al., 2019). Many data sets need to be anonymized to handle these new legal requirements, which makes the use of intelligent, self-learning algorithms more difficult or even impossible (Pumplun et al., 2019). GDPR increases the complexity of the deployment of AI (Baier et al., 2019; Pumplun et al., 2019) and can thus can lead to inhibited AI adoption. Other legal aspects that can prove to hurdles in the adoption of AI concern the intellectual property entailed in AI algorithms and the data sets used by it (Baier et al., 2019; Demlehner & Laumer, 2020). In addition to the governmental regulations, each industry has its own set of requirements that affect AI adoption (Coombs et al., 2020; Pumplun et al., 2019). This can be laws or other external circumstances that affect the company and its interaction with the environment (Pumplun et al., 2019). Highly regulated sectors, such as healthcare, may encounter additional challenges in deploying AI compared to less regulated sectors (Coombs et al., 2020).

Environmental Pressure An important driver of AI adoption is competitive pressure (AlSheibani et al., 2020; Demlehner & Laumer, 2020; Pumplun et al., 2019). Competitive pressure refers to how an organization is affected by its competitors and the action taken in response to these. Attainting a competitive advantage over rivals, means that organizations have to take action in order to reconfigure and adapt based on continuous and rapid change. The threat of losing a competitive advantage therefore acts as a force in motivating organizations to adopt IT innovations (AlSheibani et al., 2020). Competitive pressure can thus make organizations more prone to adopt AI in order to gain or maintain a competitive advantage. On the other hand, there is also a strong pull for the demand side. Customers are the ones who purchase goods and services from a company, which required that organizations need to meet and exceed the needs of its customers. When a company decides to adopt AI, it is also essential to consider its customer base's knowledge and acceptance (Pumplun et al., 2019). Customers are increasingly expecting individualized services and products, such as the recommendation engine of Netflix. This will push the companies to adopt AI in order to design individualized, intelligent products (Pumplun et al., 2019).

4.2 AI Use

The applications of AI span several diverse areas, such as marketing, production management, enterprise management, and customer service (Alsheiabni et al., 2018; Jelonek et al., 2019). AI applications can be deployed across the entire value chain of an organization, and it has the potential to revolutionize many key aspects of our daily lives (Wamba-Taguimdje et al., 2020). AI applications depending on their use can be divided very broadly into two categories: AI for automation and AI for augmentation. Automation refers to AI systems that are tasked in replacing human work, while augmentation enhances human intelligence by providing insight that can aid decision-making. Both automation and augmentation have applications in many organizational processes, or affect the organization's customers through new or improved products and services that implement AI.

4.2.1 Automation

The notion of automation is not something new, it is an established concept relating to machines replacing humans, such as robots performing tasks on an assembly line. This description is true also for the automation enabled by AI, but it does not describe the radical changes that AI causes. Recent advances in AI have enabled machines to learn, improve, and adapt, thus increasing performance over time (Coombs et al., 2020). Therefore, AI technologies are able to automate more complex tasks involving cognition, such as learning and problem-solving (Lee et al., 2019). This automation is often called Intelligent Automation (Welling, 2019). Intelligent Automation enables the automatization of tasks that were previously considered too difficult to automate, such as knowledge and service work (Coombs et al., 2020). An example is the use of virtual robots to automatically process emails (Wamba-Taguimdje et al., 2020).

In the manufacturing and construction industries, AI is used to automate budgeting and planning, as well as inventory and replenishment (Wamba-Taguimdje et al., 2020). In the service context, AI can provide customers with digital and robot services to influence their customer experience (Prentice et al., 2020). An example of this is chatbots, which are conversational software systems that emulate humans' communication capabilities (Nuruzzaman & Hussain, 2018). Chatbots can assist customers through a voice or text interface (Castillo et al., 2020). In the credit card insurance industry, chatbots are used to answer basic questions, resolve insurance claims, sell products, and ensure that the customers are adequately covered by their insurance (Nuruzzaman & Hussain, 2018). Chatbots are thus doing a job that was previously occupied by a human employee.

In addition to using AI for automating tasks within an organization, it can also create new or enhanced products and services to automate tasks for the customers. An example of this is conversational intelligent agents, such as Apple's Siri and Amazon's Alexa (Castillo et al., 2020; Prentice et al., 2020), which can automate tasks such as writing texts, making calls, and starting a playlist through voice commands. These agents can also be coupled with devices, such as Arduino and Raspberry Pi, to provide smart home automation through voice interaction (Matei & Iftene, 2019). This type of systems can automate simple day-to-day tasks at home, e.g., interactions with TV and lights. Another example is the introduction of facial recognition in smart phones, which automates the process of user authentication. These examples show the multitude of potential applications of AI, and the diversity of areas in which they can be used to automate tasks.

4.2.2 Augmentation

In recent years, AI has exceeded humans in performing certain complex tasks (Jarrahi, 2018). AI can process large amounts of information at high speed beyond humans' cognitive capabilities (Jarrahi, 2018). Hence AI can be used to overcome the cognitive limitations of humans. Augmentation refers to integrating AI with human expertise to enhance decisions and optimize actions (Schmidt et al., 2020). The focus is on AI's assistive role, indicating that it supports humans rather than replacing them.

Organizations often produce or have access to vast amounts of data. By considering this data, managers can make better-informed decisions. However, the data are often too complex to be analyzed by a human. Thus, managers can use AI to gain insights through data for better decision-making (Borges et al., 2020). Predictive analytics can learn from data and make accurate predictions and transaction-level decisions (Makarius et al., 2020). Possible use cases include interpreting previously unknown management control indicators and proposing corrective actions when sales decrease and the competition introduces new products (Bytniewski et al., 2020). AI can also be used in the analysis of opinions, attitudes, and emotions related to a particular product or a service (Jelonek et al., 2019), which is becoming more and more critical for organizations as they can get detailed insight to how their customers perceive their offerings (Bytniewski et al., 2020; Davenport & Ronanki, 2018).

In healthcare, computer vision vision can be used to process MRI images of the brain to mark tiny hemorrhages in the images for doctors (Jarrahi, 2018). AI can also detect cancer patterns (Jarrahi, 2018) or create surgical robots that can assist physicians during complicated surgeries (Makarius et al., 2020). In public relations, AI can be used to monitor social media and predicting media trends (Galloway & Swiatek, 2018). In marketing, AI can be applied to customer segmentation to classify customers based on preferences and lifestyle (Mishra & Pani, 2020). In fashion industries, AI is used to anticipate customer habits, predict future trends, and optimize recommendation systems (Wamba-Taguimdje et al., 2020).

AI can also be applied to products and services that organizations offer to enhance their customer's intelligence. An example is Netflix's recommendation engine, which uses various parameters based on the customer data, such as location, content watched, and the data searched by the user, to give personalized recommendations (Netflix (2020). *Machine Learning*, 2020-12-03). These personalized recommendations increase the likelihood of customers choosing to watch something they genuinely will like.

4.3 Impacts of AI

The question of how AI can lead to competitive performance is of interest to every business executive. To answer this question, the impacts of AI at both the process- (first-order) and firm-levels (second-order) should be studied. How does AI change business processes, and how does this lead to competitive performance? The next subsections address the first- and second-order impacts of AI.

4.3.1 First-Order Impacts

The first order effects of AI use are related to the changes it causes at the process level of an organization. Key performance indicators (KPIs) concerned with efficiency, effectiveness, capacity, productivity, quality, profitability, competitiveness and value are common measures of the performance improvements at the process level, and are used to monitor the output of an organization (Wamba-Taguimdje et al., 2020). To assess the impacts of AI on the process level, three different effects are discussed: process efficiency, insight generation and business process transformation.

Process Efficiency Using AI to automate tasks or augment human intelligence in organizations can improve business process performance by increasing efficiency indicators (Coombs et al., 2020; Kirchmer & Franz, 2019). Automation of tasks through AI involves replacing human work with a machine. By automating tasks, organizations may relieve some employees of repetitive routine tasks, which enables them to focus on other knowledge-intensive activities that add more value to the organization (Makarius et al., 2020), thus increasing their productivity (Balasundaram & Venkatagiri, 2020; Bauer & Vocke, 2019; Bytniewski et al., 2020; Finch et al., 2017a). Moreover, machines can perform tasks quicker and with greater precision than humans, increasing organizations' throughput, particularly in manufacturing industries and supply chain operations (Balasundaram & Venkatagiri, 2020; Finch et al., 2017a). Furthermore, AI use can reduce the time required to complete some key business processes (Coombs et al., 2020), and improve the error-rate and lag times by automatizing a series of tasks (Wamba-Taguimdje et al., 2020). For example, using AI in car manufacturing to automate visual recognition of barcodes and license plates improves efficiency compared to when performed by a human employee (Demlehner & Laumer, 2020). The replacement of human work by machines also includes reducing or eliminating errors made by human employees, and increasing transparency. Consequently, the quality of the results is suggested to be improved (Finch et al., 2017b).

Insight Generation One of the most prominent first-order effects of AI is that it can unlock insight and patterns hidden in large volumes of data (Mikalef & Gupta, 2021). By collecting, processing, and disseminating data within and between organizations, AI can present previously unknown information and help make insight-driven decision (Jelonek et al., 2019). According to Lichtenthaler (2019), "*Even if two firms have access to the same internal and external knowledge, they may achieve different competitive positions if one firm has superior intelligence that enables specific insights as a basis for targeted competitive moves that the other firms lacks*". This suggests that organizations should foster ways by which they can leverage AI in order to gain insight that their competitors lack (Lichtenthaler, 2019).

The hidden value unlocked by AI can be used to make better-informed decisions, or even to partially automate tasks.

AI can assist managers overcome their cognitive limitations by providing an efficient way to handle the large volumes of data available (Finch et al., 2017a; Keding, 2020). When decision-makers have access to more detailed knowledge, the quality and speed at which decisions are taken will increase (Keding, 2020). AI, therefore, enables faster and better decision-making (Wang et al., 2019). Organizations that can exploit AI's informational effects are better positioned to quickly sense and respond to market dynamics (Wamba-Taguimdje et al., 2020). This capability of responsiveness is also known as organizational agility, and it consists of sensing, informed decision-making, and responding (Wang et al., 2019). AI, and deep learning, in particular, can play an active role in each of these activities. Specifically, AI applications can be steered towards systematically and effectively identifying patterns and underlying signals that humans may miss (Eriksson et al., 2020), and be trained to respond to these signals fast and accurately (Wang et al., 2019).

Business Process Transformation As an innovative and (often) disruptive technology, AI enables organizations to innovate and transform business processes (Wamba-Taguimdje et al., 2020). The goal of all business processes is to convert inputs into valuable outputs, and new technology is expected to improve these processes through radical transformation (Mishra & Pani, 2020). AI is no exception, as it can enable the redesign of business processes with the intention of radically changing how current operations are executed (Mishra & Pani, 2020). Through this process, AI is also a driver for re-engineering and redesigning the existing organizational structure (Wamba-Taguimdje et al., 2020). It influences how human resources are being used, facilitating change in business processes and the organizational structure. The implementation of AI brings a new set of skills and capabilities for managers, employees, and AI to work together (Makarius et al., 2020). As a consequence, jobs may need to be redesigned, and new jobs can emerge. By using AI, organizations can reallocate resources, which, in the long term, have the potential to redraw the organizations' organizational chart (Eriksson et al., 2020). In other words, the transformational effects of AI on business processes can be either direct, or indirect.

4.3.2 Second-Order Impacts

The second-order impacts of AI are related to the firm-level effects of AI use in operations. These effects can be divided into four categories: operational performance, financial or accounting performance, market-based performance, and sustainability performance.

Operational Performance AI can have an impact on the operational performance in several ways, such as through the introduction of new products and services and enhancing the quality of existing products and services.

Introduction of new products and services One way of reaping the benefits of AI is for companies to identify opportunities to enter the market with a new offering (Mishra & Pani, 2020). AI can search through massive amounts of data to find patterns that can show opportunities for introducing new products and services. For example, by discovering shifts in customer preferences, organizations can find opportunities for entering markets with untapped profitable segments. Besides, as an innovative technology, AI facilitates the design of new products and services (Wamba-Taguimdje et al., 2020). In this regard, there are many possibilities for creating products and services that embed AI-based functionality. For example, organizations can use AI to introduce new services around conventional products in order to enhance customer service with applications such as recommender systems, chatbots, or intelligent agents (Alsheibani et al., 2020). In sequence, the introduction of new products and services can prompt business model innovation. Furthermore, studies have shown that AI-based recommendations can aid product developers in designing new products, particularly when it comes with design aid which can enhance creativity (Mikalef & Gupta, 2021). Business model innovation can, in turn, help companies preserve their market position (Lee et al., 2019).

Enhance the quality of products and services AI can also enhance the quality of already existing products and services. Davenport and Ronanki (2018) found in a survey that more than half of the executives said that their primary goal of adopting AI was to make existing products better. There are numerous ways AI can enhance the quality of products and services. For example, Netflix uses AI to enhance the video quality of their streaming services. Spotify uses AI to enhance their product in several ways, such as providing personalized song recommendations. Personalization of products and services are becoming more and more popular these days. By using AI to analyze customer data, organizations can provide a personalized experience to each customer, possibly causing the customers to perceive the product or service to have enhanced quality. Spotify, Netflix and Amazon are some of the many companies using AI to personalize the experience for customers.

Financial Performance Over the last few years, AI has been gradually embedded in key organizational activities, prompting business growth is various sectors (Eriksson et al., 2020). Organizations that have implemented AI solutions have realized financial and accounting performance gains, such as increased revenue and cost reduction (Alsheiabni et al., 2018; Davenport & Ronanki, 2018). In a recent empirical study, Mikalef and Gupta (2021a) find that companies that have developed a structured approach to AI adoption and use, and developed an organizational capability around the novel technologies have realized performance gains. Their analysis points out to the fact that an AI capability has a positive effect on important financial and accounting performance indicators such as growth in overall financial performance. Nevertheless, to date there are few studies examining other measures of financial performance, such as return-on-investments, profitability, and gross profit margin after the introduction of AI.

Market-Based Performance *Marketing effectiveness* Organizations using AI for marketing purposes are suggested to experience several benefits. A typical way in which AI can lead to marketing performance is to segment customers based on their needs to target the segments with different and customized marketing strategies. AI can enhance customer segmentation by processing and learning from existing customer data, enabling organizations to learn about their customers' preferences and lifestyle in a whole new way. This capability enables a more precise segmentation because organizations can classify customers on a finer level (Mishra & Pani, 2020). Consequently, organizations can target their marketing better (Afiouni, 2019), and it opens for the possibility of delivering one-to-one marketing by personalizing the experience (Mishra & Pani, 2020). Thus, AI enhances the marketing effectiveness and accuracy by targeting the right customers with the right marketing strategy. Also, as customer behavior changes, segmentation suggestions from the AI system are regenerated so that organizations can effectively adapt their marketing strategy (Afiouni, 2019).

Customer satisfaction Customer satisfaction is related to how satisfied a customer is with a company's offerings, and it directly affects the loyalty and retention of customers. By using AI, companies can learn more about their customers' behaviors and, in turn, use this enhanced understanding to proactively prevent any negative experiences (Riikkinen et al., 2018). In doing so, companies can provide offerings that reduce customer attrition, such as providing personalized services or offers. For example, by using AI in the interaction with customers, customer satisfaction can increase because customers get better informed and find better-customized solutions guided by AI (Schmidt et al., 2020). However, the use of AI can also lead to customer dissatisfaction. For example, customers interacting with AI-powered chatbots can find the experience frustrating and ineffective (Castillo et al., 2020). Hence, it is important in the design of AI systems that have a direct interaction with customers, that their experiences and perceptions are considered.

Sustainability Performance AI's disruptive potential can drive business model innovation toward sustainability (Toniolo et al., 2020). Sustainable business models describe how organizations create, deliver, and capture value in a way that contributes to the sustainable development of the company and society (Toniolo et al., 2020). In other words, companies should conduct their business while at the same time focusing on environmental and social matters. AI has the potential to impact individuals and society in a disruptive and long-term manner (Alsheibani et al., 2020).

Environmental AI can affect environmental sustainability, such as by minimizing energy costs, reducing energy consumption and, in turn, reducing negative environmental impacts (Borges et al., 2020; Toniolo et al., 2020). Also, the use of AI tools can help organizations to reduce pollution and waste (Toniolo et al., 2020). A growing body of research is also examining the impact that AI applications have in supporting circular economy strategies, by enabling organizations to pursue strategies that promote recycling, reduction of emissions, and re-use of materials (Rajput & Singh, 2019).

Social By considering social responsibility, organizations can improve their reputation and increase their market share, which in turn can affect their competitive advantage (Toniolo et al., 2020). The adoption of AI opens up many new challenges for organizations in fulfilling their social responsibilities. Examples are challenges regarding privacy and discrimination. Recall that the fundamental enabler of AI systems is data. Organizations need to ensure the privacy of data on their customers and employees (Lee et al., 2019). Also, they must ensure that the the use of AI does not result in discriminatory actions or results. As AI is based on data, the results can be biased or discriminatory if the underlying data is. AI systems understand neither the inputs they process nor their outputs (Keding, 2020). They learn by interpreting patterns in previous data to predict the future. Thus, the results may reflect suspicious patterns, such as sexism and racism, found in the underlying data (Keding, 2020). For example, in recruitment processes: if the AI system explores the existing recruitment process, and this process lacks diversity (e.g. gender and ethnicity), then the results of the system will continue to embrace this underlying discrimination (Afiouni, 2019). On the other side, as AI systems are objective, they can reduce human bias in processes, such as recruitment and customer segmentation (Afiouni, 2019; Toniolo et al., 2020). Also, employees' safety and working conditions can be enhanced with the introduction of AI. Using AI robots in manufacturing where hazards may be present, the safety conditions for employees can increase (Toniolo et al., 2020). Besides, automating repetitive routine tasks causes employees to use their capabilities and competencies elsewhere, possibly leading to more meaningful and creative jobs (Toniolo et al., 2020). This change can affect how employees perceive their working environment.

Unintended Consequences and Negative Impacts While research predominantly focuses on positive effects of AI deployment and use, several recent examples showcase that in the absence of appropriate AI governance practices, negative and unintended consequences can occur. One of the most prominent examples is the failure of organizations to identify and eliminate bias in the data or AI algorithms, which can result in discrimination or unfavorable outcomes to particular ethnic groups, genders, or population clusters. For instance, there have been several news reports on biased AI outcomes concerning gender (Dastin, 2018; Vigdor, 2019) and racial discrimination (Zuiderveen Borgesius, 2020). Such outcomes have negative effects on the image of the companies they involve, and in some cases have resulted to financial losses and significant fines (Engler, 2021). Such outcomes increase the pressure towards organizations that use AI to apply practices to reduce bias in data and algorithms throughout all stages of deployment. In fact, due to the surge of several noteworthy cases of bias and discrimination as a result of AI outcomes, governmental agencies such as the European Commission, are now proposing concrete regulations that will dictate how AI applications are developed and used.

Negative impacts due to AI use, however, are not restricted to biased outcomes, but include a number of other aspects such as black-box algorithms, lack of transparency and accountability, security concerns, as well as harm to society and the environment (Yudkowsky, 2008). An example of the effects such unintended consequences have had includes the growing requirement for organizations to introduce explainability in how AI algorithms reach certain outcomes (Arrieta et al., 2020). In addition, this move has sparked a general need to provide more transparency of the entire process from data collection to outcome generation (Loyola-Gonzalez, 2019). A lack of explainability practices and low transparency hampers individuals trust in AI systems and leads to non-use (Samek & Müller, 2019). In addition, cases of AI use for customer and citizen interaction (e.g. chatbots) that have not taken into account human-centric principles have resulted in frustration and complaints from users, hampering the corporate image (Marcondes et al., 2019).

4.4 Theories and Frameworks in Empirical Studies

In this section we examine the theoretical perspectives that were used in the sample of articles we analyzed. While not all articles built their investigations on a theoretical grounding, a surprisingly high number of papers did. In the table presented below (Table 6), we document those that have been employed, describing how they have been applied in the study of AI, and their overall scope of application. Despite still being at a nascent stage, the papers looking at different facets of AI in organizations demonstrate considerable variety in the use of theoretical perspectives. Specifically, we see that many articles use firm-level theories in studying aspects that contribute to the effective adoption and deployment of AI applications in the organizational setting, such as the TOE framework and the Resource-Based View (RBV) of the firm. As research in this domain is still at an early stage, it expected that the majority of work will be revolved around understanding how to deploy these novel technologies in operations, and complementary organizational resources need to be deployed to support these.

However, we find that several articles also examine the processes of AI development, and the knowledge-intensive practices that surround AI maturation. As AI applications involve a lengthy process of development and refinement, by tweaking algorithms, data, and analysis methods, they create an opportunity for organizations to learn by doing. Several studies have applied relative theoretical perspectives, such as organizational learning theory and theory of artificial knowledge creation to elucidate this process. In addition, as AI applications are heavily data-dependent, other articles such as that of Gregory et al. (2020) have worked on developing new theoretical perspectives such as the network effect, in an attempt to understand how AI platforms become more useful and of value as users and data increase. Finally, some studies have focused on the individual as the unit of analysis, with theoretical perspectives such as dual process theory looking into the interactions of human and AI for optimizing decision-making, and value co-destruction building on a dark-side angle of how negative interactions reduce use of AI systems.

5 Research Agenda

From the synthesis in Section 4, several research gaps are identified in relation to the study of AI use in organizations. Through challenging assumptions and identifying areas where there is a significant lack of knowledge, this section aims to provide a framework for guiding future research. The goal is not to present an exhaustive list of potential research directions, but rather, to highlight some important gaps in our understanding of how AI is shaping the way organizations are conducting business and competing. We therefore define five research themes, with each presenting a number of research directions (D) that can help expand our knowledge. The research framework is presented in Fig. 3, with the themes being represented in the enumerated circles.

5.1 Theme 1: AI Adoption and Diffusion

D1.1 Difficulties in the Process of Adopting and Deploying AI

Although the proposed business value that organizations can derive from AI is argued to be significant for all kind of business operations, there is still a very small percentage of companies that to date have adopted and deployed AI applications beyond pilot projects (Anon, 2020). Companies face a number of challenges when it comes to adopt and deploy AI (Alsheibani et al., 2018). According to Alsheibani et al.

Table 6 Used theories in papers.

Theory	Description	Application	Reference(s)
Contingency theory	Contingency theory posits that there is no one best way to adopt and implement AI, but rather, different contingencies of the internal and external environment need to be considered when doing so.	Organizational adoption and use of AI	Eriksson et al. (2020)
Dual process theory	Understanding how to develop accurate AI-based predictions in uncertain situations requires an examination of the mental processes that underly the cognitive decision-making process.	AI use for decision-making	Dellermann et al. (2017)
Resource-Based View (RBV)	Defines the different types of complementary resources that organizations should foster in order to be able to realize business value from their AI investments.	AI-business value	Mikalef and Gupta (2021a); Wamba-Taguimdje et al. (2020)
TOE framework	The process by which a firm adopts and implements AI is influenced by the technological context, the organizational context, and the environmental context	Organizational AI adoption	Demlehner and Laumer (2020)
Value co-destruction	AI value depends on complex interactions between actors and the functions provided by AI applications	AI individual use	Castillo et al. (2020)
Organizational Learning Theory	There exists a dynamic interplay between learning that occurs in AI algorithms, and learning that occurs within the organization.	Organizational learning through AI projects	Afiouni (2019)
Theory of Artificial Knowledge Creation	Explains how tacit and explicit knowledge held by individuals and organizations can be simultaneously enlarged and enriched through the recursive and reflexive amplification of tacit and explicit knowledge enabled by AI.	Knowledge creation in AI projects	Quinio et al. (2017)
Network effect	A platform exhibits data network effects if the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user	AI platform value	Gregory et al. (2020)

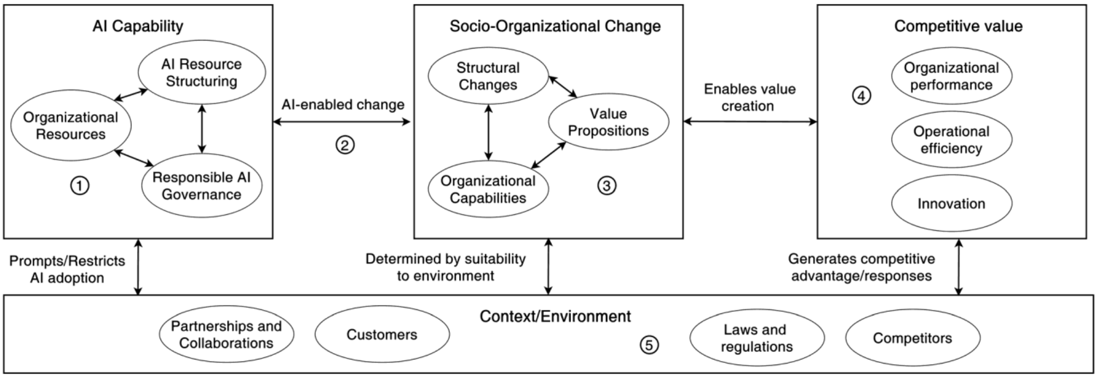

Fig. 3 AI and business value research framework

(2018) technological readiness, organizational readiness, and environmental readiness (environmental conditions such as government regulations) are important aspects that influence the adoption of AI. Other difficulties can include the costs in infrastructure, hiring capable employees and relying on external partners. Hence, the different dynamics that have a role in allowing organizations to adopt AI and in turn develop an AI capability require a deeper understanding. Due to the nature of AI that requires employees from different business units to work together to build AI applications, the socio-technical arrangements and the process through which AI applications are developed and deployed warrants further investigation (Holton & Boyd, 2019).

In addition, conflicts between shareholders and managers could have important consequences on the actual use of AI in operations. Specifically, the conflicting views where shareholders encourage automation for reducing costs (Dedrick et al., 2013), while managers promote augmentation may cause a paralysis in actual deployment (Shollo et al., 2020). Moreover, the use of AI might challenge cultural norms and act as a potential barrier for managers or even customers to accept AI technologies (Dwivedi et al., 2019). Hence, further enlightenment in these areas is needed as it is crucial identifying the difficulties and the cultural obstacles and knowing how to overcome them. Finally, the modes of human-AI symbiosis and the changes these induce in organizational structures require further investigation (Shrestha et al., 2019). AI is argued to lead to significant adjustments to how business and IT functions work, collaborate, and exchange knowledge, so finding optimal ways of doing so is critical for successful AI deployments.

D1.2 Responsible AI Governance While investing in technological infrastructure for AI may be an important part, organizations hoping to use AI in core operations must be able to govern the necessary resources and have thorough practices and mechanisms for orchestrating and following up on projects from ideation to completion (Papagiannidis et al., 2021). In addition, AI applications require several phases of maturation, and are subject to continuous improvement and development. A core requirement for most types of AI applications (e.g., in public sector) is taking into account ethical aspects and principles of responsible design. Hence, the concept of AI governance is inextricably associated with responsible and ethical principles being embedded throughout the process of design, deployment, and evaluation. Therefore, being able to break down the concept of AI governance and outline which key activities underpin the notion is an important research quest.

Past studies on IT governance have shown that having established such practices not only helps optimize output, but also enables organizations to achieve business and IT fit (Tallon & Pinsonneault, 2011). Nevertheless, AI poses an additional concern since the effects towards, as well as the interactions with humans shifts fundamentally. This poses a requirement to examine not only how AI applications are developed so that they are aligned with responsible principles (European Commission, 2019a), but also to anticipate and plan for their effects as the gradually become embedded in everyday activities. In their recent work, Amer-Yahia et al. (2020) outline what they refer to as “intellectual challenges”, which comprise of major themes organizations must consider when they plan to deploy AI applications that concern the changing nature of interaction between humans and technology. An important area of inquiry therefore concerns what responsible AI governance comprises of, as well as what are the effects of implementing such practices at different levels of analysis.

5.2 Theme 2. AI and Socio-organizational Change

D2.1 How Does AI Change Organizational Culture? Organizational culture has been consistently noted as being an important part of AI adoption and use (Mikalef & Gupta, 2021). Innovative cultures are in a better position to adopt AI. But can an innovative technology like AI lead to alteration to the organizational culture itself? This question has yet to be examined, particularly in relation to the ripple effects the adoption and use of AI may have on different aspects of organizational culture, like learning, collaboration, and communication patterns. In addition, an interesting point to explore is if the adoption of innovative technologies like AI affects the organization's ability to innovate further. Does the introduction of AI change the mindset of the employees to being more open to innovations? An interest field of inquiry therefore concerns if and through what mechanisms innovation outcomes are achieved as part of AI deployments. With the introduction of new and disruptive digital technologies, many prominent cases of organizations have documented an increase of innovation output (Nambisan et al., 2017). Future research therefore needs to examine through what arrangements organizations are able to harness the possibilities of AI technologies in order to drive innovation.

Taking a contrarian view, the dark side effects of AI also warrant further investigation in the context of organizational culture. The introduction of AI and displacement or shifting of several conventional job roles is likely to lead to increased tensions, conflict, and feelings of distrust towards the technology itself and the units that promote its deployment (Huang et al., 2019). Therefore, a major challenge for practitioners is how to be able to manage the human factor internally when planning their AI implementations. Negative perceptions can result in rigidity in digital transformation and lead to inertia, thus significantly impacting organizational performance. There is, as a result, a need for future research to examine how IT managers can plan for and deploy AI applications to minimize potential friction and facilitate trust and acceptance of newly deployed solutions.

D2.2 What is the Role of AI-driven Automation in Decision-making? Automating processes through the use of AI is argued to reduce the workload of employees in certain activities and increase efficiency of process completion (Acemoglu & Restrepo, 2018). At the same time, AI is able to automate decision-making when provided with appropriate data and business rules (Duan et al., 2019). Delegating such authority to AI applications however raises the issue of how to prevent bias that AI models might have, and how to ensure that new decision-making structures are improved, rather than debased, with the introduction of AI (Cirillo et al., 2020). While a number of studies have opened up the discussion about what the optimal decision-making structures are and how organizations can ensure that the introduction of AI enhances them, there is still a lack of empirical studies examining the effects of such arrangements (Shrestha et al., 2019). Such studies require an understanding of the impacts from the individual level, up to the business and organizational level of analysis, in order to fully capture the nature and types of effects that blended human-AI arrangements have.

D2.3 How Does AI Change the Organizations Structure? The connection between AI adoption and organizational structure is one of a reciprocal nature. Organizational structure may affect the ability to adopt AI, and AI adoption may affect the organizational structure. Pumplun et al. (2019) found that a company's organizational structure may affect its ability to adopt AI and propose that "*Departments who keep relevant data to themselves, an overreliance on status quo as well as slow and bureaucratically shaped corporate structures will have a negative effect on the adoption of AI in companies*". This proposition suggests that organizations structured in functional silos, will encounter more challenges when adopting AI. A reason for this can be that these structures do not facilitate a holistic approach to solve problems. On the contrary, agile organizational structures are more flexible and can respond quickly to change, thus supporting innovation. However, such arrangements have received little empirical attention to date. Therefore, future research needs to engage in the study of how organizational structures affect AI adoption. Nevertheless, such relationships are likely to have a dynamic and reciprocal nature. As identified during the systematic literature review, AI influences how human resources are used, possibly redesigning the organizational chart (Eriksson et al., 2020) (Wamba-Taguimdje et al., 2020). Previous roles and structures are likely to change, and new roles may emerge. Therefore a promising avenue for future research is to examine through longitudinal approaches how organizations transform in order to embrace AI technologies.

5.3 Theme 3. AI-driven Value Propositions

D3.1 How Does the Orientation of AI Impact Value Propositions? The potential use cases for AI technologies within the organizational sphere are manifold, and a plethora of value-adding applications have been suggested both for private and public organizations (Davenport & Ronanki, 2018). One broad categorization that can be made involves the distinction between the use of AI for internal- and external-oriented functions. Internal functions involve using AI for improving internal business processes, such as decision-making, or for streamlining internal business processes. On the other hand, external functions include using AI in products and services that are in direct contact with customers Some examples of the later include the use of AI to recommend songs of interest to listeners by Spotify. It is therefore expected that the value-adding possibilities of AI applications are very diverse in nature. To date, there are to the best of our knowledge no studies that differentiate on performance metrics depending on the use case of AI. Furthermore, such an area of inquiry also raises the question of what the appropriate metrics are in order to be able to capture effects of AI and how to benchmark different similar applications.

D3.2 What is the Role of Complexity in AI Application Inimitability and Value?

While high complexity in AI applications may lead to black-box systems with limited transparency, high complexity can also result in difficult to imitate projects, leading to a longer period where firms can sustain an edge over their rivals (Wamba-Taguimdje et al., 2020). Nevertheless, the notion of complexity is compound, and involves aspects such as how many features are included in the model (Monostori, 2003), the diversity of data sources used, the interactions with other systems and processes, as well as the breadth and depth of activities they span. There are instances where large cooperation's, such as Alibaba's fraud risk management system (Chen et al., 2015), initiated high complexity projects that yielded significant returns. Nevertheless, some of these projects had little success and the value creation for the business was little if none. Hence, the correlation between the complexity of an AI system and the value creation for the business requires further exploration. Understanding when value creation is adding based on the complexity of the AI system could allow organizations to identify what aspects of their developed AI projects lead to a competitive advantage. As a result, a deeper understanding of how complexity adds or retracts value in the case of AI applications presents an interesting field of study, as well as developing deeper theorizing on the phenomenon of digital complexity (Benbya et al., 2020).

5.4 Theme 4. Competitive Value of AI

D4.1 What are the Effects of AI on Financial Performance?

One of the key expectations from practitioners before adopting AI applications is that they can help improve financial performance indicators, such as revenue, growth, and help reduce costs (Alsheiabni et al., 2018; Eriksson et al., 2020). Nevertheless, there is a long chain of causal associations, and to date it is still not clear if and how AI can help organizations achieve financial performance gains. From our sample of articles there were none that studied the long-term financial consequences of AI adoption. Instead, the focus was on identifying short-term operational trends. Thus, it is important, particularly for small and medium-sized enterprises to elucidate the financial impacts that AI applications have in the long-term. As there are large financial investments tied to AI adoption, it is critical for firms that do not have large slack resources to know exactly the timeline in which AI applications start generating positive financial returns, and through what means and mechanisms. Prior studies have documented that there are large associated costs incurred by some organizations due to technology adoption, and which have resulted in significant financial losses (Chakravorty et al., 2016). It is therefore important to understand where the equilibrium lies between investing in the necessary AI resources, and the expected financial return.

D4.2 What are Appropriate Key Performance Indicators (KPIs) to Measure AI Success?

Measuring the impact of an AI project is challenging as the results are often difficult to capture with purely quantitative measures. While businesses use KPIs to measure performance, AI applications are often gauged in their success in completely different measures. Some examples of AI success measures include calculating various metrics such as Mean Squared Error, Confusion Matrix and F1-score (Kawaguchi et al., 2017). These metrics are good for determining the overall performance of a model, but they say very little about the overall project success. More organizational-focused KPIs could prove more valuable, after AI applications have been deployed and used in practice (Ehret & Wirtz, 2017). Nevertheless, such measures are typically very context specific. In addition, the selected KPIs should be quantifiable and provide managers with insights about the impact of the AI project in the business (Glauner, 2020). There is as a result a large gap in understanding what appropriate measures are to identify AI outcomes, and help guide key stakeholders.

D4.3 How Can AI Drive Innovation?

New products and services have been developed building on the functionality and affordance enabled by AI (Plastino & Purdy, 2018). Some prominent examples include Netflix's recommendation systems, Amazon's chatbot Alexa, and Tesla self-driving cars. Although AI is the technological innovation behind these services and products, there is little understanding regarding the socio-technical dynamics that lead to innovation to be generated. While undoubtedly the novel technologies that support AI have an important impact on the creation of such innovation output, the role of managers and knowledge workers, as well as their interactions needs to be understood in more detail. As new digital solutions are now one of the main components of innovations, it is imperative to understand the nexus of associations that surrounds technology-driven innovation. To date, research on the business value of AI has not built sufficiently on the growing body of knowledge on digital innovation (Nambisan et al., 2020). Thus, there is a need to understand the phenomenon of AI and its role in driving innovation in a more structured and theory-driven manner, that can allow for more nuanced understanding of how such outcomes are achieved.

5.5 Theme 5. AI and the Extended Organization

D5.1 Extended Organizational Boundaries and Partnerships

All businesses, despite their size and industry must interact with the external environment in order to remain competitive and evolve (Yang & Meyer, 2019). A sought-after option by many such organizations is engaging in different forms of organizational relationships, such as mergers, acquisitions, joint ventures and alliances. Yet, when it comes to AI applications literature largely sees the development of AI as an activity that happens in the focal organization. As organizations typically have complementary key datasets, or interlinked organizational processes, it is important to examine how these relationships dictate the types of AI applications that are developed, as well as how they prompt organizations to engage in different forms of organizational engagements. Large corporations have access to AI resources that are unavailable for the majority of the businesses, especially for small and medium-sized enterprises (Garbuio et al., 2011). Despite the managers' efforts for pioneering AI initiatives, it is not always possible to achieve goals due to limitations in key resources (Pellikka & Ali-Vehmas, 2016). A possible model to mitigate such limitations could be to engage in such strategic alliances. Doing so enables the organizations to have access to resources which they would not be able to acquire by themselves in other circumstance, while at the same time, both companies are able to increase their business value and benefit from each other's capabilities. Nevertheless, research regarding governance schemes for effectively cooperating under such AI-specific partnerships is still at an early stage in research. Building on this avenue helps understand that dynamics and conflicts of interest in such collaborative arrangements, as well as optimal ways of organizing. Furthermore, a prominent area of study is how organizations develop the necessary IT infrastructure to facilitate such inter-organizational collaboration around AI.

D5.2 What is the Role of AI in Shaping the Reputation of the Organization?

Maintaining D5.2 What is the role of AI in shaping the reputation of the organization?a good reputation with customers and partners is essential for organizations. It can affect several business areas, such as market value, ability to attract more skilled employees, and customers' loyalty (Eccles et al., 2007). An organization's reputation is highly linked with the ability of customers and stakeholders to trust the organization, and in turn has significant effects on overall financial performance. Yet, the introduction of AI technologies can influence the level of trust among critical external entities, such as customers and business partners. While AI technologies may have many of the same capabilities as humans, in cases where there is a lack of transparency on where and how AI is used, issues of distrust may arise. Some early studies have shown that in order for humans to garner feelings of trust towards AI outcomes, they need to understand how such technologies work, and have clearly defined indications of safety and reliability (Marcus & Davis, 2019). Thus, organizations adopting AI must be aware of the role of trust, how to build trust, and in turn, how trust influences their reputation and interaction with external stakeholders. Thus, a promising area for further research is to understand how the introduction of AI affects the trust people have in the organizations and, in turn, how it affects the organization's reputation. Such research can examine the technical features of AI, how communications patters influence trust-formation, as well as if there are any cultural differences among individuals in how they perceive AI applications (Felzmann et al., 2019).

5.6 Cross-cutting Challenges

The themes presented above that form our proposed research agenda, and the corresponding directions described within these themes, also raise several important concerns regarding the extended information value chains of organizations and the related activities within these (Abbasi et al., 2016; Koutsoukis & Mitra, 2003). In Table 7, we present some of the core challenges within the information value chain, and their relationship to our directions presented above. Specifically, we follow the distinction regarding the sequence of activities within the information value chain that differentiate between data, information, knowledge, decisions, and actions.

The table indicates that there are several cross-cutting challenges among the future directions which we defined. For example, when looking at the data artefact, issues regarding how data infrastructures are designed and deployed, as well as how they need to be adapted to the socio-technical context present a challenge that span several research directions within the first theme. Further challenges such as that of integrating data from a variety of sources, as well ensuring high quality input to AI algorithms, present serious obstacles for contemporary organizations (Ransbotham et al., 2018).

Table 7 Information value chain challenges and research directions

Information Value Chain	Challenge	Direction(s)
Data	Data infrastructures	D1.1, D1.2
	Data integration	D1.1, D1.2, D3.2
	Data quality	D1.2
Information	Information representation	D1.2
	Information access	D1.2, D2.3, D5.1
	Information processing	D1.1, D2.2, D4.3
Knowledge	Innovation management	D3.1, D4.3
	Business intelligence	D2.2, D4.3
Decisions	Decision structures	D2.1, D2.2, D2.3, D4.3
	Accountability	D1.2,
Actions	Value measurement	D4.1, D4.2
	Competitive advantage	D4.1, D5.2

Similarly, defining the procedures that surround information access, processing, and representation constitute tough obstacles for private and public organizations, as they concern technical facets of AI, as well as organizational and procedural aspects that span the entire organization (Dwivedi et al., 2021; Schaefer et al., 2021). As AI applications span multiple units within organizations, being able to deal with the technical requirements, as well as the necessary organizational changes that are needed to generate business value, is a challenge that organizations of all size-classes with be required to face (Mikalef & Gupta, 2021). The same applies also concerning how knowledge that is derived from AI applications or infused into such applications, is managed within organizations. Being able to harness the knowledge that AI applications can deliver is critical in generating business value out of AI applications, so it is important that organizations are structured appropriately in order to leverage such technologies in ways that contribute to value generation (Collins et al., 2021).

A final consideration regarding the cross-cutting themes of AI in organizations has to do with how decision-structures are shifted, as well as what competitive actions such technologies enable. There has been an ongoing debate about the different configurations of decision-making structures that utilize the strengths of human and AI agents, as well as their potential to generate business value (Shrestha et al., 2019). Adding to this, to be able to evaluate the value of AI applications, it is also important to have appropriate indicators of the value they deliver, as well as associate their use with the ability to attain a competitive advantage (Dwivedi et al., 2021).

6 Conclusion

AI is increasingly becoming important for organizations to create business value and achieve a competitive advantage. However, many AI initiatives fail even though time, effort, and resources have been invested. There is a lack of a coherent understanding of how AI technologies can create business value and what type of business value can be expected.

In this paper, we present a narrative review to identify how organizations can deploy AI and what value-generating mechanisms such AI uses have. The result of this analysis consists of three parts. First, several enablers and inhibitors of AI use are identified. These antecedents of AI adoption consist of technological, organizational, and environmental resources and conditions. Second, different use cases for AI are distinguished. Organizations can use AI technologies to automate tasks or augment humans, either for internal or external purposes. Internal purposes mean using AI to improve internal business processes, where the customer is not in direct contact with the AI-solution. Furthermore, external purposes mean using AI in products and services that are in direct contact with the customers. Lastly, the impacts of AI are discussed, specifically how organizations change and how this leads to competitive performance. Several implications of AI at both the process- and firm-level are identified.

The findings in this article have several implications for how to manage AI in organizations. By considering the enablers and inhibitors found, organizations can better assess their ability to adopt AI successfully and know which changes to make. Moreover, by knowing how AI can be used, organizations can make better decisions about where in their value chain to implement AI solutions. Lastly, knowing the possible effects of AI adoption can better prepare organizations to introduce AI in their line of work. We conclude this study by presenting a research agenda that identifies areas that need to be addressed by future research to understand AI technologies' value-generating mechanisms in the broader organizational environment. While this study may not follow an exhaustive approach in documenting and presenting the themes in the paper, we attempt to present themes through the IT-business value perspective. In addition, although a systematic approach was used in searching for and analyzing the paper contents, we did not follow a specific method for documenting and reporting results, such as PRISMA (Moher et al., 2015).

Appendix

Table 8 Keywords used in selection of papers

Thematic Category	Keywords
AI technologies	Artificial intelligence, cognitive technology, robotic automation, cognitive insight, process automation, machine learning, deep learning, cognitive automation, neural network, supervised learning, unsupervised learning, natural language processing, computer vision, machine vision, expert systems, cognitive application, image recognition, reinforcement learning, deep mind technologies, adaptive algorithms, recurrent neural networks, machine perception, machine intelligence, heuristic search techniques, decision tree, data mining, convolutional neural network, cluster analysis, classification, chatbots, autonomic computing, semantic analysis, image recognition, simulation intelligence, challenges of AI, integrate AI, cost of AI, deployment of AI, AI and big data, influence of AI, AI transformation, Bayesian learning system
Organizational	Business value, organizational challenges, organizational opportunities, adoption, business benefits, business process redesign, organizational change, firm performance, organizational performance, competitive advantage, process innovation, business transformation, business process management, digital transformation, business strategy, business gains, business performance, cognitive engagement, business opportunities, transformation process, business activities, data-driven decisions, competitive performance, business efficiency, reduce business costs, business management, business decision, business challenges, commercial value, business value proposition, business growth, business success, customer value, customer fragmentation, customer service, corporate value, leadership, swot analysis, obstacles, deployment, assimilation

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2)

Acemoglu, D., & Restrepo, P. (2018) The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review, 108*(6), 1488–1542

Afiouni, R. (2019). Organizational learning in the rise of machine learning. *International Conference on Information Systems, Munich, Germany*

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at firm-level. *Artificial Intelligence, 6*, 26–2018

Alsheibani, S., Cheung, Y., Messom, C., & Alhosni, M. (2020). Winning AI strategy: six-steps to create value from artificial intelligence. *Americas Conference on Information Systems*, Online

AlSheibani, S., Messom, C., & Cheung, Y. (2020). Re-thinking the competitive landscape of artificial intelligence. *Proceedings of the 53rd Hawaii international conference on system sciences*

Alsheibani, S., Messom, D., Cheung, Y., & Alhosni, M. (2020). Reimagining the strategic management of artificial intelligence: Five recommendations for business leaders. *Americas Conference on Information Systems*, Online

Amer-Yahia, S., Basu Roy, S., Chen, L., Morishima, A., Abello Monedero, J., Bourhis, P., & Demartini, G. (2020). Making AI machines work for humans in FoW. *ACM SIGMOD Record, 49*(2), 30–35

Anon. (2020). *AI is essential to India's future of work, study finds*. M2 Presswire

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115

Baby, C. J., Khan, F. A., & Swathi, J. (2017). Home automation using IoT and a chatbot using natural language processing. *2017 Innovations in Power and Advanced Computing Technologies* (i-PACT)

Baier, L., Jöhren, F., & Seebacher, S. (2019b). Challenges in the deployment and operation of machine learning in practice. In *Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm, Sweden*

Balasundaram, S., & Venkatagiri, S. (2020). A structured approach to implementing Robotic Process Automation in HR. *Journal of Physics: Conference Series , 1427*(1), 012008. IOP Publishing.

Bauer, W., & Vocke, C. (2019). Work in the age of artificial intelligence–challenges and potentials for the design of new forms of human-machine interaction. *International Conference on Applied Human Factors and Ergonomics*

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). Complexity and information systems research in the emerging digital world. *MIS Quarterly 44*(1), 1–17

Beyer, M., & Laney, D. (2012). *The importance of "Big Data": a definition.* Gartner

Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2020). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, 102225

Bytniewski, A., Matouk, K., Chojnacka-Komorowska, A., Hernes, M., Zawadzki, A., & Kozina, A. (2020). The functionalities of cognitive technology in management control system. *Asian Conference on Intelligent Information and Database Systems*

Castillo, D., Canhoto, A. I., & Said, E. (2020). The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, 1–26

Chakravorty, S. S., Dulaney, R. E., & Franza, R. M. (2016). ERP implementation failures: a case study and analysis. *International Journal of Business Information Systems, 21*(4), 462–476

Chen, J., Tao, Y., Wang, H., & Chen, T. (2015). Big data based fraud risk management at Alibaba. *The Journal of Finance and Data Science 1*(1), 1–10

Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., & Chadha, A. S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *NPJ Digital Medicine, 3*(1), 1–11

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management, 60*, 102383

Coombs, C., Hislop, D., Taneva, S. K., & Barnard, S. (2020). The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems, 29*(4), 101600

Damper, R. I., Marchand, Y., Adamson, M. J., & Gustafson, K. (1999). Evaluating the pronunciation component of text-to-speech systems for English: A performance comparison of different approaches. *Computer Speech & Language, 13*(2), 155–176

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved 15/06 from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

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

Dedrick, J., Kraemer, K. L., & Shih, E. (2013) Information technology and productivity in developed and developing countries. *Journal of Management Information Systems, 30*(1), 97–122

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2017). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *International Conference on Information Systems* (ICIS), Seoul, South Korea

Demlehner, Q., & Laumer, S. (2020). Shall we use it or not? Explaining the adoption of artificial intelligence for car manufacturing purposes In *Proceedings of the 28th European Conference on Information Systems (ECIS)*, Online

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T. … Eirug, A. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994

Eccles, R. G., Newquist, S. C., & Schatz, R. (2007) Reputation and its risks. *Harvard Business Review, 85*(2):104

Ehret, M., & Wirtz, J. (2017). Unlocking value from machines: business models and the industrial internet of things. *Journal of Marketing Management, 33*(1–2), 111–130

Engler, A. (2021). *Auditing employment algorithms for discrimination*. Retrieved 15/06 from https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/

Eriksson, T., Bigi, A., & Bonera, M. (2020). Think with me, or think for me? On the future role of artificial intelligence in marketing strategy formulation. *The TQM Journal, 32*(4), 795–814.

European Commission. (2019a). *Ethics guidelines for trustworthy AI*. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419. Accessed 23 Apr 2021

European Commission. (2019b). *Proposal for a Regulation laying down harmonised rules on artificial intelligence*. Retrieved 15/06 from https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

Felzmann, H., Villaronga, E. F., Lutz, C., Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society ,6*(1), 2053951719860542

Finch, G., Goehring, B., & Marshall, A. (2017). The enticing promise of cognitive computing: high-value functional efficiencies and innovative enterprise capabilities. *Strategy & Leadership, 45*(6), 26–33

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

Galloway, C., & Swiatek, L. (2018). Public relations and artificial intelligence: It's not (just) about robots. *Public Relations Review, 44*(5), 734–740

Garbuio, M., King, A. W., & Lovallo, D. (2011). Looking inside: Psychological influences on structuring a firm's portfolio of resources. *Journal of Management, 37*(5), 1444–1463

Geva, M., Goldberg, Y., & Berant, J. (2019). Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets. *arXiv preprint arXiv:1908.07898*

Ghadage, Y. H., & Shelke, S. D. (2016). Speech to text conversion for multilingual languages. *2016 International Conference on Communication and Signal Processing* (ICCSP)

Glauner, P. (2020). Unlocking the power of artificial intelligence for your business. In: Innovative Technologies for Market Leadership. *Future of Business and Finance* (pp 45–59). Springer International Publishing

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2020). *The role of artificial intelligence and data network effects for creating user value*. Academy of Management Review(ja)

Higgins, J. P. (2008). *Cochrane handbook for systematic reviews of interventions version 5.0. 1*. The Cochrane Collaboration. https://www.cochrane-handbook.org

Holton, R., & Boyd, R. (2019). Where are the people? What are they doing? Why are they doing it? (Mindell) Situating artificial intelligence within a socio-technical framework. *Journal of Sociology, 57*(2), 179–195.

Huang, M. H., Rust, R., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence (AI). *California Management Review, 61*(4), 43–65

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586

Jelonek, D., Mesjasz-Lech, A., Stępniak, C., Turek, T., & Ziora, L. (2019). The artificial intelligence application in the management of contemporary organization: Theoretical assumptions, current practices and research review. *Future of Information and Communication Conference*

Kawaguchi, K., Kaelbling, L. P., & Bengio, Y. (2017). Generalization in deep learning. *arXiv preprint arXiv:1710.05468*

Keding, C. (2020). Understanding the interplay of artificial intelligence and strategic management: four decades of research in review. *Management Review Quarterly, 71*(1), 91–134.

Kirchmer, M., & Franz, P. (2019). Value-Driven Robotic Process Automation (RPA). *International Symposium on Business Modeling and Software Design*

Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele University, Issue

Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2017). Partnering with AI: how organizations can win over skeptical managers. *Strategy & Leadership, 45*(1), 37–43

Koutsoukis, N.S., & Mitra, G. (2003). *Decision modelling and information systems: The information value chain*, vol 26. Springer Science & Business Media

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Mature, 521*(7553), 436–444

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: the case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(3), 44

Lichtenthaler, U. (2019). An intelligence-based view of firm performance: profiting from artificial intelligence. *Journal of Innovation Management, 7*(1), 7–20

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access, 7*, 154096–154113

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research, 120*, 262–273

Marcondes, F. S., Almeida, J. J., & Novais, P. (2019). A short survey on chatbot technology: failure in raising the state of the art. *International Symposium on Distributed Computing and Artificial Intelligence*

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust* Vintage

Matei, A. M., & Iftene, A. (2019). *Smart Home Automation through Voice Interaction* RoCHI

Mikalef, P , & Gupta, M. (2021). Artificial Intelligence Capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, Online*. https://doi.org/10.1016/j.im.2021.103434

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management, 16*(3), 547–578

Mishra, A. N., & Pani, A. K. (2020). Business value appropriation roadmap for artificial intelligence. *VINE Journal of Information and Knowledge Management Systems, 51*(3), 353–368

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1), 1–9

Monostori, L. (2003). AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. *Engineering Applications of Artificial Intelligence, 16*(4), 277–291

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing innovation management research in a digital world. *Mis Quarterly, 41*(1), 223–238

Nambisan, S., Lyytinen, K., & Yoo, Y. (2020). *Handbook of digital innovation* Edward Elgar Publishing

Netflix. (2020). Machine Learning. Retrieved 2020-12-03 from https://research.netflix.com/research-area/machinelearning

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., & Krasanakis, E. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1356

Nuruzzaman, M., & Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. *2018 IEEE 15th International Conference on e-Business Engineering* (ICEBE)

Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). 2021). Structuring AI resources to build an AI capability: A conceptual framework. *European Conference on Information Systems*, Marrakesh, Morocco

Pellikka, J., & Ali-Vehmas, T. (2016). Managing innovation ecosystems to create and capture value in ICT industries. *Technology Innovation Management Review, 6*(10)

Plastino, E., & Purdy, M. (2018). Game changing value from Artificial Intelligence: eight strategies. *Strategy & Leadership, 46*(1), 16–22

Prentice, C., Lopes, D., Wang, X. S., & &,. (2020). The impact of artificial intelligence and employee service quality on customer satisfaction and loyalty. *Journal of Hospitality Marketing & Management, 29*(7), 739–756

Pumplun, L., Tauchert, C., & Heidt, M. (2019b). A new organizational chassis for artificial intelligence-exploring organizational readiness factors. In: *Proceedings of the 27th European Conference on Information Systems* (ECIS), Stockholm & Uppsala, Sweden

Quinio, B., Harfouche, A., Skandrani, S. R., & Marciniak, R. (2017). A framework for artificial knowledge creation in organizations. *International Conference on Information Systems*, Seoul, South Korea

Rajput, S., & Singh, S. P. (2019). Connecting circular economy and industry 4.0. *International Journal of Information Management, 49*, 98–113

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence: Closing the gap between ambition and action. *MIT Sloan Management Review, 59*(1)

Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). Artificial intelligence in business gets real. *MIT Sloan Management Review*

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing, 36*(6), 1145–1168. https://doi.org/10.1108/IJBM-01-2017-0015

Samek, W., & Müller, K. R. (2019). Towards explainable artificial intelligence. In: *Explainable AI: interpreting, explaining and visualizing deep learning* (pp. 5–22). Springer

Schaefer, C., Lemmer, K., Samy Kret, K., Ylinen, M., Mikalef, P., & Niehaves, B. (2021). Truth or Dare?–How can we Influence the Adoption of Artificial Intelligence in Municipalities? *Proceedings of the 54th Hawaii International Conference on System Sciences*

Schmidt, R., Zimmermann, A., Moehring, M., & Keller, B. (2020). *Value creation in connectionist artificial intelligence–A research agenda*

Shollo, A., Aanestad, M., Lyytinen, K., Mahring, M., Nielsen, P. A., & Sørensen, C. (2020). Panel 4: a nordic sociotechnical perspective on artificial intelligence phenomena. *Proceedings of the 28th European Conference on Information Systems* (ECIS): A Virtual AIS Conference

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83

Tallon, P. P., & Pinsonneault, A. (2011). Competing perspectives on the link between strategic information technology alignment and organizational agility: insights from a mediation model. *Mis Quarterly, 35*(2), 463–486

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems, 37*(1), 6

Toniolo, K., Masiero, E., Massaro, M., & Bagnoli, C. (2020). Sustainable business models and artificial intelligence: Opportunities and challenges. In: *In Knowledge, People, and Digital Transformation* (pp 103–117). Springer

Vigdor, N. (2019). *Apple card investigated after gender discrimination complaints*. The New York Times. Retrieved 15/06 from https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal, 26*(7), 1893–1924. https://doi.org/10.1108/BPMJ-10-2019-0411

Wang, H., Huang, J., & Zhang, Z. (2019). The impact of deep learning on organizational agility. *In proceedings of the 40th International Conference on Information Systems* (ICIS), Munich, Germany

Wanner, J., Herm, L. V., Heinrich, K., Janiesch, C., & Zschech, P. (2020). *White, Grey, Black: Effects of XAI Augmentation on the Confidence in AI-based Decision Support Systems*

Welling, M. (2019). Artificial intelligence versus intelligence engineering. *Harvard Data Science Review, 1*(1)

Yang, W., & Meyer, K. E. (2019). How does ownership influence business growth? A competitive dynamics perspective. *International Business Review, 28*(5), 101482

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global Catastrophic Risks, 1*(303), 184

Zheng, N., Liu, Z., Ren, P., Ma, Y., Chen, S., Yu, S., & Wang, F. (2017). Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering, 18*(2), 153–179

Zuiderveen Borgesius, F. J. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights, 24*(10), 1572–1593

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994.

Ida Merete Enholm works as a fullstack-developer in Oslo (Norway) building digitalsolutions for a variety of customers. She has received her M.Sc. in ComputerScience from the Norwegian University of Science and Technology (NTNU). HerMaster's Thesis was focused towards responsible governance of AI.

Emmanouil Papagiannidis is a PhD candidate at the Department of Computer Science. In the past, he has received his B.Sc. in Applied Informatics from the University of Macedonia (UOM), his M.Sc. in Computer Science from the University of Newcastle and his M.Sc. in

Data Science from Newcastle University. Moreover, Manos has worked as a software developer in Greece (Athens) and the United Kingdom (Newcastle) building websites, managing systems and improving or fixing existence code by adding new features and introducing new technologies.

Patrick Mikalef is an Associate Professor in Data Science and Information Systems atthe Department of Computer Science. In the past, he has been a Marie Skłodowska-Curie post-doctoral research fellow working on the research project "Competitive Advantage for the Data-driven Enterprise" (CADENT). He received his B.Sc. in Informatics from the Ionian University, his M.Sc. in Business Informatics for Utrecht University, and his Ph.D. in IT Strategy fromthe Ionian University. His research interests focus the on strategic use of information systems and IT-business value in turbulent environments. He has published work in international conferences and peer-reviewed journals including the Journal of the Association for Information Systems, British Journal of Management, Information and Management, European Journal of Information Systems, and Information Systems Frontiers.

John Krogstie John Krogstie holds a PhD (1995) and a MSc (1991) in information systems from the Norwegian University of Science and Technology (NTNU), where he is currently a full professor in information systems at the computer science department (IDI). John Krogstie is the Norwegian representative and previously Vice-Chair for IFIP TC8 and was chair of IFIP WG 8.1 on information system design and evaluations (2010-2015). His research interests are information systems modelling, information systems engineering, quality of models and modelling languages, Value of big data and AI, sustainable smart cities and sustainable digitalization in general. He has published more than 350 refereed papers in journals, books and archival proceedings since 1991. H-index as of August 2021 is 53.

PAPER 2

Responsible AI Governance: A Systematic Literature Review

Emmanouil Papagiannidis, Patrick Mikalef, Kieran Conboy

(The Journal of Strategic Information Systems)

This paper is awaiting publication and is therefore not included.

PAPER 3

Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes

Emmanouil Papagiannidis, Ida Merete Enholm, Chirstian Dremel, Patrick Mikalef, John Krogstie

(Information Systems Frontiers)

Information Systems Frontiers (2023) 25:123–141
https://doi.org/10.1007/s10796-022-10251-y

Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes

Emmanouil Papagiannidis[1] · Ida Merete Enholm[1] · Chirstian Dremel[1] · Patrick Mikalef[1] · John Krogstie[1]

Accepted: 27 January 2022 / Published online: 20 April 2022

Abstract

In recent years artificial intelligence (AI) has been seen as a technology with tremendous potential for enabling companies to gain an operational and competitive advantage. However, despite the use of AI, businesses continue to face challenges and are unable to immediately realize performance gains. Furthermore, firms need to introduce robust AI systems and mitigate AI risks, which emphasizes the importance of creating suitable AI governance practices. This study, explores how AI governance is applied to promote the development of robust AI applications that do not introduce negative effects, based on a comparative case analysis of three firms in the energy sector. The study illustrates which practices are placed to produce knowledge that assists with decision making while at the same time overcoming barriers with recommended actions leading to desired outcomes. The study contributes by exploring the main dimensions relevant to AI's governance in organizations and by uncovering the practices that underpin them.

Keywords AI governance · AI data governance · AI challenges and outcomes · Performance gains · Competitive advantage

1 Introduction

As businesses adopt Artificial Intelligence (AI), they are faced with new value propositions, but they also have to deal with new challenges, such as reducing the gap between intent and action(Amershi et al., 2019; Enholm et al., 2021; Mishra & Pani, 2020). Artificial intelligence has been perceived as a tool with which we can layer many different functions or as a solution to problems that are beyond the ability of traditional applications to solve. (Smuha, 2019). In order to gain a competitive advantage over their competitors (Raisch & Krakowski, 2021), businesses have implemented and deployed AI solutions to automate their processes, increase efficiency and reduce costs (Frank et al., 2019; Gregory et al., 2020). To achieve these goals, AI governance is essential. According to Butcher and Beridze (2019), AI governance "can be characterized as a variety of tools, solutions, and levers that influence AI development and applications". Yet, further research is needed to better determine how AI Governance can be introduced into a company and whether AI governance can assist a company in achieving its objectives.

While AI has the potential to generate business value in terms of performance, productivity and effectiveness, it is not autonomous, as it works in concert with human capabilities (Zhang et al., 2021). Consequently, organizational capabilities are the results of combining and deploying multiple complementary resources within a firm to achieve competitive advantage (Mikalef & Gupta, 2021). When a firm optimizes its firm-level resources and adopts AI technological innovations, it can enhance its transformed projects' business value which drives business value and impacts performance (Wamba-Taguimdje et al., 2020). Simultaneously, the AI algorithms can be considered performative in the sense that they assist in decision-making, the extent to which their use can form organizational processes, or even take autonomous decisions (Faraj et al.,

✉ Emmanouil Papagiannidis
emmanouil.papagiannidis@ntnu.no

Ida Merete Enholm
idamen@ntnu.no

Chirstian Dremel
christian.dremel@ntnu.no

Patrick Mikalef
patrick.mikalef@ntnu.no

John Krogstie
john.krogstie@ntnu.no

[1] Norwegian University of Science and Technology, Trondheim, Norway

2018; Grønsund & Aanestad, 2020) that leads to new organization capabilities through AI. The use of AI, for instance, could create more substantial customer acquisition or higher customer lifetime value and lower operating costs or reduce credit risk.

The main goal of this work is to analyze AI governance when designing and implementing AI applications in order to achieve organizational goals. In particular, this study examines how AI Governance helps top-level managers achieve their goals by introducing robust systems that automate processes and enhancing tasks that traditionally were done by intuition or simple data analysis without negatively impacting employees. The main challenge for adopting AI in organizational operations is that AI technologies vary in scope and complexity, hindering familiarity, especially for non-technical employees (Holmstrom, 2021). Hence, it is crucial to define actions for overcoming barriers and challenges (technical and non-technical) to align AI applications to the organization's objectives. As an example, employees might resist new technologies due to fears of being replaced by AI. Based on the results, companies will be able to gain a better understanding of how AI technologies are used, identifying focal points and mechanisms of value generation (e.g., augmentation or automation of decision-making or processes) and what challenges AI technologies present to organizations. Hence, we argue that AI value realization is not yet fully understood and called for and specific governance practices may help in doing. This study, therefore, builds on the following research questions:

RQ1. Which practices underpin AI Governance?
RQ2. What are the antecedents and effects of AI Governance?

To answer the research question, we collected data through a multi-case study, conducting interviews with multiple respondents within three companies in the energy sector. The interview questions focused on methodologies companies currently use, mechanisms and processes used in AI application development, the collection of data, and the consequences of AI application in decision making (AI risk). During this multi-case study, employees from various departments were interviewed, primarily the business department and the IT department since these two departments play a crucial role when developing an AI application. We also built on secondary data sources, such as reports and internal documents, which help to explore AI governance dimensions and practices as well as compare, triangulate and verify results. Among the outcomes of the study, AI was found to be most helpful for (1) reducing maintenance costs, (2) increasing flexibility and robustness of the development process, (3) improving confidence in the results and final products, and (4) gaining a competitive edge over the competition. Lastly, we proposed a model where we discussed challenges, recommended actions, and desired outcomes.

The rest of the paper is structured as follows. The subsequent section presents the background of this study and the relevant work in the domains of technology governance, and then specifically focuses on AI governance practices. Section 3 details the methodology that is applied for gathering and analyzing the data. In Sect. 4, we present each case separately followed by a cross-case analysis. The paper concludes with a discussion of the findings and limitations in Sect. 5, where we interpret and analyze the data.

2 Background

2.1 IT and Information Governance

IT governance is an area of corporate governance that falls under the responsibility of the board of executives and focuses on the implementation and transformation of IT to meet current and anticipated business and client needs and is broader than IT management, which refers to the management of existing IT services and internal supply of IT (Saunders et al., 2020; Wilkin & Chenhall, 2010). In other words, IT Governance is a formal way to align IT strategy with business strategy. Governance frameworks for IT provide a structure (who is governed, what is governed, how is governed) for ensuring that IT investments support business objectives (Tiwana et al., 2013). Through embracing IT Governance, organizations can achieve measurable results towards their strategies and goals. However, implementing a comprehensive IT governance program requires a lot of time and effort (Debreceny, 2013).

In the digital era, information governance has an even more central role, as it promotes a more purposeful path to obtaining information. (Cath, 2018). Research previously conducted in similar areas sought to answer questions like what information governance practices are firms adopting and what are the effects of information governance on performance. According to a study conducted by Intel (Tallon et al., 2013a, b), Big data governance policies achieved the main goal of maximizing business value while minimizing technical and organizational risks related to data privacy (Tallon et al., 2013a, b). Furthermore, research studies have been conducted and supported by empirical evidence on developing AI capabilities by creating a unique set of resources that can effectively leverage investments and generate business value that leads to competitive advantage (Mikalef & Gupta, 2021).

In their empirical research, Tallon and colleagues (Tallon et al., 2013a, b) discovered that Information governance is associated with a range of intermediate or process-level benefits and many of these intermediate effects could possibly affect firm-level performance. The authors suggest a need for extending structures and practices employed in IT governance and to decompose information governance into a range of structural, procedural, and relational practices. In this paper, the structural, procedural, and relational practices are used as the main dimensions to explain how to govern information and boost firm performance (Appendix).

2.2 Governance of AI Projects

While IT governance intends to manage IT assets, hardware and software components, that assist in establishing the automation of well-defined tasks, data governance aims to manage data assets as facts having value or potential utility that are documented (Fadler & Legner, 2021). Furthermore, sophisticated forms of analytics involve artificial intelligence and automated decision-making, requiring new roles and responsibilities, but also leading to new risks. Governance should therefore not be limited to the content, but should also include its analysis, as AI should be considering a dynamic frontier of computing (Berente et al., 2021). In addition to IT and data governance, analytics governance mechanisms are needed to overcome challenges, such as the alignment among business users and analytics practitioners (Fadler & Legner, 2021). AI increasingly influences many aspects of society, from healthcare and marketing to human rights. Allowing the development of AI applications that are not under any supervision could be harmful (Chatterjee et al., 2020; Mishra & Pani, 2020); thus, it is important to promote a trustworthy AI that is lawful (complying with laws and regulations), ethical (ensuring ethical principles and values) and robust (from a technical and social perspective). For example, the use of AI in healthcare poses various issues, including a loss of privacy in health information, diminished human oversight in decision-making, and increasing prejudice across the board (Johnson et al., 2021; Trocin et al., 2021b). Governing AI projects can be interpreted differently depending on the perspective of different individuals and algorithmic management should be a concern. Because of the extent to which algorithms and the institutional frameworks allow them to get acquire management jobs to define AI's impact on key organizational processes such as delegation, coordination, and decision-making (Holmström & Hällgren, 2021).

In contrast, researchers from Microsoft (Amershi et al., 2019) approach AI governance from a technical perspective, while European Commission (EC) (Smuha, 2019) and Singapore principles approach AI governance from a human-and ethics-centric perspective. To extend this point, researchers at Microsoft (Amershi et al., 2019) have a deep focus on the technical aspects of AI. They emphasized the best practices that Microsoft teams have implemented over the years to create a united workflow that has software engineering processes and offers insights about several essential engineering challenges that an organization may face in creating large-scale AI solutions for the marketplace. According to their findings, AI governance consists of three main aspects: (1) discovering, managing, and versioning the data required for machine learning applications is more complex than a typical software application, (2) the required skills for building models and customizing them can vary based on the project, and (3) AI components might be difficult to manage if distinct modules, as well as models, exhibit non-monotonic error behavior. The European Commission's and Singapore governments' principles see AI governance as a way to promote Trustworthy AI through guidelines. Based on these guidelines, a framework has been created that offers guidance on fostering and securing ethical and robust AI. Further, the guidelines aim to go beyond the ethical principles by guiding how such principles can be operationalized in sociotechnical systems (Smuha, 2019). Fairness and explicability are key principles that an AI application must have, which can be achieved by governing data, reducing bias and collecting diverse data. Hence, AI can be trusted when making suggestions or taking decisions. Meanwhile, AI should be human-centric by safeguarding the well-being and safety of individuals. This calls for human oversight over AI with human agents making decisions and holding themselves accountable. As a result, it is argued that in the existing literature researchers investigate IT governance and data governance and they suggested frameworks or procedures for improving performance or minimizing risks caused by AI. There is, however, a gap in AI governance, which deals with both IT governance and data governance, and has a direct relationship with AI (Mikalef et al., 2020). Therefore, the literature would benefit from an investigation into how AI governance can increase organizational performance, while at the same time neglecting negative consequences of AI use.

2.3 Typologies of AI Organizational Value

The value of AI in organizations varies based on the sector and the organization's activity(Collins et al., 2021). Machine learning (ML) technologies reduce the cost of repetitive, time-consuming tasks while it enhances automation and assists with predicting events or trends. But these technologies also have the ability to bring societal inequalities into organizational processes (Teodorescu et al., 2021). Lebovitz et al. (2021) discovered a knowledge gap between AI and specialists in their research, allowing managers to better understand the risks and benefits of each technology. When the underlying information is unknown, their research demonstrates the dangers of using ground truth labels objectively in ML models; thus, the organization value that AI offers has some constrains. Furthermore, in a multi-method study that comprised an analytical model, experimental testing, and a simulation study, Fügener et al. (2021) investigated how AI counsel effects complementarities between people and AI, concentrating on what humans know that an AI does not (unique human knowledge). They observed that human judgments converge on similar responses, which enhances individual accuracy. Individual unique human knowledge decreases when the group's overall individual accuracy improves (Fügener et al., 2021). Nonetheless, as revealed in a two-year ethnographic study (Van den Broek et al., 2021) when AI economic value could not be easily realized, human engagement in the development phases remained crucial. Despite the researchers' objective to keep domain experts "out of the loop," they observed that developers and experts collaborated to create a new hybrid practices that merged ML with domain experience (Van den Broek et al., 2021). Finally, when it comes to the introduction and deployment of AI, senior executives with a comprehensive understanding of the technologies have a direct positive effect on their organizations' overall strategic direction and goals resulting in long-term economic benefits (Li et al., 2021).

3 Methodology

AI Governance in both the public sector and private sector is a set of practices that still have not been consolidated. The inadequate empirical data on mechanisms and procedures that firms deploy led us to engage this research using an exploratory, comparative case study approach that boosts generalizability while at the same time giving room for extending theory via cross-case analyses (Ramesh et al., 2017). As AI will receive more attention in the following years because of the numerous challenges it poses, we sought revelatory cases that throw light on the phenomenon for the purpose of gaining a better understanding of it (Lewis et al., 2011). In addition, there is no established framework or theoretical model that is commonly accepted by the industry and describes in detail the overall governance firms should adopt. For carrying out our multiple case studies, we followed established guidelines for case study research as illustrated by Baskarada (2014), Stewart (2012) and Eisenhardt (1989). Also, we make use of the Information value chain schema to facilitate the interplay between people, processes and technologies over the information value chain, as proposed by Abbasi et al. (2016).

Trustworthiness in the evaluation process and the findings themselves were of the utmost importance; thus, we enhanced the research methodology by strengthening credibility, dependability, transferability and confirmability (Korstjens & Moser, 2018; Sikolia et al., 2013). To ensure validity in our findings, we used triangulation across multiple sources and methods through the convergence of information. In terms of transferability, the firms have common traits and operations, but they have some key differences in their business strategy. Dependability was achieved by being consistent in the analysis process and being in line with the accepted standards. Finally, confirmability was achieved by conducting interviews with different employees in the same firm who have key positions and belong to the same or different departments. What is more, data were analyzed and coded independently by three authors bringing various insights and points of view so that the authors could identify similarities and differences in their results, creating a comprehensible and coherent framework. Hence, in order to develop a theory based on empirical data, it was necessary to establish three iterations of data analysis.

3.1 Case Selection

The selection process of the cases was conducted based on the common characteristics in respect to industry, use of AI systems, size of development teams and cultural environment. All firms operate in the same industry and have similar capabilities in terms of collecting, analyzing and interpreting data for making business decisions. The most common perspective among the selected firms is that AI must be developed, expanded and adopted in the following years as it will be crucial for gaining or maintaining their competitive advantage over rivals or new companies entering the frame and seeking a piece of the pie. Also, the nature of AI projects undertaken by firms indicates that they face similar challenges, so they require similar solutions. Comparing the selected companies is fair because (1) they are all allocated in Norway, (2) they have similar AI teams in terms of size and experience, although the size of the companies ranges, and (3) their cultural differences are limited. Therefore, choosing these three firms from the industry allows us to compare the cases for commonalities and key differences and spot how AI Governance has been implemented. Also, a generalized and standardized framework would assist companies and the state in adopting AI and planning ahead for the resources, infrastructure and necessary processes that are required. In Table 1, the cases are presented with an overview of their size, revenue and AI strategy that they follow or plan to follow in the upcoming years.

3.2 Data Collection

Conducting interviews is an excellent mechanism for gathering information, especially when the researcher does not have a priori guiding theory or assumptions (Qu & Dumay, 2011). Also, interviews can be used to refine a theory or understand a phenomenon (Tallon et al., 2013a, b). As shown in the background section, previous researchers

Table 1 Overview of companies

	Company A	Company B	Company C
Country	Norway	Norway	Norway
Sector	Energy	Energy	Energy
Employees	200	530	100
Turnover 2020	180 million dollars	260 million dollars	23 million dollars
AI Vision	Use AI to become one of the top players in the market	Use AI to increase flexibility and business capabilities	Create AI products that are customer oriented and boosts customer value
AI Technologies	Both cloud and local ML pipelines combined with intelligence dashboards – Python, Grafana	ML pipelines combined with intelligence dashboards – Python, Grafana, Power BI	ML pipelines combined with intelligence dashboards – Python, Grafana, Tableau

decompose information governance into a range of structural, procedural, and relational practices, which could be used as part of our baseline to understand how to build practices that enable AI Governance. A case study approach is chosen because it allows for in-depth analysis using interviews as generating method for collecting data. By exploring these data, new knowledge can be generated allowing for meaningful insights that explain similar situations (Oates, 2005). Also, the research is qualitative as it involves the use of qualitative data, which can be used to understand and explain the research question (Michael, 1997), as it involves the use of experiences, beliefs, and attitudes of the key respondents through the semi-structured interviews (Wynn & Williams, 2012).

Every case was initiated by contacting the human resources department or those who should have been able to handle this type of communication, for instance, managers. A brief introduction was sent via email to establish an understanding of the purpose of this research project and in some cases, quick telephone calls where necessary in order to provide some extra information. We described ideal candidates for interview as employees that (1) have a key position in the firm, for example, managers and leading developers, (2) have a good understanding of AI technologies and (3) have contributed to the overall development of AI either through their domain knowledge or their software development skills. A total of 15 individuals were interviewed, including both domain and technical experts who have worked in their current positions for at least one year, but have relative experience of at least five years. This means they are experienced, and they gained a solid understanding of AI development over time. Furthermore, participants shared how they understand specific issues, according to their own thoughts and in their own words (Pessoa et al., 2019) as members of either the business department or the IT department, as input from both departments is needed in order to understand how AI governance is designed. Table 2 shows information about

Table 2 Responders' role and length of interviews

Firm	Respondent	Role	Years in firm	Interview time
A	1	Chief AI officer	3	90 min
	2	AI Software Developer	3	55 min
	3	Machine Learning Engineer	3	45 min
	4	AI Software Developer	3	43 min
	5	Project Manager	4	49 min
	6	Machine Learning Engineer	3	35 min
	7	Machine Learning Engineer	3	45 min
B	1	Data Analyst	9	49 min
	2	Head of AI department	1	25 min
	3	Head of Data Analytics department	4.5	59 min
	4	Digitalization Engineer	10	55 min
	5	Head of Digitalization department	2	43 min
C	1	Data Scientist	2	65 min
	2	Head of Analytics department	3	60 min
	3	Operation Manager	3	60 min

the interviews, such as the firm candidates' number and their current position.

The interviews formed on open-ended questions that led to interesting conversations, where the interviewees had the opportunity to adopt their questions based on the answers or even ask questions that were not part of the interview guidelines. Before each interview, we explained to each interviewer individually what we hope to accomplish through the interviews and what we expect to be the outcome of our research, while at the same time we encourage them to add anything they believe is relevant or that we missed during the interviews. The questions were split into three categories:

1. The business value and the organizational context where we try to see how AI grew over time.
2. The data management where the interviewees were explaining how their firm deals with data services and governance practices.
3. The control and technical aspects focused on control processes and mechanisms that ensured AI systems were acting upon set goals.

Each interview lasted approximately 55 min on average, with the range being between 25 and 90 min via Zoom, which was used to record each session and then the audio was transcribed using Otter AI. The audio files were transcribed in a verbatim way so that the text remains identical to the audio, meaning that all raw data are transparent, and the findings and results could be reproduced and tracked down rigorously. As part of the process, we had to go through the text and the audio to make sure everything was looking good since we wanted our text to match the audio and the only way to guarantee that was by checking all results manually.

In addition, we used related data publicly available on the company's site (e.g., annual reports, vision and firm structure) because we consider them to have merit in our research. These documents served both as validations for our findings as well as information that we did not have prior to the interviews, assisting us to obtain a better understanding of the vision, objectives and regulations of each company.

3.3 Data Analysis and Theory Building

A narrative analysis is followed for analyzing the content from the interviews as the stories and experiences shared by employees are used to answer the research questions.

As a first step, we went through the interview transcripts and commented on our initial thoughts by writing memos. Although memos are usually used at the beginning of a text analysis, we continued to use them for updating our thoughts and interpretations or even adding new ideas. The generated transcripts were imported into the software NVivo, where open and axial coding were applied, and categories were formed based on the notation process (coding). NVivo has an add-on module called "NVivo Collaboration Cloud" allowing teams to collaborate by storing projects securely in the cloud. Two of the writers had an "administrator" role while the rest had a "workspace owner" role, so it was convenient to store, upload and update our project files. Each writer was responsible for updating his content to the cloud and the administrators reviewed the changes, but not the content, in case something went entirely wrong; for example, unintentionally deletion of a file. If the administrators were satisfied, then a merge was performed and everybody could work on the updated version of the project. Backup files were part of the process in case we lost our work or needed to go back to a previous version, so at the end of each week, a backup process was in place and the files were stored independently of NVivo.

In the first iteration, we tried to identify all the concepts related to AI Governance and the adopted practices by the firms. Initially, there were 200 descriptive codes, such as, "working with domain experts" and "domain experts lead projects" but after an iteration the number was reduced to 95, since many codes were merged into a more appropriate coding name such as "domain experts take lead of a project to ensure quality of the final product", where the combined codes become abstract.

The next logical step was to apply axial coding, where the main nodes that have been coded were procedural, relational, structural, AI development and AI challenges. In addition, comments and observations from different transcripts were combined to identify commonalities and patterns in the processes used when creating and deploying AI systems that assist firms minimize AI risks. Grouping the comments and observations, known as axial coding (Charmaz, 2014), allowed for better interpretations since the employees could refer to the same concept with similar terminology, depending based on their technical skills, knowledge, experience and position in the firm. In order to obtain a high level of confidence, researchers validated findings by examining reports, public information and presentations related to this research and focused on the AI aspects (Table 3).

Once all cases had been adequately analyzed, and the researchers had reached consensus, a cross-case analysis was performed. In the course of the discussion, we identified a number of patterns that were either similar or different and explored the reasons behind them through open discussion, trying to establish consistency and cohesion, arguing which interpretation seems most reasonable to our goal and how AI Governance is created among these cases and which practices companies should adopt or introduce.

Table 3 Nodes and possible items under each node

Dimension	Definition
Procedural	Practices associated with data migration, system messages, documentation and processes for expansion, dynamic model selection, pipeline evaluation, human and AI interaction, data quality sources
Relational	Practices that deal with employees and communicating goals, domain experts, AI education for employees
Structural	Practices associated with IT, optimization and automation, AI automation, ML pipelines, data access
AI culture	Understanding of AI capabilities, AI-phobia, Trust issues against AI
AI architecture	Development best practices, cloud infrastructure, unified tools
Legal regulations	GDPR, legal constrains of AI use
Domain challenges	Data challenges, domain knowledge, external challenges
Adoption problems	Fear of losing position
Competitive Advantage	Developing unique AI strategy, keep AI knowledge in house
Flexibility	Cloud services boost flexibility
Cost maintenance	Minimize costs from various operations
Scaling up	AI assists in scaling up without needing more resources
Superior AI results	Internal AI teams can give high value through solutions that are targeted in a specific problem and not generalized

4 Case Analysis

4.1 Within Case Analysis

All cases have some commonalities in their characteristics and practices. Firstly, all the cases operate in the same industry and have overlapping areas of operation. Secondly, development best practices were followed such as the use of Git, documentation and containerization platforms like Docker. Thirdly, data privacy (GDPR) is not a genuine concern (expect in the last case) since their data mainly consists of environmental data that anyone could access or buy, while legal regulations restrict them to using AI in specific areas, for instance they are not allowed to speculate on prices. Lastly, the set-up goals mainly concern reducing costs and forecasting energy demands.

4.1.1 Company A

Company A is a Norwegian company in the energy sector using environmentally friendly production and energy-related services. The main focus is on the areas of hydropower production and wind power production, meaning the center of attention is on developing renewable solutions that supports positive societal development. The company trades in different markets by forecasting how much energy is projected to be consumed each day known as intraday, while being actively involved in planning for hydropower plants. Hydropower plants are a controlled energy source that the owners can decide how much energy they want to produce, compared to wind energy that is affected by environmental variables. In this sense, optimization plays a key role.AI contributes to the reduction of predictive maintenance costs, which is challenging in Norway due to its harsh weather conditions, especially during winter.

As part of its strategy, the company developed an AI team internally and adopted or utilized cutting-edge AI technologies and techniques more extensively. A small group of recently hired developers forms the AI department and becomes part of the business development and innovation team of the company. Among the reasons for that decision was the belief that the company cannot maintain a competitive advantage without using AI in the upcoming years, and eventually, larger corporations will absorb them. The AI team brought value to the firm by forecasting energy consumption, assisting in decision-making for the end-users and automating repetitive tasks. As a result, performance was boosted and maintenance costs were down.

Control of key domain knowledge was one of the main concerns for firm. Company A did not want to give away domain knowledge to external partners, who offer specialized AI products, since they could build and sell similar AI products to their rivals:

> "If we help them (the software company) develop their software, they will take this software where we provide the data, we provide domain knowledge and sell it to everyone, especially to our rivals". (Respondent 1, Company A)

The development team aimed for automation and flexibility but they did not want to develop the entire software from scratch since it would be time-consuming to do everything. At the same time, they did not prefer to use software of other companies, so they decided to develop the intelligence that runs on top of cloud services (boosting flexibility at the same time) despite the fact that using cloud services was challenging in the beginning:

> "The real challenge was not to deploy a single model but a whole cascade of models that were dynami-

cally selected between each other". (Respondent 3, Company A)

Standardization and unification of AI technologies was an issue because the team is consisted of people from different backgrounds and with different skills creating obstacles in AI development. The problem occurred because each member of the team had his preferences about which tools and style should use during development time, making it difficult to exchange or understand others' code since the system was not unified. The team decided through internal workshops to unify the used tools (e.g., programming languages, databases) while creating a shared vocabulary through collaborative wiki pages:

> "We were responsible for our own code. There was no code sharing, there were no shared tools that people can use amongst each other as a team, because everyone else was doing their own thing". (Respondent 2, Company A)

In the beginning, data was exchanged through Excel files. These files were not secure, and at the same time they realized that they could not scale up, so APIs were used to replace Excel files. The necessary data was collected through vendors, so it was possible to compare data and ensure high quality outcomes for the trained models. To increase security, data access was only possible through intranet, but the company did not define clear data management roles, making the data request process time consuming:

> "You're getting data from somewhere, and the data for some reason, you don't have access at that particular time. And that that's something that pops up multiple time. You can of course, try to get around, having some to wait a bit, and you know, retry". (Respondent 4, Company A)

Multiple steps were taken into consideration to achieve robustness and reliability. To govern the process of data cleaning and model evaluation, ML pipelines were created in the cloud. This made it easier to oversee the overall process and apply quantifiable metrics on the ML results. Also, domain experts participated in the evaluation so they can provide their insights and feedback to make the model outputs reliable and trustworthy. Rather than increase profit margins, the model outputs emphasize reducing errors, because Company A places higher priority on prediction safety instead of profitability. In case of failure, local systems (ML pipelines) were ready to support decision-making, ensuring a reliable and robust system that could always generate output and assists employees with their everyday tasks:

> "You still need to have an option to run them, not on the cloud solution itself, but on your local system. So basically, we do have these kinds of processes, in case something fails, because things fail much more often than you would think". (Respondent 2, Company A)

Domain experts manage the projects as their knowledge and expertise are needed at each step of the development phase. For example, their insights could determine, which data should be needed for the machine learning models. In addition, domain experts help with the creation of meaningful dashboards that are responsible for alerting information to employees, explaining historical data and assisting in decision making for end users. At the same time, developers focused on alerting errors and failures, for instance, if a data stream stopped delivering data. Another way to ensure robust outcomes after deployment was to test the models against real-time datasets. Through This, they were able to make adjustments to the models, obtain a better understanding of the data, and improve the overall quality of the system:

> "When an incident happens, usually the ones who have developed the system and some stakeholders from the rest of the organization, they sit down and sort of meet … and they questioned what happened, what was the consequences, and then the developers go into find out the reasons for that". (Respondent 5, Company A)

Due to radical changes in processes and operations AI training for end-users was more than necessary. All these changes caused human agents to feel phobic when interacting with the machine, as they had the overall watch and check periodically that everything is in working order. From the employees' point of view, these automations raised concerns as they saw themselves being automated and driven away from their posts, which could result in unemployment:

> "People get scared of the fact that we will automate them away. So, we had a hard environment. We started talking about why we need the people here, their domain knowledge … so we had regular meetings explaining what AI can do and not". (Respondent 1, Company A)

To summarize, company A built AI capabilities to automate procedures and assist with decision-making by using cloud services, ML pipelines and domain experts to understand data and the outcomes of models. Flexibility, productivity, and reduction of costs were the immediate effects that the company saw as positive results allowing them to remain a competitive player while achieving their set up goals of their overall AI strategy.

4.1.2 Company B

Company B is a Norwegian renewable company that focuses on customers' needs by producing and distributing clean and renewable energy. The company's management believes that future energy consumption will differ from what it is today in many ways. Energy customers will produce their own energy and they will want to have the opportunity to combine this with smart energy solutions, meaning that the customer will more than ever be at the centre of attention where he will play a small but still significant role in the production of energy. The firm understood that the adoption of AI is vital for creating new products and services that will make them a leading provider of competence services.

Data analysts performed data surveys to evaluate which data they think to be the finest and most suited for their purposes. Within the last five years, the firm has hired a couple of analysts with machine learning experience and they have begun developing AI models in conjunction with domain experts. To build the AI capabilities data were gathered internally and externally from various vendors as it needed to verify and ensure the quality of the data since it is crucial for the AI models:

> "We have a data survey, to make sure that we have the right data for what we think would do the job. And then we build the model". (Respondent 1, Company B)

Reducing maintenance costs and errors, as well as creating flexible systems that can scale, were all top priorities. Initially, the team used cloud services, but they were not flexible enough, or at least to their liking, so they moved to influx databases that allow storing and retrieving time series data. By contrast, a containerization platform like Docker was adopted from the start to let developers to package applications into containers. Thus, these standardized executable components boosted flexibility and the cloud services were put aside. With the help of the IT department help new tools and processes were introduced to detect early problems and warnings by using different types of sensors. Based on these inputs, autoregressive (AR) models were developed to detect anomalies in the system, saving time and effort, which means fewer maintenance costs in the long run:

> "We have audio surveillance, to monitor and detect early problems with just sound and then we have the AI model. It is listening to the sound and try to detect early warnings". (Respondent 4, Company B)
>
> "We had a cost of around two million a year and it has been reduced to around ten thousand a year, almost nothing". (Respondent 3, Company B)

Nevertheless, it is expensive to add many features and takes a lot of time to develop. Despite the use of ML applications with neural networks, all the applications are considered to be weak AI (AI that is limited to a narrow task). Because of that the company still uses conventional and traditional ways in parallel with AI, while they plan to replace them over time in the future:

> "We have used this technology started with basic AI … using more machine learning and neural networks and so on and that has only been around for two years, but it was a strategic decision". (Respondent 5, Company B)
>
> "It's always a question of cost them money… so that's, maybe that's why we use Excel for many processes, because it's, it's very easy to set up and when you have set up something that works, and you have to pay in order to replace it". (Respondent 4, Company B)

Another challenge that the developers faced came from employees who refused to use the new technologies as they did not trust the results or even oppose the change. Although the AI works as assistance in most cases and helps with decision making, the employees could not accept that a new member of the firm that has no experience in their field could improve their work significantly:

> "I've got some feedback from people that "you can't come here and tell me what to, how to do it. I worked here for 20 years with the same thing". So, they are there are scared of me doing their job better, I think". (Respondent 3, Company B)

Nevertheless, when people start using the applications, they misunderstand the AI capabilities. End-users had unrealistic expectations of what the model could or should predict, and the developers spent many hours explaining what a statistical output is and how the model actually make predictions. Furthermore, they elaborated on what is possible and what is not doable, which took a long time for the end-users to digest all these new information and the training process lasted for months.

Last but not least, the data administrator is a straightforward process because there are only two roles primarily, one administrator who can perform all actions (e.g., write and read), and one reader who can only read specific data as part of their work. This simplicity in roles and the fact that they do not deal with private data in their applications led to the decision to not have a dedicated employee responsible for data management.

To sum up, company B uses AI as a tool for prediction for identifying market opportunities and reducing maintenance costs. To accomplish this, a small team of AI developers was formed, who introduced new technologies and processes with data from various vendors. The complexity of the system was kept low to prevent high development costs while the end-users were introduced to AI capabilities to enable them to trust and adopt AI in their daily work.

4.1.3 Company C

Company C is a firm that identifies itself as climate-conscious, where they assist their customers through digital technologies to reduce energy consumption. Their services cover many aspects such as charging devices and heating in the home, which is appealing for many people as their services assist in saving a considerable amount of money every month. Company C realized that there was a big gap in the market since energy-producing companies did not offer any customized services. Hence, they decided to adopt AI practices to build the necessary capabilities to create customized applications for each client. A direct effect was that customers came with constructive feedback driving the firm to become even more efficient and building new services that were highly demanded:

> "Every time a customer approaches with a question, we take those questions. And let's say a customer just comes and says like, I would like to control my water boiler at home, and I can't, and I am spending a lot of money on this. So, I would like you to improve that." (Respondent 1, Company C)

Building these AI capabilities though would be impossible if the company did not follow best development practices. In addition, cloud services are used to cover areas that the members of the development team have no expertise or the time to develop:

> "We would need to build our own data centers, which is completely out of our expertise, we would need to hire people and know how to distribute the load, then you need to secure your data etc." (Respondent 1, Company C)

To ensure robustness, the development team has created procedures that covers extensively any AI behavior change and when the timeline that these changes are allowed to be published, for example, not before a big event, in production to avoid AI failures. AI unit tests are also in place to ensure the system's outputs are reliable. To gather the data for their AI models, Company C uses APIs from different vendors. As previously mentioned, the firm uses private data, so a dedicated team was formed to deal with privacy issues by introducing procedures during the data transformation and data storage phases:

> "We have a team in the company that it's exclusively focused on privacy, and how to comply with the regulations." (Respondent 1, Company C)

Nevertheless, data roles and data management were not always in the spotlight as almost all employees could access data since the company's size was small. The growth in numbers led to the decision of introducing data management roles and restrictions on the data types and situations under which employees can access data. This was accomplished through data-gates where employees had to ask for permission from the supervisor of the system:

> "If they need to access that data, they will need to request it from their supervisor for example. And then it depends on the type of data that you use, what data you get access to, but I would say like data scientists and developers usually we get access to basically everything because we work on everything." (Respondent 1, Company C)

The AI applications focus on specific needs, which usually involve forecasting ancillary services, customer needs (AI assistants) and reducing maintenance costs by minimizing business risks at the same time. To ensure trustworthiness and confidence in their provided services, the team has implemented ways for explaining AI decisions (XAI) which allow customer service employees to communicate efficiently with customer requests that involved AI decisions or AI suggestions:

> "The machine taking decisions and that the customer wondering why the machine took the decision and asking support for this. And then we need to tell them why the machine took this decision." (Respondent 1, Company C)

It is worth mentioning that Company C never experienced any problems related to AI fear since all employees have a good understanding of what AI can offer and how it helps them in their everyday lives. Two could be the main reasons for that. Firstly, employees have an extensive onboarding training process and secondly, people who applied to the company are aware that the firm uses extensive AI products; thus, work candidates have prior knowledge of AI technology and AI products or are willing to embrace AI.

4.2 Between Case Analysis

The interviewees talked about how their company transformed over the years and the necessary steps that were taken in order to expand and maintain a competitive advantage, while minimizing AI risks. In Table 4 there is a sample of the grouped observations that are generated based on the interviews.

4.2.1 Procedural

As far as the procedural practices are concerned, all firms aimed to build new capabilities using external software. Algorithms, trading strategies and machine learning pipelines are developed by the internal AI teams, using platforms from third partners, keeping domain knowledge in house.

> "We try to build all by ourselves. We do not want third parties to build what we can because they can use the same software for different purposes". (Respondent 2, Company C)

Table 4 Nodes and grouped observations (sample) based on the interviews

Node	Observations	Code
Procedural	Having a backup [offline] AI model is recommended	Backup offline ML pipelines
	Use AI platforms mostly for deploying models	Build intelligence on top of external AI services
	Correct the source data not the cleaning process	Data quality sources
	understand concepts not just data	Data quality sources
	Create dashboards for monitoring actions and results	Enable human—AI interaction
	Create AI products that do one task	Create weak AI applications
	Ensemble models to maximize the output	Dynamic model selection
Relational	Onboard training processes	AI education for employees
	Operators should understand what the model is (and not) capable of predicting	AI education for employees
	Read data from different vendors to increase quality of model	Data vendors
	Domain experts take lead of a project to ensure quality of the final product	Domain experts lead projects
	Hire external consultants to predict the value of the project or help with specific cloud technologies	AI consultants
	Explain to customers AI decisions	Explainable AI
Structural	Automate operations that take place 24–7	AI Automation
	Repetitive and boring tasks should be automated	AI Automation
	AI solutions that focus on a very specific problem perform much better than generalized AI solutions	Locus of AI strategy
	Allocate required resources and create plan for AI development	Locus of AI strategy
	Access data through intranet for security reasons	Intranet data access
	No clear roles who is responsible for data management	Data ownership responsibilities
	Data transformation process has been standardized	ML pipelines

For all projects, data governance, data quality and data security are common elements to ensure quality and security. All firms attempted to fix potential issues in the data sources, through data collection corrections and the use of APIs, instead of extending their cleaning process:

> "We do not do much cleaning of the data; we are focusing on getting it right." (Respondent 4, Company B)

The evaluation of ML pipelines was a continuous process that took place at different points of the pipeline for ensuring robustness and quality. The outcomes of the pipelines were AI products that are considered to be weak AI, executing singular tasks or providing with suggestions for decision making (AI assistants). Nevertheless, the end-users had to follow the AI suggestions intuitively and use their domain knowledge to fill gaps that AI was not capable of. In addition, intelligence systems should include notification systems, error detection and decision-making tools. By doing so, firms measure the credibility of their systems and evaluate the performance gained through KPIs:

> "We need to always monitor the quality measures and always be on our toes and improve that." (Respondent 4, Company B)

4.2.2 Structural

As for the structural practices, AI strategy for current or feature development projects seems to be the centre for top managers as they need to design products that focus on specific needs, while adding business value. Also, managers need to allocate the right resources and plan precisely as the costs and timelines for AI projects do not follow the usual software projects:

> "We need to plan and decide how long it takes, these are the resources that we need to do it, and this is the plan, and then we will go through a decision." (Respondent 1, Company A)

Managers had to separate the nice to have features that were often requested by either clients or employees. Otherwise, these requests could delay considerably the project and skyrocket the cost of development leading the project's failure. A note of caution is that AI development is usually more expensive than traditional software development:

> "It depends on the available resources and time; it is really costly to add a lot of AI functionality. We would definitely like to have them, but it is not feasible". (Respondent 3, Company C)

Managers could estimate the for building a pipeline based on the project specifications. It is common in AI projects to reuse parts of one pipeline for another, which reduces the overall development time considerably. At the same time, pipelines provide confidence in the quality of the end result as the final product is robust, easily maintainable and extendable for new features.

> "We have all kinds of pipeline, for example, usually, we have basic, like getting the data as a first step and we do some preprocessing. Then we do feature selection, building different models, compare the performance etc.". (Respondent 6, Company A)

Data management practices consider mostly securing data, using secure databases and intranet access, and creating a few roles for data access, where usually there are two types of roles, (1) developers with full access and (2) end-users with access to specific data:

> "There is a there is a shift now. So, if you work with data, you will get access to that data ... before everything was open …, and we needed to implement these restrictions." (Respondent 1, Company C)

4.2.3 Relational

In all cases, domain experts were involved in all development phases for two reasons. Firstly, their domain knowledge was crucial to the success of the project, and secondly, they led the projects as project managers. Also, with the help of AI developers they built notifications systems, by declaring which notifications should be sent via email and which should be displayed in dashboards:

> "If something (bad) happens, we get a warning to our email. Then we can find the bugs or look more on tools and see what happened in there and fix it." (Respondent 6, Company A)

External AI consultants assisted only at the beginning, and they were only called on in rare cases when the development team was unsure how to proceed with a particular project:

> "We had consultants for cloud services that we weren't familiar with and for some ML optimizations". (Respondent 3, Company B)

Lastly, establishing an AI culture inside the firm through extensive training was not an easy process, especially true for the two first cases. Employees did not trust the outcomes, sometimes they described the recommendations as naive, and most importantly, employees saw AI and automation as a way of losing their status and position. This direct threat, as they experienced it, was handled by many workshops and internal meetings.

> "We explain to them that we are going to help them; we're not going to automate them away, and I talked quite a lot about this, when I explain sort of what we were doing and how it was going to work. So, taking away this fear that we were coming from the outside as aliens and our work is to identify patterns (basically) it helped a lot". (Respondent 1, Company A)

What is more, AI teams explained what AI is all about and how it works because most employees who started using AI as assistant in their decision-making processes misunderstood very often AI's ability to predict certain patterns (especially true when AI models were updated):

> "You need to ensure that model operates in a way that the operators understand and they agree with how it was developed … allow an operator to make changes to the decision, what you often see is that the performance gets much worse." (Respondent 6, Company A)

4.2.4 Enablers and inhibitors

Firms encounter various enablers and inhibitors when they innovate their business model. One of the main enablers for AI governance is unification in the choice of technologies and infrastructure because there are different tools for developing AI products. For example, Company A had legacy code written in different programming languages making compatibility among applications an issue. The need to unify and standardize the set of used tools was more than a necessity:

> "Developers were programming in MATLAB or Python, and everyone was doing their own thing". (Respondent 4, Company A)

Furthermore, it became essential to increase the speed of models and scale up because the company increased the amount of data while creating new intelligence based on the data. These changes were boosting efficiency and employees liked automation that lifted the heavy load of the work:

> "One of the big changes and additions that everyone started programming, and automating stuff is that we went fully on cloud in all our systems, and it enabled us really be very flexible with our resources". (Respondent 2, Company A)

AI culture promotes the acceptance of AI, meaning that employees use and trust AI. The lack of AI understanding could lead to AI phobia, which is a huge inhibitor when digital transformation process is in place. Another inhibitor could be lack of domain knowledge or lack of data for creating business intelligence. On top of that legal regulations forbid certain uses of AI, for example, the prediction of energy prices.

4.2.5 Outcomes

The outcomes were similar in all cases. That could be because their desired goals were similar. The need for reducing maintenance costs and forecasting energy consumption were the top priorities since most of the business value come from these two outcomes:

> "It is similar in other industries. It is said that they are estimating, based on big data, that reduction in maintenance costs is about 20%—30%". (Respondent 4, Company B)

Flexibility and robustness were products of the development process as their AI systems have to be able to adapt and estimate market trends. As an example is Company C, which strives to understand its customers' habits so it can adapt to each one, while at the same time, AI decisions should be robust and non-costly for the customer to use:

> "If there is a break, and someone wants to charge his car, and then start heating up… in an hour that price is high, then the cost would be pretty high… the customers is going to be angry." (Respondent 2, Company C)

This superiority in results, boosts confidence in decisions and the potential customer value is high, especially for firms that have a more direct relationship with their clients. As a result, companies gain a significant competitive advantage over the competition as they can reduce the overall product cost and provide clients with exceptional services that adopt in their specific needs.

Table 5 shows challenges and recommended actions that firms faced and followed collectively in order to achieve desired outcomes.

A proposed model is constructed based on the foregoing discussion. Our model (Fig. 1), which includes the structural, procedural, and relational components as key components, illustrates the techniques that companies have used over the last five years. Enablers include existing AI culture and architecture within a company, whereas inhibitors are mostly legal constraints, domain challenges, high development costs, and AI-phobia. Companies that seek to use AI should ensure that these problems have been examined and addressed in advance, since numerous impediments can lead to failure and waste of company resources. The model's most essential results are a competitive advantage, cost reduction, and dependable AI systems, all of which are critical to any business's success, particularly in competitive marketplaces.

5 Discussion

In this study, we set out to explore the underlying activities that comprise an organization's AI governance. Specifically, we built on the prior distinction between structural, relational, and procedural dimensions of governance in order to understand how organizations are planning around their AI deployments. Through a multi-case study of three organizations that have been using AI for several years, we conducted a series of interviews with key respondents and identified a set of activities that were relevant under each of the three dimensions, as well as challenges they faced during deployments of AI and how they managed to overcome them. Our analysis essentially points out the various obstacles that AI governance is oriented to overcoming, and the mechanisms employed to operationalize them.

Specifically, we find that the obstacles that are identified during the process of deploying AI are observable at different phases and concern different job roles. When it comes to difficult management responsibilities that a business owner must do, AI solutions can always provide a variety of responses and probabilities for each of these alternatives. However, AI lacks the ability to make decisions in specific contexts. To make the ultimate decision, a business owner or manager must employ intuition to reconcile the choices provided by AI (Kar & Kushwaha, 2021). In addition, they span various levels of analysis, from the personal, such as fear of AI and reluctance of employees to adopt it, to organizational-level ones, such as organizational directives on how to comply with laws and regulations. What is more, the study reveals not only that AI governance is a multi-faceted issue for organizations but that it spans multiple levels, therefore requiring a structured approach when it is deployed. In addition, different concerns emerge at different phases of AI projects, so AI governance also encapsulates a temporal angle in its formation and deployment.

The significance of governing AI can be critical in attaining digital innovation. The firms we looked at were leveraging AI to help them reinvent their operations. Instead of having an information collection approach, these firms followed an information analysis approach. Information analysis refers to the opportunity of developing unbiased approaches for evidence-based data analysis (Trocin et al., 2021a), where AI can foster digital process and service innovation as companies did in this study. Also, AI has the potential to foster a digital innovation process by developing new and evidence-based approaches for data collection (Mariani & Nambisan, 2021). First, it enables organizations to modify particular parameters to appeal to a wider audience when content is released online, and second, it allows them to gather online behavioral

Table 5 Challenges, recommended actions and desired outcomes

	Challenges	Recommended actions	Outcomes
Development	AI cloud is challenging to build	Offline recommendation system Develop intelligence on top of external platforms	Boost flexibility
	AI development does not follow necessarily traditional software development	Standardized executable components Unify technological tools Create shared libraries	Robustness Reduce amount of workload
	Prediction techniques vary based on sector	Allow human interaction in high uncertainty to prevent high AI bias	Robustness
	Lack of data	Choose AI algorithms based on data volume and data types Generate data from existing data Read data from different sources Buy data from vendors using APIs	Boost flexibility Robustness
	Lack of domain knowledge by AI developers	Allow domain experts to lead	Save money and time Robustness
Employees	Misunderstand of AI capabilities	AI training to understand what the models can do and what cannot do	Better communication between departments Easier adoption of AI
	Employees do not adopt AI	AI training to understand how to use the new technologies	Better communication between departments Easier adoption of AI
	Employee's fear losing their position because of AI	AI training to explain why their expertise cannot be replaced	Better communication between departments Easier adoption of AI
	Different vocabulary for different departments	AI training to be familiar with different terms and processes Create different dashboards for different concepts	Better communication between departments Easier adoption of AI Measure performance
Value	Classical optimization tools are still better than AI models	Automate operations that 1. take place 24–7 2. there is a 1–1 correlation between workload and number of employees 3. are repetitive and boring document code and process	Save money and time Scaling up becomes easier Reduce amount of workload
	Hard to predict effort and costs	Avoid nice to have features as they will delay the whole process considerable use KPIs to quantify performance	Save money and time Scaling up becomes easier
External environment	Giving out knowledge to external partners	Develop intelligence on top of external platforms instead of using external solutions	Maintain competitive advantage
	Distance with third parties can affect development	Develop internal AI team to speed up processes considerably	AI Development is focused on your specific problem not to a generic solution maintain competitive advantage
	Legal constrains and GDPR	Create clear data management roles	Security

data and store it for a set period of time (e.g. one year) in accordance with GDPR regulations (Trocin et al., 2021a). It is worth mention that emotional intelligence is not part of these systems although understanding how people deal with emotional challenges is crucial for AI systems to emulate human reasoning (Luong et al., 2021). Finally, the COVID-19 pandemic has introduced new challenges and opportunities for digital transformation and innovation. For example, the United Kingdom intends to employ health information technology and execute proposals for a national learning health and care system as a result of a serious public health shock. Hence, each UK country's digital health and care strategy should be re-evaluated in light of the pandemic's lessons (Sheikh et al., 2021).

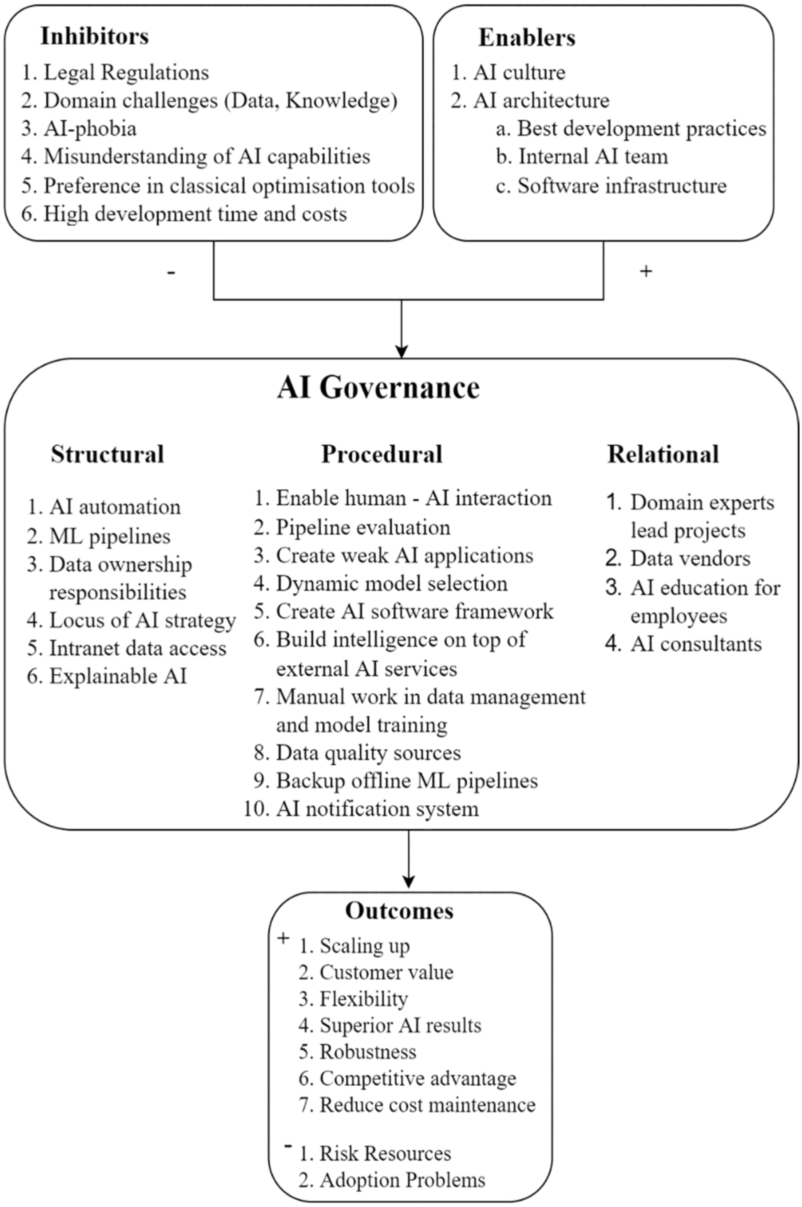

Fig. 1 Proposed model

5.1 Research Implications

This study contributes to IS literature. Despite the considerable debate in the scientific community about what is considered AI and how companies should incorporate AI in their everyday operations, we tried to understand the processes firms use to govern AI. However, not all companies have managed to build AI solutions that have had significant organizational effects and resulted in added business value. In this article, it is argued that although it is important to adopt AI, it is equally vital to create the necessary processes and mechanisms for developing and aligning AI applications with the requirements of the business environment. One of the main challenges we identify is that AI governance requires continuous adaptation and modification as new data emerges or conditions change, for instance how employees perceive AI. Thus, there is a form of ephemerality which places an increased focus on establishing processes, mechanisms, and structures to ensure that it is functioning as required and that it aligns well with the goals of the organization.

Furthermore, there is a multitude of angles that a firm can approach AI governance; for instance, companies in this study tried to create ML pipelines and interactive dashboards, but not all of them had a real focus on explainability of the results since they are still in early stages and focus on parts that they believe are more urgent. In the industry there is a recent article by Microsoft, which focuses primarily on the technical aspects of workflow implementation, outlining the key phases in the lifecycle of machine learning applications (Amershi et al., 2019). Yet, this research concentrates on the development challenges and the practical solutions a firm could follow to build an AI through solid and effective organizational practices. In this sense, AI governance in this article is not seen as a process but as a set of important aspects that need to be considered when designing and deploying practices and mechanisms, in order to ensure that the main challenges are overcome successfully and that AI applications are operating as planned. Our proposed model suggests that although there are inhibitors and barriers and despite the different ways of approaching AI governance, it offers positive outcomes, if best practices are followed, and this study identified specific procedural, structural and relational components that are necessary for achieving this.

Our exploratory work opens up a discussion about what AI governance comprises of, and how it can be dimensionilized. Furthermore, it explores the link between the challenges such governance practices help overcome, and the actors and practices they involve. This stream of research is particularly important in the value-generation of AI-based applications, as it paints a more detailed about how relative resources are leveraged in the quest for business value (Mikalef et al., 2019). In addition, the work sheds some light on the process-view of AI deployments by opening up the dialogue about the different phases of AI deployments and the unique challenges faced within each of these.

5.2 Practical Implications

Based on the findings, a firm needs to incorporate new procedures when adopting AI in order to maintain an advantage over the competition and boost efficiency. A unified system is required for building AI pipelines, which is consistent with the tools that developers use. Hence, the system will be more robust as it will be easier to maintain and improve different components of the system. In addition, managers should create procedures that employees are aware of and follow and give clear guidelines; otherwise, time and resources might be wasted, which could be invested in other projects that would add more business value.

Firms should use AI for automating tasks that are repetitive, which is appreciated by employees since they do not want to do monotonous work, but at the same time managers should have extended conversations with employees of other departments ensuring them that AI will not replace them (AI education). This could be crucial for the company’s internal stability as people might lose trust in the leadership, they might leave the company taking their expertise with them or resist using new technologies and try to undermine the value of AI.

Lastly, firms can use dashboards as an effective way to allow communication between human and machine. Dashboards are a great information management tool that is used to track KPIs, metrics, and other essential data points relevant to a business. That way the black-box nature of models and AI in general can be less problematic because the use of data visualizations simplifies complex data sets and provides end-users useful information that can affect business performance. In other words, humans will be able to evaluate results and detect any outliers or anomalies in processed data. This in turn facilitates greater transparency and a more direct way of revising the models used to analyze data.

5.3 Limitations and Future Research

In the current work, we investigate how to govern AI, which practices should be adopted and how to minimize AI risks. However, there are certain limitations that characterize this research. First, the data are collected through interviews with companies that do not require extensive use of sensitive data; thus, there might be bias in our data or provide an incomplete picture of the entire challenges around relevant practices. Second, while we conducted several interviews with key employees within the organizations, our data collection was based on a snapshot in time and may not accurately reflect the complete breadth of practices. Lastly, all cases are from the same sector. Hence, generalizability could be an issue that should be taken into consideration.

As future research, it would be interesting to gather more empirical data through interviews, from firms that belong to different sectors, and theorize the notion of AI governance from a positivist perspective, which could be tested with empirical data on the antecedents and their effects. It would also be beneficial for the field to know which resources firms deploy most in order to achieve their organizational goals and how they govern these resources to boost their performance, and how AI governance practices impact specific types of resources.

Appendix

Interview Guidelines

Introduction

1. What is your current role and background in the company?

Business value/organizational context

1. Could you mention briefly the history behind AI use in your company? How long it took you to adopt AI (timeline)?
2. How did (i) the use of AI grow over time, (ii) how did the AI team grow over time (iii) how did the value/effectiveness of AI grow over time?
3. Are there any changes brought by AI that you did not anticipate?
4. Do you plan to use AI in other aspects of your company?
5. Do you prioritize reducing risks or potential margin profits and why?

Dealing with Data

1. Do you deal with Data Privacy?
   1. If yes, how do you do that?
   2. If not, why not?
2. How do you handle data?
3. Could you describe the cleaning process?
4. Do you use cloud services?
   1. If yes, then what type of server do you have?
   2. What about external services like Azure?
5. How are your organization's models audited for security or privacy vulnerabilities?
6. Do you follow any best practices for Trustworthy AI?
   1. If yes, which one?
   2. If not, why not?
7. Which people have access to your AI features? Describe the main roles.
8. Have you established any governance practices? For example, have you defined roles and responsibilities?
9. Who is in charge of the data management and what were the requirements for that position?
10. Have you quantified decision bias in your company's model predictions?
11. Could you describe the infrastructure of your system?

Control and Technical Aspects

1. What types of data do you collect? How do you ensure to use data and AI algorithms such that they are in line with your organizational objectives?
2. Are there any procedures or processes for managing the data you use in your organization (for AI purposes)?
3. Where is data stored? How is it shared etc.? (In what cloud service are data stored?)
4. Do you specify, monitor and evaluate the (i) behavior and (ii) outcomes of your AI systems and potentially the combination with human decision-makers? Which actions are taken upon this?
5. Which control processes and mechanisms are in place to ensure that AI systems are acting upon your set goals? Does this differ depending on the use cases?
6. What processes do you have to ensure robustness?
7. Do you develop any kind of internal AI software framework?
8. What development practices do you follow as a team?
9. Did you try to incorporate external AI software?
10. What practices have you adopted to ensure scalability?

Acknowledgements Acknowledgements have been removed for the purposes of the review process.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Declarations

Conflict of Interest We declare that:
No funding was received for conducting this study. Also, the authors have no financial or proprietary interests in any material discussed in this article.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2), 3–32. https://doi.org/10.17705/1jais.00423.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP).

Baskarada, S. (2014). Qualitative Case Study Guidelines. *The Qualitative Report*, *19(40)*, 1–25. https://ssrn.com/abstract=2559424.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/misq/2021/16274.

Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? *The RUSI Journal*, *164*(5–6), 88–96.

Cath, C. (2018). *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*. The Royal Society Publishing.

Charmaz, K. (2014). *Constructing grounded theory*. Sage.

Chatterjee, S., Ghosh, S. K., & Chaudhuri, R. (2020). Knowledge management in improving business process: an interpretative framework for successful implementation of AI–CRM–KM system in organizations. *Business Process Management Journal*.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383.

Debreceny, R. S. (2013). Research on IT governance, risk, and value: Challenges and opportunities. *Journal of Information Systems*, *27*(1), 129–135.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, 1–26.

Fadler, M., & Legner, C. (2021). Toward big data and analytics governance: redefining structural governance mechanisms. Proceedings of the 54th Hawaii International Conference on System Sciences,

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, *28*(1), 62–70.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., & Moro, E. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*(14), 6531–6539.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *Management Information Systems Quarterly (MISQ)-Vol*, *45*(3), 1527–1556. https://doi.org/10.25300/misq/2021/16553.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2020). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, *46*(3), 534–551. https://doi.org/10.5465/amr.2019.0178.

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, *29*(2), 101614.

Holmstrom, J. (2021). From AI to digital transformation: The AI readiness framework. *Business Horizons*.

Holmström, J., & Hällgren, M. (2021). AI management beyond the hype: exploring the co-constitution of AI and organizational context. *AI & Society*, 1–11.

Johnson, M., Albizri, A., & Harfouche, A. (2021). Responsible Artificial Intelligence in Healthcare: Predicting and Preventing Insurance Claim Denials for Economic and Social Wellbeing. *Information Systems Frontiers*, 1–17. https://doi.org/10.1007/s10796-021-10137-5.

Kar, A. K., & Kushwaha, A. K. (2021). Facilitators and Barriers of Artificial Intelligence Adoption in Business–Insights from Opinions Using Big Data Analytics. *Information Systems Frontiers*, 1–24.

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really "true"? The dangers of training and evaluating AI tools based on experts' know-what. *Management Information Systems Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/misq/2021/16564.

Lewis, M. O., Mathiassen, L., & Rai, A. (2011). Scalable growth in IT-enabled service provisioning: A sensemaking perspective. *European Journal of Information Systems*, *20*(3), 285–302.

Li, J., Li, M., Wang, X., & Thatcher, J. B. (2021). Strategic directions for AI: The role of CIOs and boards of directors. *MIS Quarterly*, *45*(3), 1603–1644. https://doi.org/10.25300/misq/2021/16523.

Luong, T. T., Sivarajah, U., & Weerakkody, V. (2021). Do agile managed information systems projects fail due to a lack of emotional intelligence? *Information Systems Frontiers*, *23*(2), 415–433.

Mariani, M. M., & Nambisan, S. (2021). Innovation analytics and digital innovation experimentation: the rise of research-driven online review platforms. *Technological Forecasting and Social Change*, *172*, 121009.

Michael, D. M. (1997). Qualitative Research in Information Systems. *MIS Quarterly Executive*, *21*(2), 241–242.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), 103434.

Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019). Developing an artificial intelligence capability: A theoretical framework for business value. International conference on business information systems, Seville, Spain.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361. https://doi.org/10.1016/j.im.2020.103361.

Mishra, A. N., & Pani, A. K. (2020). Business value appropriation roadmap for artificial intelligence. *VINE Journal of Information and Knowledge Management Systems*, *51*(3), 353–368. https://doi.org/10.1108/vjikms-07-2019-0107.

Oates, B. J. (2005). *Researching information systems and computing*. Sage.

Pessoa, A. S. G., Harper, E., Santos, I. S., & Gracino, M. C. D. S. (2019). Using reflexive interviewing to foster deep understanding of research participants' perspectives. *International Journal of Qualitative Methods*, *18*, 1609406918825026.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, *8*(3), 238–264. https://doi.org/10.1108/11766091111162070.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, *46*(1), 192–210.

Ramesh, B., Cao, L., Kim, J., Mohan, K., & James, T. L. (2017). Conflicts and complements between eastern cultures and agile methods: An empirical investigation. *European Journal of Information Systems*, *26*(2), 206–235.

Saunders, C., Benlian, A., Henfridsson, O., & Wiener, M. (2020). *IS Control & Governance*. M. Quarterly.

Sheikh, A., Anderson, M., Albala, S., Casadei, B., Franklin, B. D., Richards, M., Taylor, D., Tibble, H., & Mossialos, E. (2021). Health information technology and digital innovation for national learning health and care systems. *The Lancet Digital Health, 3*(6), 383–396. https://doi.org/10.1016/s2589-7500(21)00005-4.

Sikolia, D., Biros, D., Mason, M., & Weiser, M. (2013). *Trustworthiness of grounded theory methodology research in information systems* Proceedings of the Eighth Midwest Association for Information Systems Conference, Illinois.

Smuha, N. A. (2019). The eu approach to ethics guidelines for trustworthy artificial intelligence. *Computer Law Review International, 20*(4), 97–106.

Stewart, J. (2012). Multiple-case study methods in governance-related research. *Public Management Review, 14*(1), 67–82.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013a). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems*, *30*(3), 141–178.

Tallon, P. P., Short, J. E., & Harkins, M. W. (2013b). The Evolution of Information Governance at Intel. *MIS Quarterly Executive*, *12*(4), 189–198.

Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness in automation require a deeper understanding of human-ml augmentation. *MIS Quarterly*, *45*(3), 1483–1500. https://doi.org/10.25300/misq/2021/16535.

Tiwana, A., Konsynski, B., & Venkatraman, N. (2013). Information technology and organizational governance: The IT governance cube. *Journal of Management Information Systems, 30*(3), 7–12.

Trocin, C., Hovland, I. V., Mikalef, P., & Dremel, C. (2021a). How Artificial Intelligence affords digital innovation: A cross-case analysis of Scandinavian companies. *Technological Forecasting and Social Change, 173*, 121081.

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021b). Responsible AI for digital health: a synthesis and a research agenda. *Information Systems Frontiers*, 1–19.

Van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: an ethnography of developing AI for hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/misq/2021/16559.

Wamba-Taguimdje, S.-L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal, 26*(7), 1893–1924. https://doi.org/10.1108/bpmj-10-2019-0411.

Wilkin, C. L., & Chenhall, R. H. (2010). A review of IT governance: A taxonomy to inform accounting information systems. *Journal of Information Systems, 24*(2), 107–146.

Wynn, D., Jr., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly*, *36*(3), 787–810. https://doi.org/10.2307/41703481.

Zhang, D., Pee, L., & Cui, L. (2021). Artificial intelligence in E-commerce fulfillment: A case study of resource orchestration at Alibaba's Smart Warehouse. *International Journal of Information Management*, *57*, 102304.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Emmanouil (Manos) Papagiannidis is a doctoral student in the Department of Computer Science. He previously earned a B.Sc. in Applied Informatics from the University of Macedonia (UOM), an M.Sc. in Computer Science from the University of Newcastle, and an M.Sc. in Data Science from the University of Newcastle. Manos has also worked as a software developer in Greece (Athens) and the United Kingdom (Newcastle), where he built websites, managed systems, and improved or fixed existing code by adding new features and introducing new technologies.

Ida Merete Enholm works as a full-stack developer in Oslo (Norway) building digital solutions for a variety of customers. She has received her M.Sc. in Computer Science from the Norwegian University of Science and Technology (NTNU). Her Master's Thesis was focused on responsible governance of AI.

Christian Dremel is a "Alain Bensoussan" Fellow at the Norwegian University of Science and Technology, lecturer at the University of Bamberg, and senior research fellow at the University of St.Gallen. In practice and academia, he addresses sociotechnical aspects of digital transformation, digital innovation, and the strategic use of information systems. His research has been published in journals such as the Information & Management, MIS Quarterly Executives (MISQE), Electronic Markets and Technology Forecasting and Social Change, and presented at conferences such as the International Conference on Information Systems (ICIS) and the European Conference on Information Systems (ECIS).

Patrick Mikalef is an Associate Professor in Data Science and Information Systems at the Department of Computer Science. In the past, he has been a Marie Skłodowska-Curie post-doctoral research fellow working on the research project "Competitive Advantage for the Data-driven Enterprise" (CADENT). He received his B.Sc. in Informatics from the Ionian University, his M.Sc. in Business Informatics for Utrecht University, and his Ph.D. in IT Strategy from the Ionian University. His research interests focus the on strategic use of information systems and IT-business value in turbulent environments. He has published work in international conferences and peer-reviewed journals including the Journal of Business Research, British Journal of Management, Information and Management, Industrial Management & Data Systems, and Information Systems and e-Business Management.

John Krogstie (1967) holds a PhD (1995) and a MSc (1991) in information systems from the Norwegian University of Science and Technology (NTNU), where he is currently a full professor in information systems at the computer science department (IDI). At IDI he was Department Head 2017-2021. His research interests are information systems modelling, information systems engineering, quality of models and modelling languages, neuro-conceptualization, eGovernment, sustainable smart cities and sustainable digitalization in general. He has published more than 350 refereed papers in journals, books and archival proceedings since 1991.

PAPER 4

Uncovering the dark side of AI-based decision-making: A case study in a B2B context

Emmanouil Papagiannidis, Patrick Mikalef, Kieran Conboy, Rogier Van de Wetering

Industrial Marketing Management)

Industrial Marketing Management 115 (2023) 253–265

Contents lists available at ScienceDirect

Industrial Marketing Management

journal homepage: www.elsevier.com/locate/indmarman

Uncovering the dark side of AI-based decision-making: A case study in a B2B context

Emmanouil Papagiannidis [a,*], Patrick Mikalef [a,d], Kieran Conboy [b], Rogier Van de Wetering [c]

[a] Norwegian University of Science and Technology, Trondheim, Norway
[b] National University of Ireland, Galway, Ireland
[c] Open University of the Netherlands, Heerlen, Netherlands
[d] SINTEF Digital, Trondheim, Norway

ARTICLE INFO

Keywords:
Artificial Intelligence
AI dark side
AI trading
B2B
AI challenges
AI decision making
Case study

ABSTRACT

Over the last decade, many organizations worldwide have been assimilating Artificial Intelligence (AI) technologies to increase their productivity and attain a competitive advantage. As with any technology, intelligence systems come with potential downsides. Despite the efforts to mitigate any negative consequences of AI, businesses and employees continue to confront the dilemmas of adopting AI, so it is essential to explore in detail the rising concerns around such technologies. In this paper, we used a single case study to investigate the dark aspects of AI in a Norwegian energy trading firm. We gathered data through semi-structured interviews and secondary data. Specifically, we interviewed AI managers, traders and developers who have worked on deploying and using AI tools over the last three years. Our aim is to identify the dark side of AI use in trading, how AI trading bots affect the relationship between traders and AI developers and how the firm adjusts to this new reality. The findings indicate that negative or unintended consequences of AI can be grouped into three clusters related to (1) the nature of the work; (2) conflicts and effects; and (3) responsibility. The paper concludes with future research and practical implications that can help organiziations mitigate the negative aspects of AI use.

1. Introduction

Artificial Intelligence products have the potential to vastly improve our professional and personal lives. Firms seek to make the most of AI opportunities, but with new opportunities come new challenges and risks, constituting a dark side of AI. Challenges include privacy concerns, data security, and ethical dilemmas such as staff replacement and AI fairness (Wirtz, Weyerer, & Sturm, 2020). There is little understanding of the AI challenges confronting the public or private sector, and there is no consensus on how to address them in the future (Sun, Li, & Yu, 2022). Even though AI research has started focusing on these concerns, few studies have dealt with them in depth. Understanding these challenges is important because, while the development of AI may be one of the most outstanding achievements in human history, it can be a double-edged sword. Prominent opinion leaders have voiced their concerns about this matter. According to Elon Musk and Stephen Hawking, AI may be beneficial in the future, but it may pose a threat to humanity; for instance, AI might lead to massive job losses or be vastly deployed in warfare (CBNC, 2017, 2021). It may drive people to the sidelines or even create unanticipated damage; hence, we have to consider what AI means for our society and how to mitigate its risks in advance. For example, AI might be exploited by unauthorized users, or the AI itself might cause substantial financial losses in an instant; thus, the dark side of AI must be investigated further.

In previous studies (Akter, Wamba, Mariani, & Hani, 2021; Li, Liu, Mao, Qu, & Chen, 2023), researchers began exploring the potential negative consequences of AI and how they can be mitigated by incorporating AI into business analytics (BA) capabilities, specifically focusing on data, governance, and training resources. However, there is a limited amount of research available on AI in the context of B2B interactions and its practicality in such scenarios. Rana, Chatterjee, Dwivedi, and Akter (2021) have noted the scarcity of studies examining the impact of data, system quality, and end-user training on competitiveness, both directly and indirectly, when utilizing AI-BA capabilities and

* Corresponding author.
E-mail addresses: emmanouil.papagiannidis@ntnu.no (E. Papagiannidis), patrick.mikalef@ntnu.no (P. Mikalef), kieran.conboy@nuigalway.ie (K. Conboy), rogier.vandewetering@ou.nl (R. Van de Wetering).

https://doi.org/10.1016/j.indmarman.2023.10.003
Received 21 September 2022; Received in revised form 21 August 2023; Accepted 6 October 2023
Available online 12 October 2023

considering the potential negative effects associated with AI. Another study conducted by Castillo, Canhoto, and Said (2021) demonstrates the growing trend of AI technology providing support or even substitution for frontline employees (FLEs), resulting in substantial investment in AI for automated customer service agents, which has led to one of the highest shares of AI investment (USD 4.5 billion worldwide in 2019 (Castillo et al., 2021)) aiming at automated customer service agents. Using conversational agents or chatbots, as well as voice-controlled digital assistants (e.g., Alexa), will fundamentally change the nature of service interfaces from being predominantly user-driven to being more autonomous and technologically driven, even in B2B. However, the authors pointed out that value can also be collaboratively co-destroyed during the interaction process (Castillo et al., 2021). It is possible that the autonomy of AI could result in suboptimal outcomes if the technology is adopted in unintended ways or if FLEs act upon biased data.

The negative effects do not stop there. They might affect innovation or the potentional to create superior value as mentioned by Gligor, Pillai, and Golgeci (2021) who argue that when relationship bonds are reduced, they can lead to adverse B2B outcomes; thus, it is essential to examine the negative aspects (dark side) of such relationships. Rai (2020) expresses his concern about the power asymmetries in a B2B context and he proposes research on explainable AI, while Behera and Bala (2023) suggest future research on organizations' ethical decision-making because when decision-making is destructive, it can lead to dysfunctional and undesirable behaviours. Gligor and Esmark (2015) also conclude that future research is required on how managers should establish policies, avoid negative effects and provide guidance to employees to help them create and maintain positive relationships. Based on the above, we argue that a gap exists on the dark side of AI in a B2B context. We build on an in-depth case study of a company that operates in the energy sector and has been utilizing AI solutions to improve energy trading. This company engages in B2B operations by acting as a liaison between energy producers and industrial customers.

Based on the above discussion, the study's research objectives are as follows. First, we aim to identify the negative aspects of AI usage in a B2B context. As a large proportion of organizations that are deploying AI within a B2B context, it is important to understand the potential negative or unintenteded consequences that may emerge. Second, adopting an inward view, we aim to understand how the deployment of AI influences the relationships between different key stakeholders within the organizational boundaries. As AI has introduced significant changes in the structuring and organizing of the company, it is important to understand how the relationships between groups of employees have shifted. Lastly, we examine the adaptation of specific employee categories, like traders, to the emerging landscape and its impact on their job perceptions and objectives, aiming to understand the dynamics between traders and AI developers influenced by AI and to assess the adjustment of AI traders in this novel context and we give future research directions based on the themes we came up with.

Based on our findings, we are able to shed light on managers' actions, traders' reactions, and the relationships between AI developers and traders. Our findings indicate that from an organizational perspective, a cultural shift towards AI use and AI training were among the new procedures and policies managers implemented to minimize complaints and fears about AI replacement and potential errors while managers tried to maximize profits. We also found that traders adapt to their new roles and responsibilities by becoming hybrids between traders and AI supervisors. Traders had mixed feelings about using an AI system that replaced a great deal of their work, while AI developers felt pressure to show real results and assist traders in their new role. Finally, AI developers worked hard to maintain good relationships throughout departments since they did not want their colleagues to blame them for losing part of their work.

The paper is structured as follows: Section 2 presents the theoretical background and discusses the application of AI in B2B and its potential negative aspects. Section 3 outlines the data collection and analysis methodologies employed. In Section 4, the findings from the data analysis are presented. Finally, Section 5 concludes the study with a discussion of the results, their interpretation, evaluation, and the limitations of the research.

2. Theoretical background

Our inquiry into AI begins with a definition of intelligence in the human context, which is defined as a person's ability to learn, cope with new situations, grasp and handle complicated concepts, and impact one's surroundings through knowledge (Demlehner & Laumer, 2020) or as the ability to perceive and interpret information, transform that information into knowledge, and then apply that knowledge to goal-directed activities (Paschen, Wilson, & Ferreira, 2020). Consequently, perceiving one's environment, problem solving, reasoning, learning, memory, and acting to achieve goals are only a few of the tasks that contribute to good intelligence adaptation. AI can be defined as a system capable of interpreting external data, learning from such data, and using it to achieve specific goals and tasks through flexible adaption (Enholm, Papagiannidis, Mikalef, & Krogstie, 2021). Real-life instances include the implementation of automation bots to alleviate workloads, the utilization of AI for content creation in marketing, and the application of AI to personalize experiences and satisfy individual customer needs. The B2B marketing ecosystems have witnessed the impact of AI automation and AI products, which offer diverse applications in industry (Stone et al., 2020). For example, B2B marketing companies consider AI predictions of customer purchase behaviour to be one of the most significant parts of increasing revenues (Moradi & Dass, 2022; Paschen et al., 2020).

With AI omnipresent in everything from smartphones to household finances to law and justice systems (ethical AI robots), both the public and businesses need to understand that although AI technologies indisputably offer a lot of advantages, disadvantages exist too (Vanderelst & Winfield, 2018). One disadvantage of AI-based technologies in a business environment is the requirement to involve clients in the customer service process, which increases complexity and, eventually, increases the chance of failure. Customers may become irritated and frustrated after spending a lot of time and effort during an interaction when co-created services fail to meet customer expectations, since AI chatbots lack the human intervention that plays a significant role in configuring costumers' needs and maximizing their satisfaction (Grönroos & Voima, 2013). To extend this, based on the work of Plé and Cáceres (2010), value decreases when differences exist in resources, practices, or AI agents' activities and services. Camilleri and Neuhofer (2017) discovered numerous value formations in their work about the tourism industry. To be more precise, they discovered that value could be co-created and co-recovered, and co-reduced and co-destroyed in a sharing economy context. In addition, recent research by Grundner and Neuhofer (2021) has highlighted the potential drawbacks of AI, encompassing areas such as job displacement, privacy risks, machine ethics, security concerns, and negative developments in superintelligence systems.

Other concerns or drawbacks are customer data privacy and security, a possible decline in social interactions among end users during their experiences, and technological limitations, which may dissatisfy both customers and employees. There may even be implications associated with different business settings and their capacity to establish trust and guarantee that each transaction is tamper-proof (Risius & Spohrer, 2017) as research suggests that trust builds close relationships among B2B participants that are mutually beneficial (Gligor & Holcomb, 2013). B2B relationships that suffer from diminished relationship bonds lead to a lack of innovation or inferior value creation (Gligor & Esmark, 2015). By analyzing big data, for example, firms can better understand their B2B partners and anticipate their needs and behaviours (Hallikainen, Savimäki, & Laukkanen, 2020) but it lets firms rely less on information obtained from their partners and eventually they might neglect them. That implies that if a firm loses sight of the importance of managing B2B relationships, it may experience erosion in the quality of its relationships, which can eventually lead to adverse outcomes (Gligor et al., 2021).

2.1. AI use in B2B marketing

AI use and adoption in B2B marketing is driven by two primary motivators (Keegan, Canhoto, & Yen, 2022). The first motivator is the fact that AI is capable of handling large datasets and identifying new patterns in data, which can be used to generate new insights (Cortez & Johnston, 2017), increase efficiency (Bag, Gupta, Kumar, & Sivarajah, 2021) and enhance decision making (Borges, Laurindo, Spínola, Gonçalves, & Mattos, 2021). This illustrates how AI technology improves marketing campaigns' effectiveness, improving the firm's performance (Liu, 2020). For instance, AI is transforming B2B marketing by solving issues in the labour intensity of data collection, management and analysis by creating better customer insights and enhancing and personalizing customer experience (Kim, Kim, Kwak, & Lee, 2022). The second motivator is the alleged cost reductions that AI technologies may bring (Keegan, Canhoto, & Yen, 2022) and the fact that AI is meant to be quicker and less prone to errors than humans (Davenport, Guha, Grewal, & Bressgott, 2020). Even the most basic AI system, however, needs a substantial upfront investment, a lot of computing power, access to several datasets, and regular upgrades. It is vital to emphasize that industrial AI is currently being developed in a conventional B2B setting and it is a transitory endeavour. As part of the design, development, and implementation of such technological solutions, businesses need to interact with customer companies and hire, temporarily at least, consulting firms and IT suppliers (Li, Peng, Xing, Zhang, & Zhang, 2021). As a consequence, partners and stakeholders, from many parties with varying interests and requirements, collaborate to produce value.

In order to produce value through collaboration, B2B marketing operations often involve agencies, which are specialized third-party marketing firms responsible for adapting a company's marketing message to target other companies (Huang & Rust, 2018). These agencies possess extensive knowledge about effectively reaching decision-makers for high-priced products and services, while B2B marketing ecosystems have seen major changes in the last decade due to the introduction of new technologies and process automation (Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021). Among these changes, the utilization of AI techniques and software has emerged as one of the most notable advancements, aiming to increase efficiency and optimization and streamline processes through the implementation of intelligent agents or systems (Saura et al., 2021). What is more, B2B marketing teams recognize the importance of meeting their customers' evolving needs and integrating customization into the sales process to add value. For that reason, many businesses rely on mass customization for marketing purposes to increase product variety, and through that customer satisfaction, without increasing costs (Kamis, Koufaris, & Stern, 2008). To accomplish this, they leverage AI systems in advanced manufacturing with the primary goal of economic growth, leading to technological advancements and greater competitiveness in both local and international markets (Chen, 2017). To take advantage of mass customization though, manufacturing processes should embrace digital production and adopt AI systems (Moradi & Dass, 2022).

2.2. Industrial AI in B2B

Past research (Ives, Palese, & Rodriguez, 2016) indicates that AI has led to more profound adjustments and transformations to industrial organizations than earlier digital and technological innovations. Industrial AI is referred to as a broad spectrum of enterprise activities utilizing machine and deep learning (Li et al., 2021). In general, it refers to AI technology utilized to solve issues related to complex industrial operations, collaborations, and marketing activities. For example, networks have to swiftly respond to abnormalities and adjust to shifting traffic in order to sustain excellent operations (Kromkowski et al., 2019). AI is being used by telecommunications companies to monitor and enhance their networks and give the best performance to their consumers (Liang, Li, Long, Kui, & Zomaya, 2019). Another example would be in the sector of oil, gas, and energy due to safety and environmental concerns. Energy firms are now able to boost their efficiency without raising prices thanks to AI advancements (Ahl et al., 2020). This includes applications such as image processing for identifying maintenance requirements and predictive models for energy demands (Pradhan, Ghose, & Shabbiruddin, 2020).

Furthermore, a recent analysis by MIT (2018) highlights the crucial role of AI technology in enhancing business and professional service outcomes. The combination of AI and big data may assist B2B organizations in discovering and using vital information and expertise leading to a competitive advantage (Li et al., 2021; Paschen, Paschen, Pala, & Kietzmann, 2021). Furthermore, B2B marketers should focus on both consumers and people who make purchasing choices because both customers and those who make the purchases play a significant role in B2B marketing. It is envisaged that adding AI processes would result in even greater marketing efficiency through personalization (Abrell, Pihlajamaa, Kanto, Vom Brocke, & Uebernickel, 2016; Li et al., 2021) because it is critical to understand supplier and customer firms' requirements to properly create and apply industrial AI in B2B marketing (Wang, Ma, Zhang, Gao, & Wu, 2018). AI can aid marketers in gathering important information from consumers, retaining existing customers, and increasing customer satisfaction (Meire, Ballings, & Van den Poel, 2017), while assisting B2B managers to acquire more accurate customer-related innovation for retaining existing clients and exploring new opportunities to attract new customers (Han et al., 2021).

What is more, AI is rapidly being used to improve B2B market performance by speeding up decision-making processes. While this phenomenon has been widely embraced in the B2B sector, little academic research has been conducted on it in the context of industrial marketplaces (Dwivedi & Wang, 2022). That means there is a research gap that could be filled and offer illumination about customers' or competitors' habits and decisions, helping firms to improve their products and services. The majority of AI research focuses on consumer marketing at the moment, but industrial data is rarely evaluated to solve challenges such as organizational behaviour, product innovation, supply chain management, and B2B customer relationship management (Davenport et al., 2020). Besides the B2B marketing ecosystem changes, estimating the net client lifetime value (Chan & Ip, 2011) is considered critical, especially for digital marketing B2B companies. One of the most tried and tested B2B marketing tactics is personal selling, because it focuses on face-to-face networking and contacts to close purchases. This is the least scalable approach to promoting your firm to other businesses, but it has the highest conversion rate. AI, in this matter, helps firms make better decisions and create more effective content since firms may conduct targeted marketing operations, resulting in increased ROI, due to the benefits of understanding the audience better.

Simultaneously, firms attempt to maximize profit and other quantitative indicators like inventory investment and storage capacity (Chen & Chen, 2008). Data mining techniques are known to have influenced the development of intelligent systems that help the marketing strategy process (Martínez-López & Casillas, 2013). The objective is to aid managers in dealing with ambiguity and uncertainty while providing sensible marketing strategy guidance (Li, 2000). As a result of these AI encounters, B2B marketing businesses may learn from them and adjust their views and strategy accordingly (Cruz, 2009). Consequently, in the last decade, AI applications have been introduced for trading purposes in B2B marketing businesses. Businesses that leverage data to make better, faster, and more accurate trade decisions have a competitive advantage. Organizational digital networks generate a significant volume of data and have immense value for enterprises. For instance, important commercial events, natural disasters, athletic events, political crises, or just popular topics tend to increase messaging activity in networks by creating large amounts of social media data. These types of network activities are crucial for trading in B2B marketing environments since the increase in communication might be a sign of a broader problem, and it can be documented and researched to identify meaningful trends and patterns (Candi, Roberts, Marion, & Barczak, 2018). To be more specific, data stored in structured and unstructured databases, when combined with other real-time data streams (feeds from social media or sensors), may assist management and stakeholders in understanding the severity and intensity of an unfolding crisis.

Such data is usually stored or produced from outside sources. External data may be classified into two types. The first category consists of data that is directly linked to organizations, such as online social media content and mobile devices. The volume of data collected from professional and personal contacts (such as social media and cellphone data) is concerning since AI can discover personalized patterns, which threatens individual privacy. The second type of data is not directly related to an organization but can impact its performance. Collecting socio-cultural data, for example, can assist in enhancing corporate operations by making judgments that are more in line with contemporary cultural developments (Farrokhi, Shirazi, Hajli, & Tajvidi, 2020); thus, customer knowledge, user knowledge, and external market information are all key components of the knowledge management process and are accountable for the development of B2B marketing knowledge (Abubakar, Elrehail, Alatailat, & Elçi, 2019) contributing to the accumulation of big data.

Big data proves valuable in extracting critical information from structured and unstructured data inputs, such as web browsing behaviour, demographic characteristics, and purchase trends, and delivering relevant consumer knowledge for rational decision-making. With big data, B2B marketing firms could build models that generate new content and target the right audience with the right content. Customers will be happier since they will find what they need according to their needs. That means user knowledge is required for the creation of new products and method innovation and improvement (Bag et al., 2021). B2B marketing firms need to have a thorough awareness of the external market to stay ahead of the competition. By analyzing unstructured data such as news, social media content, and specialized external sources, AI may help B2B marketers improve analytic and decision-making abilities, and promote creativity (Paschen, Kietzmann, & Kietzmann, 2019). Hence, knowledge management and decision-making styles are critical for corporate success in the digital age (Bag et al., 2021). As a result, B2B organizations can utilize AI to turn enormous amounts of data into knowledge and, ultimately, expertise to develop efficient sales plans and tactics.

2.3. AI use in B2B marketing and its dark side

Limited research has been conducted to better understand the processes behind the dark side of B2B relationships and find solutions to minimize their negative effects (Sharma, Kingshott, Leung, & Malik, 2022). From an organizational standpoint, a company's reputation and overall profit are likely to be impacted by the launch of AI-enabled goods. The effect of the impact, positive or negative, is determined based on the success of the system. For instance, the effectiveness of AI-enabled chatbots affects how satisfied customers are (Ashfaq, Yun, Yu, & Loureiro, 2020). Chatbots that are unable to offer the requested information would have a negative impact on customer satisfaction, meaning that customers will distrust chatbots if they do not perform as well as they are expected to; thus, customers will not trust the sellers or the businesses who use AI as a result (Yen & Chiang, 2021). Furthermore, because AI is still a relatively new technology, firms frequently miss the opportunity to address the question of how AI plans will affect the human workforce. The most common fear of AI is the fact that people might lose their current positions and be unemployed due to the technological advances in AI (Li & Huang, 2020). The argument has merit since AI can do massive tasks in parallel to reduce costs and improve performance, especially for B2B that operate a type of AI-CRM (Chatterjee, Rana, Tamilmani, & Sharma, 2021). Because of that, employees fear being replaced and cannot realize how to coexist with AI, leading them to reject the idea of embracing AI in their everyday work.

From a social standpoint, AI has negative repercussions, ranging from data security concerns to ethical dilemmas (Boyd & Wilson, 2017). The issues involve AI legislation, regulations, bias, and fairness. Mikalef, Conboy, Lundström, and Popovič (2022) connect the responsible AI's aspects with the dark side of AI. One of the most common aspects is fairness and bias, which could have a huge social impact; therefore, these potential issues provide difficulties for AI governance at the social level. Consequently, many studies on fairness and bias concentrate on various choices, such as HR recruiting, budget distribution, or a set of medical testing (Mikalef et al., 2022). This focus could be justified due to the fact that a business cannot afford to ruin its position in the market since the risk of using AI could be high in some cases. AI risk assessment needs to be part of a larger business risk management framework that incorporates regulatory, financial, credit, and information technology challenges, because AI risk assessment procedures will not be ad hoc if an organizational risk management framework is implemented throughout the whole firm, i.e., there will be a continuous plan enforced with management-approved regulations (Barta & Görcsi, 2021). This signifies that data governance and safety measurements could not be ignored since these can prevent the negative consequences of the dark side of AI (Cheng, Su, Luo, Benitez, & Cai, 2021; Mikalef et al., 2022).

Another aspect of the dark side relates specifically to AI trading. Professional investors can trade financial assets in secret and anonymously by utilizing dark liquidity pools (or dark pools). Dark pools enable investors to conceal their market moves from rival traders by not releasing pre-trade information such as pricing, volumes, and the number of open orders. However, dark pools may potentially jeopardize financial markets' informational efficiency and the fair pricing of securities by preventing pre-trade information from being available to all market participants (Lagna & Lenglet, 2020). This dichotomy, which characterizes dark pools, raises an important question about how dark liquidity providers can persuade investors that trading in the dark is secure. In addition, there are legal concerns when dealing with AI. These concerns intersect with ethical considerations as there are many who expressed their views on the ethical aspect of AI. Using the proprietary trading sector as an example, emerging threats to the safe use of existing legal concepts of market abuse in dealing with misconduct by more autonomous AI trading bots should be examined. Autonomous AI trading has the potential to exhibit unparalleled flexibility and develop capabilities that human specialists can only aspire to. Because of self-learning, AI traders may operate in unexpected ways though, for both good and evil. Various ethical and legal issues emerge when dealing with accountability issues for algorithmic misbehaviour. AI's misbehaviour, for example, might someday undercut current market abuse restrictions (Azzutti, Ringe, & Stiehl, 2022). As a result, it is difficult to determine who is truly responsible and allocate responsibility for decisions that could be taken automatically by the AI or in combination with a human agent.

3. Methodology

Through case studies, complex issues in their real-life contexts can be explored thoroughly and in a multifaceted way (Rashid, Rashid, Warraich, Sabir, & Waseem, 2019). Many fields, such as business, law, and social sciences, recognize the value of the case study approach. Moreover, case analysis can be very helpful for gaining a more in-depth understanding of an issue, event or phenomenon in its natural setting. As a research strategy, a case study has traditionally been viewed as lacking rigour and objectivity compared to other kinds of social research (Gibbert, Ruigrok, & Wicki, 2008). This is a major reason for carefully justifying the process of designing and implementing a research study. Despite this skepticism about case studies, they are widely used because they can offer insights that might not otherwise be available (Yin, 1981). Moreover, case studies are frequently used to develop more structured instruments that make surveys and experiments possible in the preliminary, exploratory stages of a research project. Case studies provide useful information for contemporary events when the relevant behaviour cannot be manipulated. As a rule, case study research draws its evidence from diverse sources, including documents, artefacts, interviews, and observations. Rashid et al. (2019) suggest interviews for refining theories or understanding phenomena. By analyzing these data, users can gain new insights that are useful for explaining similar situations (Oates, 2005).

A qualitative methodology is chosen for our case study because it allows for flexibility in our case study and encourages discussion, which can be used to comprehend and explain the research goal (Michael, 1997), as it includes key respondents' experiences, beliefs, and attitudes through semi-structured interviews (Wynn Jr & Williams, 2012). Consequently, for the purposes of this work we used interviews combined with secondary data, including reports and internal documents, in order to supplement our data and validate our findings. We analyzed the comments and observations from different transcripts to discover common themes and patterns representing the dark side of AI trading. Another reason for using axial coding (Charmaz, 2014) is to group the comments and observations, which allowed for better interpretations due to the employees' ability to refer to the same concept using similar terminology based on their technical knowledge, experience, and position in the company. To ensure high confidence levels, the researchers examined reports, public information, and presentations related to this research that focused on the dark side of AI trading.

3.1. Case context and data collection

For this particular study, a Norwegian company working in the power industry has been chosen. The company is a midsize enterprise (around 500 employees) and has been in operation for 65 years. There are three main criteria that make this company suitable for this case study. Firstly, according to the Global Economic Forum's 2019 Global Competitiveness Report organizations, Scandinavian countries have a high level of ICT adoption. Secondly, the majority of individuals have strong digital skills, making them well-equipped for digital transformation. Thirdly, the company's vision and plan for developing, expanding and using AI into its B2B marketing practices in order to maintain a competitive advantage over its competitors. To be more specific, AI is used for buying and selling energy, and the company advertises these AI driven activities. Furthermore, the company uses AI to place itself in the market as a strategic partner, while at the same time the company uses AI for ensuring efficiency and low energy prices.

As part of an interview design, a total of fourteen interviews were conducted. Each participant was questioned for an average of 45 min, allowing them to convey their understanding in their own words and based on their own ideas on specific subjects. To comprehensively explore the dark side of AI trading, input was sought from both the trading and AI departments. The involvement of both departments was crucial to gain a comprehensive understanding and obtain insights into their perspectives regarding the dark side of AI trading. For this reason, the guideline questions were divided into three sections. The first part dealt with how they used to do their trading activities without AI, including the aspects of time management and the effort they had to put in to accomplish the initial business goals. The second part focused on the use of AI in trading and how it affected them personally and as a department, while we were trying to identify the various aspects of the interaction between human and machine. The last part targeted the negative aspects that might occur because of the use of AI in trading. Mostly, we investigate the fears and concerns of the employees and the expectations they had from the use of AI in their everyday work. What is more, during the interviews, we try to find any tensions between departments because of the development of AI, to be more specific, between AI developers and AI traders. Table 1 presents an overview of the participants and their respective roles within the organization.

In our case, traders do not necessarily adhere to the strict definition of a trader, thus it is important to note their activities and contribution to the company. To be more precise, while some AI traders have a background in computer science and finances, others do not. The company's strategy was to follow a quantitative approach from the beginning and this was reflected in the hiring process, where they prefer candidates with knowledge in computer science and finances. Traders' everyday activities primarily involve selling power, forecasting energy consumption, and keeping an eye on AI decisions. They also assist, or at least have assisted in the past, with the development of the AI trading system as a side task because their skills and knowledge were needed to evaluate AI and calibrate the data utilized for developing AI. Additionally, traders offered the AI team insightful comments on the outcomes of the AI while suggesting features and methods for testing the capabilities of the AI system.

3.2. Data analysis

We analyze content from the interviews employing a narrative analysis process because the experiences and stories shared by employees are used to answer the research questions. Our analysis is inductive in nature, as we gather our data (interviews, reports, etc.) and we came up with the general conclusions based on these data. The transcripts that were generated were imported into the software NVivo. During the analysis, we utilize an axial coding process as it involves relating data together to uncover codes, categories, and subcategories encapsulated in the voices of participants (Michael, 1997). Essentially, axial coding is a method of constructing links between data and it is used to enhance the depth and structure of existing categories. According to Charmaz (2006), axial coding aims to reassemble data and is a step that follows open coding. In our case, there are four groups of nodes corresponding to societal, organizational, interpersonal and individual entities (Ragins & Sundstrom, 1989). We also draw upon the work of Linstead, Maréchal, and Griffin (2014), who describe a longitudinal resource development model of power in organizations and the dark side of organizational behaviour, including the concept of organizational misbehaviour. To elaborate further, the leadership and expertise that individuals bring to a position within an organization are the emphasis on the individual level. The interpersonal analysis focuses on the connections between people in light of their roles within the organization, while the analysis of organizational nodes focuses on selection and promotion methods. The social level focuses on the development of roles and expectations throughout society as a whole. The analysis (Table 2) can be perceived as a system with four nodes in which each node interacts with the others, creating interconnectedness. Actions taken at any node have the potential to impact and be influenced by events at other nodes. Table 2 presents the observations we generated, the themes

Table 1
Respondents' stats.

Respondent ID	Role	Years in company
R1	Quality control AI manager	5
R2	ML Engineer	4
R3	Trade AI Manager	5
R4	Chief AI Officer	5
R5	ML Engineer	3
R6	ML Engineer	4
R7	ML Engineer	4
R8	AI Trader	3
R9	Data Scientist	4
R10	Trade AI Manager	5
R11	Data Scientist	4
R12	AI Trader	3
R13	AI Trader	3
R14	Data Scientist	16

Table 2
Themes, observations, and nodes for the dark side of AI trading.

Themes	Observations	Nodes
Nature of work	Deskilling	Individual
	AI false expectations	Organizational
	Unemployment	Social & environmental
	Mobilize human capital	Organizational
	Losing interest in work	Individual
	Hacking attempts	Organizational / Social & environmental
Responsibility	Lack of AI decision explainability	Organizational
	Absence of AI accountability	Organizational / Interpersonal
	Manipulating the market	Social & environmental
Conflicts and effects	Portfolio risks	Individual
	Enforce patterns / overconsumption	Social & environmental
	Conflict of interest between managers and traders	Interpersonal
	Sell overseas / lack of energy	Social & environmental
	Conflicts among AI developers and traders	Interpersonal
	Conflicts among AI traders and non AI traders	Interpersonal

formed by grouping these observations, along with the corresponding nodes.

4. Findings

The interviewees discussed the negative aspects of AI with respect to trading. In particular, they provided their views about the use of AI in their daily work and what the future holds for them as a result of the continued evolution of AI capabilities. The findings provided a range of contrasting views in terms of the existing and potential benefits but are wary about the potential consequences for their careers if AI becomes mainstream.

4.1. Nature of work

To begin with, employees had high expectations of AI technologies but like any technology or software there were many times when things could have worked better during production time.

> "When it goes to production, things will go wrong, mostly because some data is missing or incorrect… (that is why) we always have mechanisms, like fallbacks and manual mechanisms to correct things…so the system did not produce the value that we expected." (R1)

AI might fail in different ways, such as adding business value or giving unexpected results.

> "Many organizations face the problem of getting things into production…they do some experiments and develop some prototypes but AI is not adopted and used…(personally) I was expecting that AI would be more profitable." (R1)

> "We tried to gather different projects that we thought AI would give some value to us. Then we started with the easier ones and the ones that we believe would give financial gain, but not all of them were successful." (R2)

Similarly respondent 3 stated:

> "We have done some tests on whether we can improve inflow forecasting to hydropower…but the tests that we have done so far gave negative results on the horizons we wanted." (R3)

Since the company decided to develop the software, they had the opportunity to maximize AI outputs as they did not use a generic AI software that provides outputs for unspecific goals. That was a double edged sword as the small AI team could make fatal mistakes because AI models failed to show adaptation when unexpected weather conditions applied.

> "We optimize AI for our company, which I am quite sure is the reason why we were the 'best'. We worked closely with the domain experts, who could help us evaluate our systems in the best possible way. Recently (though) we failed…we forecast for wind parks…but we did not understand how poorly the model worked…So it took time until we realized that something is very wrong here and we have a very high cost of it." (R4)

> "We observe things the system is not aware of. Like the icing conditions on the wind farm and if there is any change in the way the company is operating the windmills. For example, the company have implemented a new software…and the AI models did not know that. Meaning that the models did not learn from the new data." (R14)

Due to intraday trading, i.e. selling and buying on the same trading day, traders do less manual trading and instead tend to rely on high speed AI agents. Consequently, traders lose their competence, and deskilling will be a challenge in the upcoming years since the traders will do less and less trading, transforming the nature of work.

> "You cannot have people doing this manually because the good bits, they just vanish before you see them, maybe before even they get to your screen. So you need algo-traders." (R4)

Respondents R3 and R13 reinforced that statement:

> "We are on the dashboard, we have tried to focus on some key values…So that is sort of daily monitoring." (R3)

> "We have 24/7 operators that are sitting at the production central… (monitoring) processes or predictions." (R13)

As a result, traders do not necessarily become better by learning through AI.

> "I think maybe it is almost the other way around; they see behaviours that maybe they did not expect. They analyse it a little bit, and they use them to improve the algorithm itself, instead of themselves becoming better traders." (R8)

Meanwhile, the company aimed to make the transition from manual to automatic trading smoothly. Traders are still in use as their expertise is required to develop these systems since AI developers are unlikely to be experts in the field. Therefore, the board decided to use AI to do the heavy lifting and manual trading is still on to counter the feeling of being replaced. AI brings this tension to a point where, in the best-case scenario, employees are transferred to another position within the company or, in the worst-case scenario, lose their jobs and have to find new ones in a market being taken over by technology.

> "You could expect because we automate…people will become jobless…We need to automate and then you should automate in the best possible manner. And that is about doing it."

(R4)

However, that will not be the case in the near future. Most trade interactions today occur within a split second, so keeping up with the pace is essential. Currently, transactions occur almost every hour in the energy market, but this will likely change drastically soon.

> "We will switch from our hourly production times to quarterly. So that way, every 15 minutes, there is going to be products that you can trade on…It is not easy to monitor and manage, and do all of these things, because you need people that have high education, and experience as well."
>
> (R8)

Respondent 10 stated:

> "In my mind, we should do it as much automatically as possible. So the operators are not using that much time taking manual decisions, and of course, I would prefer that we use AI for the most part."
>
> (R10)

Respondent 13 expressed his concerns regarding the potential downsizing of positions as a result of AI:

> "I believe some jobs might be changing based on how AI is being used. It might reduce some of the positions due to being more efficient, but at the same time, it might need to be watched a bit. So it will be a change for someone."
>
> (R13)

It is worth noting that a trader who had previously worked in a similar role in another firm, in the same industry, mentioned that AI technology was not well received there, and most traders resisted changing. Their resistance was due to facts or patterns they observed in the market that AI models did not consider highly important; thus, traders were very skeptical, if not against the use of a system that did not incorporate what they believed was important.

> "When we argued with other traders about which trade they should perform, they might say 'Oh, yeah, but even though the expected price is higher than the market price, we should not buy now because the price has gone down the last five days'. So they argue that the trend of the market was very important, whereas our backtests did not support this. So although we were closer to the price due to our distribution model, they had arguments against it based on more short-term things that they observe in the market."
>
> (R12)

Their primary concern was that AI would certainly replace their work, and eventually, they would lose their jobs since AI could do exactly what they do now, if not better. The issue was addressed through workshop sessions. Employees were trained and educated on AI throughout the entire development process, allowing them to understand, at least, how AI works, how traders will use it, and why it was adopted. Last but not least, the AI team built models for other companies. The obvious benefit is revenue, but there might be some problems. For example, if the models fail to predict accurately even for a short period of time, energy prices will rise. The effect may be more severe in northern European countries with more prolonged and colder winters.

> "Short notices on production emergency requirements and incorrect estimations of supply and demand will lead to bad reputation and high prices for consumers."
>
> (R12)

Therefore, companies need to utilize AI to replace manual trading, and hence the role of human traders will be reduced or marginalized. In this case, the company shifted traders to another department and set up a monitoring room where traders monitor patterns and detect any anomalies that may occur. This provided an opportunity for traders to explore the market and discover new patterns because they did not have to conduct as much trading as they had in the past while developing new talents.

> "There will have to be much more back and forth between the trading desk and various developers to ensure that we really have good enough monitoring tools and good enough systems and pipelines that are robust. I think trade will shift more to monitoring tasks, and also coming up with strategies and more control tasks because they would obviously be better at something like that, compared to someone like me."
>
> (R8)

The main benefit of AI is that it reduces overload and makes it possible to scale up in areas where human capabilities cannot grow without the participation of new employees. Because AI's scalability is reliable and supports rapid growth, managers may choose to let go of some employees or not hire new ones in the wake of AI automation.

> "It is much easier to scale up so that you can trade on several portfolios."
>
> (R8)

Hence it is likely that traders will end up doing something that they are not interested in or they have not signed up for in the first place and this might lead them to lose interest in their job. Furthermore, due to the nature of the work cybersecurity issues might arise. Respondent 11 mentioned that he had been aware of similar companies whose AI systems had been hacked, causing financial and reputational loss to the company, its clients and the surrounding social environment.

> "You have to ensure that systems are secured because they have many activators and control over the value chain. Security and AI have to go together. Otherwise, the risk is high. Last year, two companies (in this sector), this is publicly known, were hacked, but I do not know the consequences in monetary value."
>
> (R11)

Nevertheless, traders in the company try to use AI to develop new strategies and tactics that are impossible to think of without using AI. Back-testing is applied, and traders came up with new insights into how they could operate under different scenarios.

4.2. Responsibility

As far as the responsibility part is concerned, many disagree in terms of who is responsible for the AI outcome and use. The concerns arise because it is hard to identify if the AI developers are the ones to blame since they build the software, or the traders who eventually use the AI, although in this case, they are monitoring mostly AI. The lack of explainability tools may contribute to that as it is hard to deal with concerns surrounding transparency and bias.

> "In the energy business, legal complication starts when we are trading with AI. Questions arise, for example, who is responsible or how should we put a new trading algorithm into operation. You have to make sure that you do not place any bids that are bigger than the ones you can take."
>
> (R3)

Respondent 8 added more on this topic:

> "There are going to be humans involved, which means there are going to be a lot of things that are very difficult to understand for a machine. Someone might implement a way to trick the machine in some way or trick the algorithm into gradually lowering or increasing its price to get a good sale or a good transaction out of it. That is something that would be much easier for a trader to detect, because they could see the 'tells'."

(R8)

Companies might be held responsible for price speculations as AI might affect prices and lead to a dramatic increase, meaning that in energy markets speculations are not allowed and there are legal consequences. Furthermore, the company did not have clear distinct roles, which contributed to the question of who was held responsible in case of something going wrong.

> "We have been doing this with no distinct roles. But there are, of course, distinct positions within the company that work on different things. So they will have access to different parts of some database. But it is dependent on the problem that is solved and who deals with it."
>
> (R6)

To mitigate some of the issues, the company decided to promote robustness and reliability through infrastructure and by standardizing processes.

> "In the beginning, there was not an infrastructure which promoted robustness and reliability in the whole process…we have these error detection methods now, which do it automatically."
>
> (R7)

On the same matter, even managers have different views, which highlights the fact that there are blurred boundaries on where responsibility lies. This raises the issue of who is held accountable in cases where there are important deviations from forecasted values.

> "Who is responsible? That is almost a good question when it comes to the AI, but I guess that it should be the head of the Energy Management department who supervises the trading."
>
> (R10)

> "So there is a responsibility when you use AI. You have to have some kind of explainability so that you can actually explain why this trade was done…but then again, who should be held accountable for AI's actions remains a good question."
>
> (R9)

Nevertheless, employees do not find this a major problem since trading in their sector has to follow strict regulations, and there is a great deal of documentation on how AI should behave. All the limits should make the algorithms simple enough, meaning that, ideally, most traders can understand how the processes work and how the decision making is done.

> "You need to define some limits. The algorithm ideally has to be simple enough but not naive, it cannot be too complicated."
>
> (R8)

Obviously, if the algorithms were too naïve, the employees would challenge the outcomes and adoption issues may occur. In case of unexpected events, AI might produce undesirable outcomes, especially in unknown scenarios; thus, the human controller is required to identify and outmanoeuvre the issue. In order to do that the trader should have critical thinking and be able to judge if the AI outcomes are reliable, robust and profitable in the long run.

The responsibility part is a real concern if the legal aspects that the government has enforced are taken into consideration.

> "But I am worried that our algorithm will place orders, for instance, in a way that creates a pattern on the orders. Then you might make a fake impression that there is a lot of buyers in the market, while there is only one algorithm placing all of these orders and that is not legal…someone could implement it in a way to trick the machine or trick the algorithm to gradually lower or increase its price to get a good profit."
>
> (R12)

Hence, the employees raise the question of who is truly responsible for such undesirable outcomes. Violating these rules could mean a lot of legal trouble on the horizon and it is a dark side of AI trading that employees are not willing to face. Considering the upcoming energy crisis, which seems to be a huge problem in the near future, it would be difficult to decide who is accountable, at least for the economic effects on society.

4.3. *Conflicts and effects*

One of the main effects that traders identified in terms of AI outputs was trust due to results in comparison with other traditional methods.

> "We are exploring how we could use AI methods to solve hydropower scheduling problems, especially when it is important to represent uncertainty. But that is on a research level. I am a little bit skeptical, though. I think it will take many decades to use AI in advanced decision making because models cannot compete with the classical optimization tools that we use now."
>
> (R3)

> "We get questions of why and how it comes up with the results, which is maybe one thing that can sometimes be a bit difficult to explain because most of the time, it can make sense, if the data looks a specific way. But sometimes the results do not really make sense."
>
> (R2)

At the same time, employees should be able to provide predictions even if AI data is not updated. That means the quality of outputs might not be as expected. That makes the company decide the development of tools that promote trust in AI.

> "Sometimes you have to do forecasting even if you have missing data, using old data to fill them. It depends on the application how we deal with the missing data. But we need to handle this, we cannot say that we cannot predict. So, maybe the old data are a little bit off, (meaning that) the data we want to use might not be right."
>
> (R5)

> "What is done is that usually a person is in control and he is given options so he can see the choices and he can evaluate (the outcome) and decide if he should use whatever this computer produced for him. This is done as a tool to build trust and confidence in the solution because that is something that is extremely important for us, at least to make sure that the operators feel heard and taken into consideration. Also, it is important to have meetings where we discuss what we are doing and what we should be doing in the future."
>
> (R8)

Another issue that arises is the communication among departments. Traders have to understand new terms and effectively explain what they believe is the problem to AI developers so that they can take all the necessary steps to resolve potential problems. In comparison to other dark sides of AI trading, that minor problem could cause a lot of misunderstanding and confusion and cause a lot of discomfort to traders, who would feel unable to pass on their message effectively.

> "Whenever we talk to experts, we cannot just talk in terms of what a feature does in the model, what are the statistical errors, what models we are using etc. We need to use vocabulary that they understand."
>
> (R7)

"There are very different terms that traders use. They use classical training terms and I was not familiar with those terms. So it took me a while to get used to it…For them it was difficult too, because it was difficult to talk about machine learning terms with them because I could not explain to them properly."

(R9)

Nevertheless, some traders resisted accepting AI outputs. Respondent 12 described as follows other traders' perceptions:

"However, the fundamental models and the way that prices are actually formed might not support all traders' decisions. You have many traders focused on different areas. The importance of these things (i.e. what traders believe is important) might be, for them, very high, but for the actual price formation in the market, it might not be like that at all. So my impression is that they focus on this, ignoring the models, and they just try to justify their view."

(R12)

Another effect is the way employees thought about AI. Especially in the beginning AI was perceived as a magical way of solving all kinds of problems.

"They think AI is magical. They think that if we have some data, and use AI, it will magically solve the problem…Someone comes with an idea, and you have a conversation to find out what we can actually do."

(R3)

"I think people have an overly positive expectation that AI will always come up with profitable technical trading strategies."

(R12)

As noted above, respondent twelve had been part of another company who were making their trade manually, at least in the past. Traders there were quite skeptical and there was a lot of conflict among employees, who were open to the idea of involving new technologies in their work. That was the case when some traders adopted some forecasting models for energy consumption, which triggered a lot of conflict. This behaviour may be a product of how the traders feel about themselves in terms of competence and their understanding of quantitative systems.

"Sometimes when the market was volatile, it went against us. Everyone else was telling us that you should not believe the algorithms now. And they were telling us this all the time…do they feel safe? Well, it depends on their competence, regarding quantitative systems, what underlies these systems etc."

(R12)

In some cases, the traders felt that they had to take extreme actions to either prove their quality as a trader or make a vast amount of money based on their trade decisions. Their personal income is affected by the amount of money they make every year, and a huge bonus may be earned based on that. Hence, it makes sense for them to take huge risks even if that could lead to losing their job, as the money they could earn is more than enough to make the risk worthwhile, if not desirable.

"We would like to take high risk. That means one out of five years, we will lose all our money but that is ok."

(R12)

Conflicts might arise among traders and managers too. Managers need results and a system that is robust and reliable. An automated system can be robust and reliable since it performs with the same quality and speed daily. However, traders do not want to see all parts of their work being automated since they will lose the most important and interesting part of their job, which is trading assets.

"As a manager of course I would prefer that we are using AI for the most part. But of course, since you are trading with physical assets, we need to ensure that AI is making the right trades."

(R3)

What is more, consumption might be affected based on energy prices. For example, companies might choose to sell overseas instead of in the local market because the energy price was higher in these countries. Hence, short-term profitability can shape the nature of consumption, supply and demand, as energy companies can not generate an infinite amount of energy. Some evidence can be found in countries like Germany, where Nordic companies decided to sell their energy to Germany due to the fact that prices were higher there.

"A trader might be very interested in how the German prices are related to the Nordic prices because if the German energy price is very high compared to the Nordics, then it is a sign that the Nordics can export power because energy consumption is always flowing the direction of the price in the power market… and you have all of these market manipulation rules that you have to follow, so if you have an AI placing orders, then you need to be very careful that the way AI is placing orders is not manipulating the prices and hence the consumption."

(R12)

It is worth noting that the firm has moved the decision-making mostly to AI to minimize high risk decisions and in order to be more robust and persistent in decision making. Nevertheless, the managers understood that setting up an AI team and developing AI products is not enough because it is equally challenging to adopt AI in the trade department, without losing the confidence and loyalty of the employees. Another crucial side of AI in trading is the false expectations that the employees develop. It is common to ask for features and capabilities that are not realistic. Traders might ask for a magical solution that will solve huge problems using some data. At the same time, developers should remember that overwhelming traders with information is not ideal, because it could cause discomfort using AI as a trader.

5. Discussion

This study explores the dark side of AI trading and how it affects the employees. Specifically, we gathered data from interviews and organized our observations into three themes: (1) the nature of work, where we investigate how it affects the traders in their work, (2) responsibility, where we investigate who is considered to be responsible for AI decisions and (3) conflicts and effects, where we investigate which conflicts arise between AI developers and traders or the social effects that might occur. The purpose is to underline the dark sides of AI and what the consequences are that underpin them. To be more specific, we found that traders had to go through a process of evolving and adding different values to the business through their expertise. Notably, traders express significant fear of being replaced by AI technology, particularly as it assumes a prominent role in their vital trading functions. Furthermore, responsibility plays a huge role as someone should be accountable for AI decisions and explainability tools should be implemented as they will add a safety net for decision-makers. It is equally important to understand that there might be social effects, such as shaping clients' energy consumption behaviours.

The ability to obtain large amounts of high-quality data and manage that data effectively is crucial for AI. However, it can be challenging to extract worth from B2B data. B2B firms frequently lose out on actionable insights due to the absence of meaningful data and the fact that the data that is acquired is frequently irrelevant and poorly handled, which can result in the development of unsuccessful business strategies (Chatterjee et al., 2021). In order to effectively apply AI in B2B, it is crucial to address the issue of data orchestration, meaning acquiring, cleaning, matching, enriching, and making data accessible across technology systems (Sun, Hall, & Cegielski, 2020). When B2B firms lack the necessary data, they may be able to solve their problem by using insight from AI. Other drawbacks, albeit less pervasive, should be considered. For instance, customers may have increasing demands, such as the desire for customization and competitive pricing, which can extend the purchasing process for businesses. Additionally, a lot of businesses are continuously investing in AI to stay competitive in the market (Chen, Jiang, Jia, & Liu, 2022). Although many firms embrace the AI path, employing AI to drive B2B sales is still in its infancy and has not yet had a big impact, similarly to our case.

A few studies have looked into how AI may affect B2B sales management. AI that predicts future events based on recent data has the potential to automatically add underlying prejudices, which may encourage unfairness. With the demand for enhanced client experiences reaching unprecedented levels, a rise in operational efficiency, and a more intense competitive landscape, it is only logical for B2B vendors to pursue AI technologies. However, any AI system has to provide benefits that outweigh the cost of handling data and assembling a specialized team (Rahman, Hossain, & Fattah, 2021). If not, you work for AI; not AI for you. Using AI may be challenging in a variety of ways. Firms should make sure that they have a solid use case for the project and the necessary funding. Consider alternatives such as automated workflows. For the reasons stated above, B2B firms need to carefully evaluate how to handle and resolve any ethical quandaries that may arise while implementing AI-based B2B marketing solutions.

Lastly, B2B success can be hampered by a number of factors, including leadership and lack of organizational readiness. In our case study this was not the case but it might be true for other firms who do not have the proper organizational structure (Di Vaio, Hassan, & Alavoine, 2022). While some of these obstacles are clear, others that typically obstruct achievement are more difficult to identify. B2B sales may be substantially more complex since there are more moving parts, more decision makers, longer sales cycles with more touch points, and more potential for mistakes. B2B enterprises usually thrive on long-term relationships, which means that it can be difficult for smaller B2B firms to establish a name and clientele among individuals accustomed to doing business with certain suppliers without using an advanced AI system that would probably be very costly to have. Table 3 summarises the current state and contribution of this paper.

Table 3
Summary of current state and contribution of this study.

Current B2B state	Paper contribution
Researchers explored the drawbacks of AI and sought to address them by incorporating AI into BA capabilities. They specifically focused on data management, governance, and training resources to effectively tackle these concerns (Akter et al., 2021; Li et al., 2023).	This study examined employee adaptation in B2B departments and their changing perceptions and goals. Our study delved into the dynamics between traders and AI developers influenced by AI, exploring how AI traders adjust in this context.
The need for more comprehensive studies on the influence of data, system quality, and end-user training on competitiveness, both directly and indirectly, in the context of utilizing AI-BA capabilities has been investigated (Rana et al., 2021).	This study provides insights into managers' actions, traders' reactions, and the dynamics between AI developers and traders. Managers implemented new procedures and policies to address concerns and maximize profits, including a cultural shift towards AI adoption and providing AI training.
Addressing accountability problems related to algorithmic misbehaviour, numerous ethical and legal concerns arise (Boyd & Wilson, 2017)	This study shows how models often failed to learn from new data despite having access to abundant data, which can be attributed to the specific nature of the intended prediction. Moreover, establishing a proper organizational structure is crucial for successful AI adoption in B2B firms, enabling seamless integration and utilization of AI technologies. In the B2B context, obtaining large volumes of high-quality data and effectively managing it are essential for AI implementation. However, extracting valuable insights from data remains a challenge for many B2B firms due to data relevance and management issues.
Determining true responsibility and allocating accountability for decisions made either solely by AI or in collaboration with a human agent poses a significant challenge (Mikalef et al., 2022).	Traders, although no longer directly involved, were still relied upon as domain experts, raising concerns about responsibility and accountability. The company prioritized the use of explainable AI over higher-margin options that carried potential unknown risks. Additionally, compliance was emphasized in areas where government regulations were not explicitly mandated.
The reduction of relationship bonds can result in unfavorable outcomes for B2B interactions (Gligor et al., 2021).	This study demonstrated the socio-economic effects. The company's actions influenced clients' behaviour regarding energy consumption, as energy prices varied based on the time of day.

5.1. Research implications

There are two distinct categories in which AI dark side effects can be classified. The first category encompasses harm inflicted upon the organization itself, while the second category pertains to harm inflicted upon others (Linstead et al., 2014). Our study highlights that many organizational aspects of a firm may change, including procedures that did not exist before, such as monitoring AI decisions. Investigating how to mobilize human capital will be vital for firms that do not want to damage their public image due to firing employees. Nevertheless, managers should deal with this internal issue and find a solution that is satisfying for traders since the most important part of their job is taken by AI. Equally important for research are other dark sides of AI in trading, such as negative implications on the individual (harm done to others) or the social level (Ibáñez & Olmeda, 2021). For example, AI fear, deskilling, and unemployment are concerning aspects that firms should not underestimate. Therefore, it is stated that, although adopting AI is vital, establishing the necessary procedures and mechanisms for building and aligning AI applications with business objectives is also critical (Bag et al., 2021; Saura et al., 2021). One of the most difficult parts of AI is that it is a technology that requires ongoing modification and change as new data and conditions emerge. As a result, there is a fleetingness that emphasizes recognizing the negative elements of AI in trading in order to ensure that the business continues to function as intended and that all organizational changes are in accordance with the firm's goals. In addition, firms should take into consideration the dark sides of AI in trading when planning, designing and building AI strategies and products, as AI can add value co-creation but co-destruction too, in aspects that range from job loss to privacy concerns, machine ethics, security issues and negative developments of superintelligence (Petrescu, Krishen, Kachen, & Gironda, 2022). Therefore, it is vital for organizations to effectively govern AI in a sense that promotes business goals, since AI governance should close the gap that exists between accountability and ethics in technological advancement (Papagiannidis, Enholm, Dremel, Mikalef, & Krogstie, 2022). For example, the difficulties encountered throughout the AI deployment process are obvious at various stages and influence diverse job duties, meaning that when it comes to difficult management tasks, AI solutions may give a selection of responses, as well as the likelihood of each of these choices (Papagiannidis et al., 2022).

5.2. Research agenda

There have been few studies examining the interpretation and prediction of behaviour in B2B contexts and the use of AI to govern organizational workflow, including employee and corporate procedures. For instance, Kushwaha, Kumar, and Kar (2021) developed a model that examined customer experience and trust, shedding light on their influence on the overall reputation of systems and brands. However, they did not specifically address the direct effects of AI in a B2B context, leaving an avenue for future research in that particular domain. While Davenport et al. (2020) discussed the broader impact of AI in the future, their exploration did not delve into the specific implications of AI in B2B environments. Their findings serve as a foundation for further research in this specific field. Similarly, Farrokhi et al. (2020) devised an AI model for anomaly detection aimed at averting external crisis events that could harm a firm's reputation. However, their study did not provide insights into the establishment of a centralized AI system tailored specifically for B2B environments, thereby leaving room for further investigation. In an similar direction, Grewal, Guha, Satornino, and Schweiger (2021) explored both the positive and negative aspects of AI in B2B and B2C contexts, with a predominant focus on the positive impacts on B2C firms. While they did address concerns regarding the black box nature of AI and potential opportunistic behaviour, further research is necessary to examine the social-economic effects of AI in B2B settings. Furthermore, Graef, Klier, Kluge, and Zolitschka (2021) discussed the concept of human-machine collaboration, emphasizing the significance of a feedback-based approach to provide accurate and reliable solutions. However, they did not delve into the specific practices required for the successful adoption of AI in such collaborative scenarios. Troisi, Maione, Grimaldi, and Loia (2020) emphasized the need for a framework that transforms data into actionable knowledge, fostering continuous learning, creativity, and improvement. They introduced the concept of a hacking mindset to achieve marketing objectives in B2B. Expanding their work to include practical implementation strategies and investigating AI-driven analytics could enhance their proposed framework. While Rai (2020) explored the concept of interpretable models and the conversion of black-box models to glass-box models, their overview was broad and did not focus on specific sectors like B2B. Additionally, they did not delve into critical decision-making processes in depth.

We propose a variety of ways in which researchers could investigate the dark side effects of AI in a B2B context for the future research agenda. Research could be directed towards the positive and negative effects of AI in business-to-business and the consequences that AI could have on a company's reputation and market positioning. Consequently, it is necessary to fully understand which effects AI might cause and how to deal with them, as well as the dark side effects of AI not only in a business but also in society as a whole. It is important to note that having an AI system does not guarantee success. To understand and utilize an AI product, a firm must incorporate an AI system into its culture and educate its employees, because vulnerabilities usually appear with the introduction of new technologies and AI systems are no exception to this rule. An AI hack can cause great financial loss to a B2B firm as a B2B firm usually has a limited number of clients, and losing one could be devastating. Furthermore, understanding and trusting your AI system is crucial for B2B, and there are a few technologies designed with B2B in mind that support bias removal and transparency. As a final note, it is equally important to investigate the ways to centralize a B2B ecosystem around AI, because B2B environments are very different from B2C ones. Table 4 provides a summary of the research questions that we propose to the readers, as well as references for further investigation and guidance.

Table 4
Future research agenda.

Themes	Research question	References for guidance
Nature of work	In which ways does AI affect B2B reputation and how should AI anxiety be addressed?	(Kushwaha et al., 2021)
	In a B2B environment, how does AI impact employee expectations of AI?	(Davenport et al., 2020)
	How should AI be adopted and used in a B2B culture?	(Keegan, Dennehy, & Naudé, 2022)
	How can an AI system be properly centralized in a B2B environment in order to detect anomalies in AI decisions?	(Farrokhi et al., 2020)
Conflicts and effects	What are the dark side social and economic effects on society due to the use of AI in B2B?	(Grewal et al., 2021)
	What are the best ways to mitigate conflicts when adapting a human-machine AI collaboration, and what practices should be created?	(Graef et al., 2021)
Responsibility	How is it feasible to secure a B2B AI system from its data and model outputs being hacked?	(Troisi et al., 2020)
	How can AI transparency tools be developed for a B2B firm while mitigating fatal AI model mistakes?	(Rai, 2020)

5.3. Practical implications

Regarding the practical implications, AI developers should not overlook the importance of explainability tools for addressing AI decisions. This is crucial as it allows the accountable employees, in this case, the traders, to acquire a better sense of accountability and feel safer when using AI for decision making (Paschen et al., 2019). In this way, managers can ensure safety for their employees and the processes or mechanisms they introduce and against legal regulations that might demand an explanation of how decisions are taken. What is more, firms need appropriate infrastructure to centralize their advanced systems (Al-Surmi, Bashiri, & Koliousis, 2022). AI, in particular, should be founded and developed to decrease inequality and promote social empowerment while preserving individual autonomy and enhancing advantages that are shared equitably by all (Puntoni, Reczek, Giesler, & Botti, 2021). AI must be explainable since it is a major tool for establishing public trust and understanding of the technology (Keegan, Dennehy, & Naudé, 2022). By doing this, monitoring AI is much easier, and it allows for reallocating employees from the position that AI takes over; thus, employees do not lose their jobs. Lastly, firms should develop tools for testing AI decisions to identify new patterns that would lead to a deeper understanding of data and information (Davenport et al., 2020). By doing so, employees would come up with new ideas that may lead to new strategies and tactics boosting productivity and innovation. Backtesting is common when traders want to test a theory using various hypotheses. Therefore, managers may prioritize such processes, allowing their traders to be more enlightened about their data and how they can be in the loop of actively improving AI trading bots through their expertise (Mikalef, Conboy, & Krogstie, 2021); thus, traders can contribute by innovating ideas that will potentially boost the competence of the algorithms and as a result the competence of the company.

5.4. Limitations and future research

This paper has looked into the negative aspects of AI in trading. A few limitations exist to this study. Firstly, the information was gathered through interviews with only one organization; as a result, our data may be biased or offer an inadequate picture of the problems surrounding relevant procedures. Secondly, while we performed multiple interviews with important individuals inside the organization, our data was gathered at a specific point in time and may not accurately reflect the full range of activities. Lastly, the company had used AI for only five years and in general had a positive experience due to the slow, steady and careful development of AI to prevent risks; therefore, the results may be affected by that factor. As a result, generalization may be a problem that should be considered. A future study might acquire additional empirical data through interviews and postulate the concept of the dark sides of AI in trading from a positivist viewpoint, which could be evaluated with empirical evidence on the antecedents and consequences. It would be useful for the field to understand how companies reallocate human resources to meet organizational goals and how they regulate AI resources to improve performance while keeping in mind the negative consequences for workers.

Role of the funding source

There was no funding.

Declaration of Competing Interest

None.

Data availability

The data that has been used is confidential.

References

Abrell, T., Pihlajamaa, M., Kanto, L., Vom Brocke, J., & Uebernickel, F. (2016). The role of users and customers in digital innovation: Insights from B2B manufacturing firms. *Information & Management, 53*(3), 324–335.

Abubakar, A. M., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation & Knowledge, 4*(2), 104–114.

Ahl, A., Yarime, M., Goto, M., Chopra, S. S., Kumar, N. M., Tanaka, K., & Sagawa, D. (2020). Exploring blockchain for the energy transition: Opportunities and challenges based on a case study in Japan. *Renewable and Sustainable Energy Reviews, 117*, Article 109488.

Akter, S., Wamba, S. F., Mariani, M., & Hani, U. (2021). How to build an AI climate-driven service analytics capability for innovation and performance in industrial markets? *Industrial Marketing Management, 97*, 258–273.

Al-Surmi, A., Bashiri, M., & Koliousis, I. (2022). AI based decision making: Combining strategies to improve operational performance. *International Journal of Production Research, 60*(14), 4464–4486.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics, 54*, Article 101473.

Azzutti, A., Ringe, W.-G., & Stiehl, H. S. (2022). Machine learning, market manipulation, and collusion on capital markets: Why the "black box" matters. European banking institute working paper series 2021 - No. 84. *University of Pennsylvania Journal of International Law, 43*. https://doi.org/10.2139/ssrn.3788872

Bag, S., Gupta, S., Kumar, A., & Sivarajah, U. (2021). An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. *Industrial Marketing Management, 92*, 178–189.

Barta, G., & Görcsi, G. (2021). Risk management considerations for artificial intelligence business applications. *International Journal of Economics and Business Research, 21*(1), 87–106. https://doi.org/10.1504/ijebr.2021.112012

Behera, R. K., & Bala, P. K. (2023). Unethical use of information access and analytics in B2B service organisations: The dark side of behavioural loyalty. *Industrial Marketing Management, 109*, 14–31.

Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management, 57*, Article 102225.

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? *Policy Quarterly, 13*(4). Available at: https://apo.org.au/sites/default/files/resource-files/2017-11/apo-nid120146.pdf.

Camilleri, J., & Neuhofer, B. (2017). Value co-creation and co-destruction in the Airbnb sharing economy. *International Journal of Contemporary Hospitality Management, 29*(9), 2322–2340. https://doi.org/10.1108/IJCHM-09-2016-0492

Candi, M., Roberts, D. L., Marion, T., & Barczak, G. (2018). Social strategy to gain knowledge for innovation. *British Journal of Management, 29*(4), 731–749.

Castillo, D., Canhoto, A. I., & Said, E. (2021). The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. *The Service Industries Journal, 41*(13–14), 900–925.

CBNC. (2017). Stephen Hawking says A.I. could be 'worst event in the history of our civilization'. Retrieved 19-09-2022 from https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html.

CBNC. (2021). Elon Musk warned of a 'Terminator'-like AI apocalypse — Now he's building a Tesla robot. https://www.cnbc.com. Retrieved 19-09-2022 from https://www.cnbc.com/2021/08/24/elon-musk-warned-of-ai-apocalypsenow-hes-building-a-tesla-robot.html.

Chan, S., & Ip, W. (2011). A dynamic decision support system to predict the value of customer for new product development. *Decision Support Systems, 52*(1), 178–188.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage Publications.

Charmaz, K. (2014). *Constructing grounded theory*. Sage Publications.

Chatterjee, S., Rana, N. P., Tamilmani, K., & Sharma, A. (2021). The effect of AI-based CRM on organization performance and competitive advantage: An empirical analysis in the B2B context. *Industrial Marketing Management, 97*, 205–219.

Chen, L., Jiang, M., Jia, F., & Liu, G. (2022). Artificial intelligence adoption in business-to-business marketing: Toward a conceptual framework. *Journal of Business & Industrial Marketing, 37*(5), 1025–1044.

Chen, L.-T., & Chen, J.-M. (2008). Collaborative marketing and production planning with IFS and SFI production styles in an ERP system. *Journal of the Chinese Institute of Industrial Engineers, 25*(4), 337–346.

Chen, Y. (2017). Integrated and intelligent manufacturing: Perspectives and enablers. *Engineering, 3*(5), 588–595. https://doi.org/10.1016/J.ENG.2017.04.009

Cheng, X., Su, L., Luo, X., Benitez, J., & Cai, S. (2021). The good, the bad, and the ugly: Impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing. *European Journal of Information Systems, 31*(3), 1–25. https://doi.org/10.1080/0960085X.2020.1869508

Cortez, R. M., & Johnston, W. J. (2017). The future of B2B marketing theory: A historical and prospective analysis. *Industrial Marketing Management, 66*, 90–102. https://doi.org/10.1016/j.indmarman.2017.07.017

Cruz, J. M. (2009). The impact of corporate social responsibility in supply chain management: Multicriteria decision-making approach. *Decision Support Systems, 48*(1), 224–236.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42.

Demlehner, Q., & Laumer, S. (2020). Shall we use it or not? Explaining the adoption of artificial intelligence for car manufacturing purposes. In *28th European conference on information systems - liberty, equality, and fraternity in a digitizing world, Marrakech Morocco*.

Di Vaio, A., Hassan, R., & Alavoine, C. (2022). Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change, 174*, Article 121201.

Dwivedi, Y. K., & Wang, Y. (2022). Guest editorial: Artificial intelligence for B2B marketing: Challenges and opportunities. *Industrial Marketing Management, 105*, 109–113. https://doi.org/10.1016/j.indmarman.2022.06.001

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial intelligence and business value: A literature review. *Information Systems Frontiers, 24*(5), 1–26. https://doi.org/10.1007/s10796-021-10186-w

Farrokhi, A., Shirazi, F., Hajli, N., & Tajvidi, M. (2020). Using artificial intelligence to detect crisis related to events: Decision making in B2B by artificial intelligence. *Industrial Marketing Management, 91*, 257–273. https://doi.org/10.1016/j.indmarman.2020.09.015

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal, 29*(13), 1465–1474. https://doi.org/10.1002/smj.722

Gligor, D. M., & Esmark, C. L. (2015). Supply chain friends: The good, the bad, and the ugly. *Business Horizons, 58*(5), 517–525.

Gligor, D. M., & Holcomb, M. (2013). The role of personal relationships in supply chains: An exploration of buyers and suppliers of logistics services. *The International Journal of Logistics Management, 24*(3), 328–355. https://doi.org/10.1108/IJLM-07-2012-0067

Gligor, D. M., Pillai, K. G., & Golgeci, I. (2021). Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain. *Journal of Business Research, 133*, 79–88. https://doi.org/10.1016/j.jbusres.2021.04.043

Graef, R., Klier, M., Kluge, K., & Zolitschka, J. F. (2021). Human-machine collaboration in online customer service–A long-term feedback-based approach. *Electronic Markets, 31*, 319–341.

Grewal, D., Guha, A., Satornino, C. B., & Schweiger, E. B. (2021). Artificial intelligence: The light and the darkness. *Journal of Business Research, 136*, 229–236. https://doi.org/10.1016/j.jbusres.2021.07.043

Grönroos, C., & Voima, P. (2013). Critical service logic: Making sense of value creation and co-creation. *Journal of the Academy of Marketing Science, 41*(2), 133–150.

Grundner, L., & Neuhofer, B. (2021). The bright and dark sides of artificial intelligence: A futures perspective on tourist destination experiences. *Journal of Destination Marketing & Management, 19*, Article 100511.

Hallikainen, H., Savimäki, E., & Laukkanen, T. (2020). Fostering B2B sales with customer big data analytics. *Industrial Marketing Management, 86*, 90–98.

Han, R., Lam, H. K., Zhan, Y., Wang, Y., Dwivedi, Y. K., & Tan, K. H. (2021). Artificial intelligence in business-to-business marketing: A bibliometric analysis of current research status, development and future directions. *Industrial Management & Data Systems, 121*(12), 2467–2497.

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172.

Ibáñez, J. C., & Olmeda, M. V. (2021). Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI & Society, 37*(4), 1663–1687. https://doi.org/10.1007/s00146-021-01267-0

Ives, B., Palese, B., & Rodriguez, J. A. (2016). Enhancing customer service through the internet of things and digital data streams. *MIS Quarterly Executive, 15*(4), 279–297.

Kamis, A., Koufaris, M., & Stern, T. (2008). Using an attribute-based decision support system for user-customized products online: An experimental investigation. *MIS Quarterly, 32*(1), 159–177. https://doi.org/10.2307/25148832

Keegan, B. J., Canhoto, A. I., & Yen, D. A.-W. (2022). Power negotiation on the tango dancefloor: The adoption of AI in B2B marketing. *Industrial Marketing Management, 100*, 36–48. https://doi.org/10.1016/j.indmarman.2021.11.001

Keegan, B. J., Dennehy, D., & Naudé, P. (2022). Implementing artificial intelligence in traditional B2B marketing practices: An activity theory perspective. *Information Systems Frontiers*, 1–15.

Kim, J. H., Kim, M., Kwak, D. W., & Lee, S. (2022). Home-tutoring services assisted with technology: Investigating the role of artificial intelligence using a randomized field experiment. *Journal of Marketing Research, 59*(1), 79–96.
Kromkowski, P., Li, S., Zhao, W., Abraham, B., Osborne, A., & Brown, D. E. (2019). Evaluating statistical models for network traffic anomaly detection. In *2019 systems and information engineering design symposium (SIEDS), Charlottesville, VA, USA*.
Kushwaha, A. K., Kumar, P., & Kar, A. K. (2021). What impacts customer experience for B2B enterprises on using AI-enabled chatbots? Insights from big data analytics. *Industrial Marketing Management, 98*, 207–221.
Lagna, A., & Lenglet, M. (2020). The dark side of liquidity: Shedding light on dark pools' marketing and market-making. *Consumption Markets & Culture, 23*(4), 390–406.
Li, B., Liu, L., Mao, W., Qu, Y., & Chen, Y. (2023). Voice artificial intelligence service failure and customer complaint behavior: The mediation effect of customer emotion. *Electronic Commerce Research and Applications, 59*, Article 101261.
Li, J., & Huang, J.-S. (2020). Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society, 63*, Article 101410. https://doi.org/10.1016/j.techsoc.2020.101410
Li, S. (2000). The development of a hybrid intelligent system for developing marketing strategy. *Decision Support Systems, 27*(4), 395–409.
Li, S., Peng, G., Xing, F., Zhang, J., & Zhang, B. (2021). Value co-creation in industrial AI: The interactive role of B2B supplier, customer and technology provider. *Industrial Marketing Management, 98*, 105–114.
Liang, W., Li, K.-C., Long, J., Kui, X., & Zomaya, A. Y. (2019). An industrial network intrusion detection algorithm based on multifeature data clustering optimization model. *IEEE Transactions on Industrial Informatics, 16*(3), 2063–2071.
Linstead, S., Maréchal, G., & Griffin, R. W. (2014). Theorizing and researching the dark side of organization. *Organization Studies, 35*(2), 165–188.
Liu, X. (2020). Analyzing the impact of user-generated content on B2B firms' stock performance: Big data analysis with machine learning methods. *Industrial Marketing Management, 86*, 30–39.
Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management, 42*(4), 489–495.
Meire, M., Ballings, M., & Van den Poel, D. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. *Decision Support Systems, 104*, 26–37.
Michael, D. M. (1997). Qualitative research in information systems. *MIS Quarterly Executive, 21*(2), 241–242.
Mikalef, P., Conboy, K., & Krogstie, J. (2021). Artificial intelligence as an enabler of B2B marketing: A dynamic capabilities micro-foundations approach. *Industrial Marketing Management, 98*, 80–92.
Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems, 31*(3), 257–268.
MIT, T. R. I. (2018). Professional services firms see huge potential in machine learning. https://www.technologyreview.com/2018/11/02/139216/professional-services-firms-see-huge-potential-in-machine-learning.
Moradi, M., & Dass, M. (2022). Applications of artificial intelligence in B2B marketing: Challenges and future directions. *Industrial Marketing Management, 107*, 300–314.
Oates, B. J. (2005). *Researching information systems and computing*. Sage Publications.
Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2022). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers, 25*(1), 123–141. https://doi.org/10.1007/s10796-022-10251-y
Paschen, J., Kietzmann, J., & Kietzmann, T. C. (2019). Artificial intelligence (AI) and its implications for market knowledge in B2B marketing. *Journal of Business & Industrial Marketing, 34*(7), 1410–1419. https://doi.org/10.1108/JBIM-10-2018-0295
Paschen, J., Paschen, U., Pala, E., & Kietzmann, J. (2021). Artificial intelligence (AI) and value co-creation in B2B sales: Activities, actors and resources. *Australasian Marketing Journal, 29*(3), 243–251.
Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons, 63*(3), 403–414.
Petrescu, M., Krishen, A. S., Kachen, S., & Gironda, J. T. (2022). AI-based innovation in B2B marketing: An interdisciplinary framework incorporating academic and practitioner perspectives. *Industrial Marketing Management, 103*, 61–72.
Plé, L., & Cáceres, R. C. (2010). Not always co-creation: Introducing interactional co-destruction of value in service-dominant logic. *Journal of Services Marketing, 24*(6), 430–437. https://doi.org/10.1108/08876041011072546
Pradhan, S., Ghose, D., & Shabbiruddin. (2020). Present and future impact of COVID-19 in the renewable energy sector: A case study on India. In *Energy sources, part A: Recovery, utilization, and environmental effects* (pp. 1–11).
Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing, 85*(1), 131–151.
Ragins, B. R., & Sundstrom, E. (1989). Gender and power in organizations: A longitudinal perspective. *Psychological Bulletin, 105*(1), 51.
Rahman, M. S., Hossain, M. A., & Fattah, F. A. M. A. (2021). Does marketing analytics capability boost firms' competitive marketing performance in data-rich business environment? *Journal of Enterprise Information Management, 35*(2), 455–480.
Rai, A. (2020). EXplainable AI: From black boX to glass boX. *Journal of the Academy of Marketing Science, 48*(1), 137–141.
Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2021). Understanding dark side of artificial intelligence (AI) integrated business analytics: Assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems, 31*(3), 364–387. https://doi.org/10.1080/0960085X.2021.1955628
Rashid, Y., Rashid, A., Warraich, M. A., Sabir, S. S., & Waseem, A. (2019). Case study method: A step-by-step guide for business researchers. *International Journal of Qualitative Methods, 18*. https://doi.org/10.1177/1609406919862424, 1609406919862424.
Risius, M., & Spohrer, K. (2017). A blockchain research framework. *Business & Information Systems Engineering, 59*(6), 385–409.
Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. *Industrial Marketing Management, 98*, 161–178.
Sharma, P., Kingshott, R., Leung, T. Y., & Malik, A. (2022). Dark side of business-to-business (B2B) relationships. *Journal of Business Research, 144*, 1186–1195.
Stone, M., Aravopoulou, E., Ekinci, Y., Evans, G., Hobbs, M., Labib, A., … Machtynger, L. (2020). Artificial intelligence (AI) in strategic marketing decision-making: A research agenda. *The Bottom Line, 33*(2), 183–200.
Sun, S., Hall, D. J., & Cegielski, C. G. (2020). Organizational intention to adopt big data in the B2B context: An integrated view. *Industrial Marketing Management, 86*, 109–121.
Sun, Y., Li, S., & Yu, L. (2022). The dark sides of AI personal assistant: Effects of service failure on user continuance intention. *Electronic Markets, 32*(1), 17–39.
Troisi, O., Maione, G., Grimaldi, M., & Loia, F. (2020). Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management, 90*, 538–557.
Vanderelst, D., & Winfield, A. (2018). The dark side of ethical robots. In *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society, New Orleans, LA, USA*.
Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems, 48*, 144–156.
Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration, 43*(9), 818–829.
Wynn, D., Jr., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly, 36*(3), 787–810. https://doi.org/10.2307/41703481
Yen, C., & Chiang, M.-C. (2021). Trust me, if you can: A study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. *Behaviour & Information Technology, 40*(11), 1177–1194.
Yin, R. K. (1981). The case study crisis: Some answers. *Administrative Science Quarterly, 26*(1), 58–65. https://doi.org/10.2307/2392599

PAPER 5

From Responsible AI Governance to Competitive Performance: The Mediating Role of Knowledge Management Capabilities

Emmanouil Papagiannidis, Patrick Mikalef, John Krogstie, and Kieran Conboy

(Lecture Notes in Computer Science (LNCS))

From Responsible AI Governance to Competitive Performance: The Mediating Role of Knowledge Management Capabilities

Emmanouil Papagiannidis[1](B), Patrick Mikalef[1], John Krogstie[1], and Kieran Conboy[2]

[1] Norwegian University of Science and Technology, Trondheim, Norway
emmanouil.papagiannidis,patrick.mikalef,john.krogstie @ntnu.no
[2] National University of Ireland, Galway, Ireland
kieran.conboy@nuigalway.ie

Abstract. In a constantly changing environment, researchers and practitioners are concerned with the issue of whether responsible artificial intelligence (AI) governance can help build competitive advantage. Responsible AI governance should be viewed as a source of competitive edge rather than merely a quick fix for automating manual processes. Despite this, little empirical evidence is available to support this claim, and even less is understood about the dimensions and relationships that add business value. This paper develops a conceptual model to explain how responsible AI governance practices aligned with strategic goals lead to competitive performance gains. An investigation of 144 Nordic firms is conducted to verify our hypotheses using a PLS-SEM analysis. Findings reveal that deploying responsible AI governance will make a significant positive impact on an organizations' knowledge management capabilities directly and on competitive performance indirectly. These findings also suggest that implementing responsible AI governance improves firms' ability to acquire and distribute knowledge when there is strategic alignment with a firm's goals.

Keywords: Responsible AI governance · Knowledge management capabilities · Competitive performance · Strategic alignment

1 Introduction

Over the past few years, organizations are increasingly turning to AI to digitalize their activities. Schmidt, Zimmermann, Möhring and Keller [1] define AI as the endeavor to mimic cognitive and human capabilities on computers. AI contributes towards digital transformation by customizing solutions based on the available data [2]. Nevertheless, AI capabilities have not been used to their full advantage and companies like Google decided to govern AI in a responsible way to increase performance and limit negative consequences [3]. Another example is IBM who developed several tools to address fairness issues [4]. Responsible AI is the process of designing, developing, and deploying artificial intelligence with the purpose of enabling individuals and organizations

Published by Springer Nature Switzerland AG 2022
S. Papagiannidis et al. (Eds.): I3E 2022, LNCS 13454, pp. 58–69, 2022.
https://doi.org/10.1007/978-3-031-15342-6_5

while also having a fair effect on customers and society, allowing businesses to create trust towards AI [5]. The relationship between responsible AI governance and a company's competitive performance is an important issue that has occupied information systems research for the last few years [6]. A growing body of research in Information Systems (IS) emphasizes the importance of developing responsible applications that transform competencies into differential economic value, with some studies attempting to determine the impact of responsible AI governance on a firm's competitive position [7]. Among other things, Knowledge Management Capabilities (KMC) are expected to enable firms to seek and disseminate updated knowledge in order to meet their needs, exploit innovation, and guide firms in responding quickly to external market changes to achieve high business performance [8, 9]. Furthermore, as competition has increased the need for dynamic capabilities in organizations, future research should focus on the conceptual development of dynamic KMCs incorporating new facets to resolve real-time problems and achieve better organizational outcomes [10]. Nevertheless, little empirical evidence is available to support this claim, and even less is understood about the dimensions and relationships that add business value.

To fill this gap, we developed a conceptual model, and developed an instrument of responsible AI governance based on the guidelines of MacKenzie, et al. [11] to explain how responsible AI governance practices aligned with digitalization goals lead to competitive performance gains. We collected data through a quantitative survey-based approach in which 144 Nordic businesses participated, and we examined our conceptual model and hypotheses using a partial least squares structural equation model (PLS-SEM) analysis. Therefore, the aim of this study is to determine whether responsible AI governance improves KMC, whether strategic alignment changes the strength between responsible AI governance and KMC, in other words, it has a moderating effect, and whether KMC provide any significant competitive performance advantages. The main point is that responsible AI governance will be useful only if it is used to support or enable critical KMC that are contributed by dynamic strategy alignment [12]. What is more, we contribute to the AI literature by demonstrating how responsible AI governance enhances a company's KMC, hence improving competitive performance. Using survey data from respondents with managerial responsibilities within their organization, we show empirical proof that these claims are correct. As a result, responsible AI governance reflects a company's ability to commercialize its knowledge skills. Consequently, this study seeks to answer the following two research questions: (1) *What is the relationship between responsible AI governance and competitive advantage gains? (2) What is the effect of strategic alignment on competitive performance gains?*

The rest of the paper is structured as follows. The subsequent section presents the background of this study and describes what responsible AI governance entails. Section 3 details our hypotheses, where we examine if a robust responsible AI governance can impact competitive performance. It is our theory that the indirect effect is mediated through the firm's capability to manage its knowledge, which is affected by the strategic alignment. As a result, these renewed operational capabilities provide a competitive advantage. Section 4 presents how our study analyzes factors associated with these associations using a survey-based design, and we describe the data collection methods and measures for each concept used. Afterwards, we present the results of our empirical analysis, followed by a discussion of their theoretical and practical implications as well as some significant limitations.

2 Background

Although there is no clear definition of responsible AI governance, there is a growing consensus about it. It can be defined as a function that describes the different ways AI can be governed ethically [5]. As an alternative, it can be defined as a process that spans all stages of AI projects' lifecycles by following the principles of responsible use [13]. Responsible AI governance is important to benchmark against competitive performance gains, particularly in examining what type of effect it has within organizations' capabilities to make continuous improvements and implement changes in business products, methods and services. For instance, Microsoft developed explainability tools to interpretate machine learning models which assist with decision making [4]. Hence, there is growing support for the claim that responsible AI governance not only has an impact on external entities' perception of organizations when using AI [14], but also on the internal capabilities related to managing organizational knowledge [8]. Consequently, responsible AI governance may influence KMC since it offers a framework for understanding the implications of the use of AI and propose which standards to follow so stakeholders will have confidence in the organization's use of AI. KMC is defined as an organizational mechanism to continually and intentionally create knowledge inside the organization [15].

Developing responsible artificial intelligence applications adds benefits not only from an ethical and moral standpoint but also can provide organizations with a medium to long-term competitive advantage [16]. By showcasing an organization's commitment to ethical practices, for example, it can gain an edge in recruiting technical professionals and also retain top talent, particularly when qualified developers are in short supply. According to the EIU report [17], ethically questionable practices discourage prospective employees from applying for jobs and undermine their faith in the industry, contributing to the so-called "techlash," a result of public disbelief and animosity towards large tech companies. Furthermore, responsible AI practices and processes enable the creation of documentation on how an organization addresses the challenges associated with artificial intelligence [16], allowing for a better understanding of potential operational issues or business opportunities that may arise [17]. As a result, responsible AI commences influencing performance because trustworthiness leads to increased retention, spending, and adoption of new services [18]. A well-crafted AI application can preserve and expand one's client base by adhering to ethical and responsible standards [16]. By developing inclusive products and services, businesses will be able to retain customers and increase their credibility by providing products and services that are effective for all types of customers, ensure safety, and are transparent. For instance, the acceptance of blockchain technologies in AI services for traceability and transparency can overcome trust issues from the side of customers [19]. Additionally, the development of a responsible AI governance is also essential from a compliance perspective. Authorities have begun monitoring AI applications and introducing regulations that include principles of standards and ethical considerations, such as auditing processes and algorithmic impact assessments. As a result, a number of privacy and data protection frameworks include privacy by design as an integral part of their frameworks.

Seven dimensions comprise the notion of responsible AI governance. These dimensions are accountability, environmental, societal well-being, transparency, fairness, robustness and safety, data governance, and human-centric AI [20]. A primary aim of responsible AI governance is to reduce the possibility that a modest change in the weight of an input can drastically alter the output of a machine learning model since it takes a lot of effort to create a responsible AI governance system. It is worth mentioning that this is a self-developed construct, where items and sub-dimensions have been validated through established methods [11]. Continuous examination is necessary to guarantee that an organization is dedicated to producing unbiased and reliable AI. Therefore, while creating and implementing an AI system, it is critical for a business to have a maturity model or standards to follow. The ability of an organization to effectively adapt information to future use and respond to changes in the environment is critical, as is the importance of knowledge in improving the organization's performance. KMC reflects an organization's ability to create, transfer, integrate and leverage knowledge within the organization [21]. The items used to measure the KMC of firms were adopted from the study of Mao, Liu, Zhang and Deng [22], where they also were empirically confirmed. The respondents were asked five questions about the degree to which they are able to manage knowledge within the organization.

Strategic alignment has been a top management priority since the inception of the information technology profession, and its favorable effects on business performance have been thoroughly documented in past research. The substance of plans and planning processes can be viewed as strategic alignment [23]. Tallon and Pinsonneault [24] supported the causal link between strategic alignment and performance by concentrating on the alignment of strategy, plans, operations, and processes. We measured strategic alignment based on an adapted scale used from the work of Preston and Karahanna [25] which comprised of three items. As for competitive performance, it refers to how well a company outperforms its key competitors [26]. Respondents were asked to rate how well they outperformed their primary competitors in a variety of areas such as market share, delivery cycle time, and customer satisfaction.

3 Research Model

The research model is presented (see Fig. 1), as are the hypotheses that surround it. We argue that responsible AI governance will affect a company's competitive performance. We also argue that strategic alignment between AI and business will amplify responsible AI governance's impact on KMC. Therefore, having a sensible AI governance model aligned with a company's strategic goals will enhance KMC, which will boost the company's competitive performance.

An organization that implements responsible AI governance should focus on assessing, monitoring, and evaluating the performance of an AI application, both before and after deployment [13]. It is also essential to have clear and concise ways to document both the data and AI aspects of AI governance, including how they relate to each other [3]. For instance, all steps from data collection to data use should be documented, including

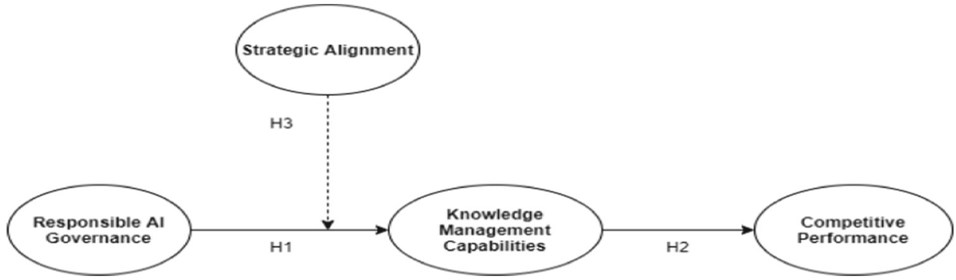

Fig. 1. Research model.

how the data was transformed [27]. Knowledge can flow within the organization more efficiently if the processes and mechanisms are well documented, therefore improving KMC. Documentation also decreases the dependence on one person's knowledge and abilities since documented processes are less likely to become obsolete. Furthermore, responsible AI governance places emphasis on designing based on inclusiveness and on enhancing human agency and autonomy. These are fundamental constituents for facilitating better use of human capital within organizations and, as a result, optimizing knowledge flows and interactions. Finally, responsible AI governance dictates that throughout the process of design, deployment and monitoring of AI applications, there is a strong focus on the safety and robustness of systems and entities that interact with AI agents. Establishing such privacy and safety policies facilitates easier cross-departmental access and knowledge sharing without the risk of critical knowledge being leaked or accessed by non-authorized employees [28]. Thus, we propose the following hypothesis: **H1:** *Responsible AI governance will have a positive effect on KMC.*

Several studies have indicated that KMC is related to organizational performance [21, 22, 29]. Information is regarded as one of a company's most valuable and critical resources, and businesses that can develop, apply, and manage the appropriate knowledge can reap a variety of benefits [29]. Having a strong KMC may aid in the improvement of product and service quality, as well as the development of new products and services. KMC helps businesses improve their processes, which is critical for competitive success. Several studies have been conducted to investigate the relationship between KMC and performance. Tanriverdi [21], for example, demonstrated that KMC has a positive effect on the corporate financial performance of multi-business firms. Thus, we propose the following hypothesis: **H2:** *KMC will have a positive effect on competitive performance.*

Strategic alignment of the information system has been linked to improved company performance (IS). It refers to the alignment of a company's business strategy with its information technology strategy. Other studies [30] investigated a wide range of factors that influence the alignment of business and IT strategies. Strategic alignment improves organizational outcomes, which indirectly increases competitive advantage. KM is also a strategy for developing new products, increasing value, and improving competitiveness by leveraging a company's intellectual assets and employee capabilities. In this scenario, if we view strategic alignment as the most important to the organization, and design a roadmap to accomplish its goals, then responsible AI governance could be the roadmap (framework) and KMC could be one of the organizational goals. That means strategic alignment could amplify the potential value that responsible AI governance has on KMC. Hence, to compete in today's highly competitive business environment, large corporations must integrate their IT with their KM policies and procedures. Thus, we propose the following hypothesis: **H3:** *S*trategic alignment *will have a positive moderating effect on the relationship between responsible AI governance and KMC.*

4 Methodology

A quantitative study was carried out to test the research paradigm proposed in this work. The survey approach was used as a strategy as a survey study collects the same type of data from a large number of key respondents such as managers, heads of departments and CEO, which can then be analyzed for trends that allow conclusions to be generalized. The study's population is Nordic enterprises because according to the Global Economic Forum's 2019 Global Competitiveness Report organizations in these countries have high levels of ICT adoption and the majority of individuals have strong digital skills, making them well equipped for digital transformation.

4.1 Data Collection

To put the study model to the test, Nordic businesses were sent an internet questionnaire-based survey. For each country, the percentage is 29.9% for Norway, 27.8% for Sweden, 27.7% for Finland and 14.6% for Denmark. According to the Global Economic Forum's 2019 Global Competitiveness Report, these countries are at the forefront of global competitiveness, ranking eighth, tenth, eleventh, and seventeenth, respectively [31]. The Nordic countries have a high rate of ICT adoption, and the majority of the population has advanced digital skills, putting them in a good position for digital transformation [32]. We utilized 58 questions to measure our items and we used a 7-point Likert scale, where a value of 1 means disagrees entirely, and 7 means agree entirely.

To ensure internal validity we used PLS' discriminant validity which establishes the distinctiveness of the constructs. In addition, we conducted a pre-study of 15 respondents to measure the statistical responses (respondent fatigue, quality of answers etc.) and we requested feedback from them to improve the survey. For external validity, we used purposive sampling as it is easier to generalize a sample of 144 respondents. Our sample is consistent since it exclusively covers Scandinavian nations, implying that they share comparable cultural traits, education level and IT infrastructure. The validity and reliability of the hierarchical research model were evaluated using a structural equation model (PLS-SEM). All analyses, in particular, were carried out using the software package SmartPLS 3. PLS-SEM is regarded as a suitable method for assessing multiple relationships between one or more dependent variables and one or more independent variables in this study because it allows for simultaneous estimation of multiple relationships [33]. As a variance-based method, PLS-SEM is adaptable and capable of evaluating both reflective and formative constructs and the ability to analyze complex models with smaller samples and theory building. PLS-SEM is widely used in data analysis for the estimation of complex relationships between constructs in a variety of subject areas, including business and management research [34].

5 Analysis

5.1 Measurement Model

We employed distinct assessment criteria to examine each of the reflective and formative constructs in the model because they are both reflective and formative. We tested reliability, convergent validity, and discriminant validity for the latent reflective components. The construct and item levels of reliability were tested. We looked at Composite Reliability (CR) and Cronbach Alpha (CA) values at the construct level and found that they were both over the 0.70 criterion. The construct-to-item loadings were checked to see if they were greater than 0.70, indicating indicator dependability. To see if AVE values were convergent, we looked at whether they were above the lower limit of 0.50, and the lowest observed value was 0.623, which is significantly higher than this threshold. To establish discriminant validity, two methods were used. The Fornell–Larcker criterion was used to ensure that the AVE square root of each construct was more significant than the highest correlation with any other construct. The second examined whether the outer loading of each indicator exceeded its cross-loadings with other constructs. The results (Table 1) show that the reflective measures are valid and that all items are good indicators for their respective constructs.

Table 1. Discriminant validity values.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Accountability	**0.78**									
(2) Data governance	0.63	**0.75**								
(3) Environmental and societal well-being	0.64	0.59	**0.81**							
(4) Fairness	0.62	0.55	0.60	**0.76**						
(5) Human-centric AI	0.58	0.73	0.63	0.52	**0.79**					
(6) Robustness and safety	0.69	0.73	0.62	0.58	0.75	**0.76**				
(7) Transparency	0.66	0.64	0.59	0.70	0.64	0.68	**0.69**			
(8) KMC	0.33	0.60	0.48	0.56	0.58	0.43	0.54	**0.82**		
(9) Strategic alignment	0.35	0.39	0.58	0.37	0.57	0.45	0.51	0.45	**0.91**	
(10) Competitive performance	0.47	0.68	0.59	0.55	0.62	0.48	0.58	0.75	0.54	**0.79**

5.2 Structural Model

Figure 2 summarizes the structural model from the PLS analysis by showing the explained variance of endogenous variables (R^2) and the standardized path coefficients (β). The structural model is validated using coefficients of determination (R^2). To determine the significance of estimates, a bootstrap approach with 10000 resamples is used (t-statistics).

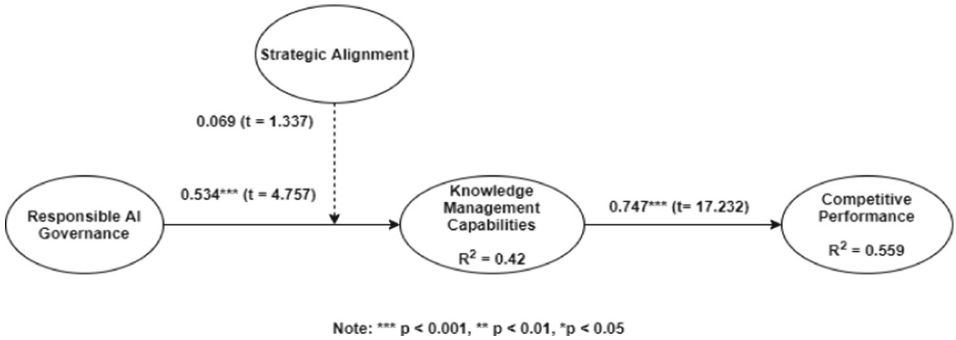

Fig. 2. Structural model.

Figure 2 depicts support for two of the three hypotheses. The responsible AI governance of a company is found to influence KMC ($\beta = 0.534$, $t = 4.757$, $p < 0.001$). Strategic alignment, on the other hand, had no such significant effect on knowledge management competencies ($\beta = 0.069$, $t = 1.337$, $p > 0.05$). As predicted, knowledge management skills are positively associated with competitive performance ($\beta = 0.747$, $t = 17.232$, $p < 0.001$). The structural model explains 62.3% of the variation in competitive performance ($R^2 = 0.559$), 66.9% of the variation in KMC ($R^2 = 0.42$), and 82.7% of the variation in strategic alignment ($R^2 = 0.42$). These coefficients of determination suggest that the data have moderate to significant predictive power.

6 Discussion

There is a growing discussion around responsible AI governance, but still, literature lacks empirical evidence and thus, there is a gap that needs to be filled. Businesses should bridge this gap in order to gain the trust of their customers, employees, and other stakeholders. If they don't, their competitive performance might suffer, and their AI initiatives could fail to deliver the expected benefits and value. The outcomes of this research contribute to IS literature through key findings which raise several theoretical and managerial implications.

6.1 Implications for Research

This study contributes by developing a construct model for responsible AI governance and by examining how it affects KMC and through that competitive performance. We provide empirical evidence on the notion that responsible AI governance has an effect on KMC and has an indirect effect on competitive performance and we validate the concept through an empirical study that builds on a large sample from Scandinavian companies. Hence, policies and goals that define and orchestrate the business plan should consider how responsible AI governance can affect directly or not the performance outcomes of a firm.

In more specific terms, responsible AI governance appears to directly impact a firm's KMC by expanding both knowledge assets as well as knowledge operating capacities and providing them with greater opportunity in terms of their capacity, competence, and ability. Through this study, we add to the AI literature by illustrating how responsible AI governance improves a company's KMC, thereby enhancing competitive performance. We provide empirical evidence that these claims are true using survey data from 160 respondents with managerial responsibilities within their firm. According to Rana, Chatterjee, Dwivedi and Akter [35], the lack of understanding of how unintended consequences of an AI system could impact the overall competitive position of a firm is vital to the development and implementation of responsible AI government frameworks that add business value. Thus, exploring and investigating how responsible AI governance frameworks should function could give companies a competitive advantage over their competition.

Despite this, our empirical results did not support the assumption that strategic alignment impacts KMC. This may be because managers develop processes, policies, and practices from the top, and from there drill down to the bottom, while in practice, responsive AI is implemented from the bottom up based on the technical skills of the AI team. This entails pushing for changes in structures and processes may be antithetical to their goals and may conflict with what the organization currently supports [36]. Also, it is imperative that managers who wish to incorporate responsible AI concerns into their work first understand what it takes to achieve this, and then take the necessary steps to develop a responsible AI system. In the absence of an AI governance framework, it is a huge undertaking to redesign organizational structures, accommodate the responsible AI work, and finally carry out management changes to implement the new organizational practices.

6.2 Implications for Practice

The results of this study can be used by managers in key positions to benchmark results and identify areas for improvement. To accomplish this, a multitude of processes must be implemented, which requires top management commitment and a clear plan for firm-wide responsible AI integration. Since many companies are still at an early stage of adopting AI practices, it is important to do it in a responsible manner in order to gain value from the building of new capabilities which can boost performance. Since AI systems are complex and expensive, additional implementation considerations should be designed into the overall design, yet the benefits can quickly be realized on a managerial and economic level.

Aside from the fact that responsible AI governance practices enhance ethical and competitive value for the company, which is good for public perception, executives should also adopt them to improve the company's performance. Of course, there is the human factor to consider, as responsible AI is concerned with how human agents make data-driven decisions in order to maximize potential business performance [37]. At the same time, responsible AI governance has an impact on a company's overall strategy and development planning because features related to responsible AI necessitate time and effort. In contrast, the development team requires appropriate management and resources to create a trustworthy system.

Finally, due to the complexity of AI projects and the fact that most firms do not yet have an established AI development department, most projects are led by AI developers.

Management should clearly invest more resources and effort in AI capability development for two reasons. Firstly, AI is developed from the ground up, which means that new capabilities will emerge from the AI team itself, implying that capable AI development teams have the opportunity and power to change outcomes in a positive and profitable way by implementing AI-driven projects that adhere to a responsible AI framework. Second, in order to drive future business value, managers must plan and invest ahead of competitors in order to remain competitive. Keeping ahead of the competition, on the other hand, does not happen overnight. It necessitates a systematic approach in which all managerial efforts are contained within a framework that clearly guides the necessary steps to achieve representative AI practices.

6.3 Limitations and Future Research

There are several limitations to this study's methodology. First, companies taking part reside in Scandinavia, where countries are known to have high standards for responsible and ethical practices, so it will be interesting to see how countries in different geographical regions tackle the same problem, such as North America or southern Europe. Our survey is limited in another way by the fact that we only captured a snapshot of what these companies do. Because we are not familiar with how they develop their AI products or make them better over time, we cannot identify how their practices change and what mechanisms they apply. Finally, we do not measure various performance metrics, such as social responsibility, reputation, or trust, which can affect the position of a firm in the market, since such measures can capture the value that an organization can obtain in the medium or long term.

References

1. Schmidt, R., Zimmermann, A., Möhring, M., Keller, B.: Value creation in connectionist artificial intelligence–a research agenda. In: Value Creation in Connectionist Artificial Intelligence–A Research Agenda (Americas Conference on Information Systems: AMCIS/Association for 2020), pp. 1–10 (2020)
2. Rodrigues, A.R.D., Ferreira, F.A., Teixeira, F.J., Zopounidis, C.: Artificial intelligence, digital transformation and cybersecurity in the banking sector: a multi-stakeholder cognition-driven framework. Res. Int. Bus. Finan **60**, 101616 (2022)
3. (2021). https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf. Accessed 16 Nov 2021
4. de Laat, P.B.: Companies committed to responsible AI: from principles towards implementation and regulation? Philos. Technol. **34**(4), 1135–1193 (2021). https://doi.org/10.1007/s13347-021-00474-3
5. Singapore-Government: model artificial intelligence governance framework. In: Model Artificial Intelligence Governance Framework (2021)
6. Cihon, P., Schuett, J., Baum, S.D.: Corporate governance of artificial intelligence in the public interest. Information **12**(7), 275 (2021)
7. Enholm, I.M., Papagiannidis, E., Mikalef, P., Krogstie, J.: Artificial intelligence and business value: a literature review. Inf. Syst. Front. 1–26 (2021)
8. Sandhawalia, B.S., Dalcher, D.: Developing knowledge management capabilities: a structured approach. J. Knowl. Manage. **15**(2), 313–328 (2011)

9. Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P., Krogstie, J.: Toward AI governance: identifying best practices and potential barriers and outcomes. Inf. Syst. Front. 1–19 (2022). https://doi.org/10.1007/s10796-022-10251-y
10. Yang, X., Yu, X., Liu, X.: Obtaining a sustainable competitive advantage from patent information: a patent analysis of the graphene industry. Sustainability **10**(12), 4800 (2018)
11. MacKenzie, S.B., Podsakoff, P.M., Podsakoff, N.P.: Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Q. **35**(2), 293–334 (2011)
12. Teece, D., Peteraf, M., Leih, S.: Dynamic capabilities and organizational agility: risk, uncertainty, and strategy in the innovation economy. Calif. Manage. Rev. **58**(4), 13–35 (2016)
13. Amershi, S., et al.: Software engineering for machine learning: a case study. In: Software Engineering for Machine Learning: A Case Study, pp. 291–300. IEEE (2019)
14. Wang, Y., Xiong, M., Olya, H.: Toward an understanding of responsible artificial intelligence practices. In: Toward an Understanding of Responsible Artificial Intelligence Practices (Hawaii International Conference on System Sciences (HICSS)), pp. 4962–4971 (2020)
15. Von Krogh, G., Nonaka, I., Aben, M.: Making the most of your company's knowledge: a strategic framework. Long Range Plan. **34**(4), 421–439 (2001)
16. Minkkinen, M., Zimmer, M.P., Mäntymäki, M.: Towards ecosystems for responsible AI. In: Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Pappas, I., Mäntymäki, M. (eds.) I3E 2021. LNCS, vol. 12896, pp. 220–232. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85447-8_20
17. Adam Cutler, A.W., Paka, A.: Staying ahead of the curve: the business case for responsible AI. In: Staying Ahead of the Curve: The Business Case for Responsible AI (2020)
18. Mikalef, P., Gupta, M.: Artificial intelligence capability: conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. Inf. Manag. **58**(3), 103434 (2021)
19. Sander, F., Semeijn, J., Mahr, D.: The acceptance of blockchain technology in meat traceability and transparency. Br. Food J. **120**(9), 2066–2079 (2018)
20. Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P., Krogstie, J.: Deploying AI governance practices: a revelatory case study. In: Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Pappas, I., Mäntymäki, M. (eds.) I3E 2021. LNCS, vol. 12896, pp. 208–219. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-85447-8_19
21. Tanriverdi, H.: Information technology relatedness, knowledge management capability, and performance on multibusiness firms. MIS Q. **29**(2), 311–334 (2005)
22. Mao, H., Liu, S., Zhang, J., Deng, Z.: Information technology resource, knowledge management capability, and competitive advantage: the moderating role of resource commitment. Int. J. Inf. Manage. **36**(6), 1062–1074 (2016)
23. Wu, S.P.-J., Straub, D.W., Liang, T.-P.: How information technology governance mechanisms and strategic alignment influence organizational performance. MIS Q. **39**(2), 497–518 (2015)
24. Tallon, P.P., Pinsonneault, A.: Competing perspectives on the link between strategic information technology alignment and organizational agility: insights from a mediation model. MIS Q. **35**(2), 463–486 (2011)
25. Preston, D.S., Karahanna, E.: Antecedents of IS strategic alignment: a nomological network. Inf. Syst. Res. **20**(2), 159–179 (2009)
26. Rai, A., Tang, X.: Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios. Inf. Syst. Res. **21**(3), 516–542 (2010)
27. Lwakatare, L.E., Raj, A., Bosch, J., Olsson, H.H., Crnkovic, I.: A taxonomy of software engineering challenges for machine learning systems: an empirical investigation. In: Kruchten, P., Fraser, S., Coallier, F. (eds.) XP 2019. Lecture Notes in Business Information Processing, vol. 355, pp. 227–243. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19034-7_14

28. Dignum, V.: Responsibility and artificial intelligence. Oxford Handb. Ethics AI **4698**, 215 (2020)
29. Tseng, S.-M., Lee, P.-S.: The effect of knowledge management capability and dynamic capability on organizational performance. J. Enterp. Inf. Manag. **27**(2), 158–179 (2014)
30. Avison, D., Jones, J., Powell, P., Wilson, D.: Using and validating the strategic alignment model. J. Strateg. Inf. Syst. **13**(3), 223–246 (2004)
31. Schwab, K.: The global competitiveness report 2019. In: World Economic Forum, vol. 9 (2019)
32. Schwab, K., Zahidi, S.: The global competitiveness report special edition 2020: how countries are performing on the road to recovery. In: World Economic Forum (2020)
33. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3, Boenningstedt: SmartPLS GmbH (2015)
34. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS 3. SmartPLS GmbH, Boenningstedt. J. Ser. Sci. Manag. **10**(3), 32–49 (2015)
35. Rana, N.P., Chatterjee, S., Dwivedi, Y.K., Akter, S.: Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. Eur. J. Inf. Syst. **31**(3), 1–24 (2021)
36. Rakova, B., Yang, J., Cramer, H., Chowdhury, R.: Where responsible AI meets reality: practitioner perspectives on enablers for shifting organizational practices. Proc. ACM Hum.-Comput. Interact. **5**(CSCW1), 1–23 (2021)
37. Duan, Y., Edwards, J.S., Dwivedi, Y.K.: Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda. Int. J. Inf. Manage. **48**, 63–71 (2019)

PAPER 6

Exploring the link between Responsible AI Governance, Legitimacy, and Firm Performance- An Empirical Examination

Emmanouil Papagiannidis, Patrick Mikalef

Exploring the link between Responsible AI Governance, Legitimacy, and Firm Performance- An Empirical Examination

Abstract

In contemporary business practices, the utilization of Artificial Intelligence (AI) has become increasingly prevalent. However, due to the proliferation of AI technologies and the occurrence of unsuccessful AI product deployments in recent times, the notion of Responsible AI Governance (RAIG) has emerged since the potential reputational damage associated with AI implementation poses a significant financial risk to organizations. Those firms who encounter difficulties in implementing RAIG principles may experience a loss of legitimacy in the eyes of the general public and stakeholders, particularly if they are unable to effectively communicate their compliance with relevant laws and regulations pertaining to AI usage. In response to these challenges, this research study synthesizes existing literature on RAIG, internal and external legitimacy, and aims to develop a conceptual model that explores the relationship among RAIG, effective communication practices regarding RAI use, legitimacy, and firm performance. The proposed research model is empirically examined using partial least squares structural equation modeling, leveraging survey data collected from a sample of 329 employees working in companies across Western Europe and the USA. The findings of the study provide empirical evidence supporting the influence of robust RAIG practices on organizational legitimacy and subsequent firm performance. This study contributes to the field by being the first to present empirical insights into the interplay between RAIG, legitimacy, and its implications for firm performance. The results shed light on the importance of adopting RAIG strategies and effective communication practices in maintaining organizational legitimacy and achieving positive business outcomes.

Keywords: Responsible AI Governance, Internal Legitimacy, External Legitimacy, Firm Performance, RAI Communication, Corporate Reputation, Survey, Quantitative Analysis

1. Introduction

In the past, corporate performance was primarily assessed based on the concept of traditional profit maximization (Zhang, 2015). Early economics literature used this measure to evaluate economic performance, considering companies efficient if they were focused on maximizing profits and inefficient if they were not profit-oriented (Riahi-Belkaoui, 1992). Profit maximization was regarded as a significant model for evaluating corporate performance (Terrien et al., 2017). However, the concept of firm performance has been redefined many times as new technologies and challenges emerged. Legitimacy theory challenges this perspective by proposing that profit should be seen as an all-encompassing measure of organizational legitimacy (Archel et al., 2009). According to this theory, while profits are important, businesses should also prioritize the preservation of goodwill within the community and maintain their legitimacy as responsible participants in society by implementing ethical and socially responsible practices (Chan et al., 2014). Legitimacy theory emphasizes that organizations should consider the rights of the public as a whole, not just the rights of investors (Beyers & Arras, 2021). For companies to exist and thrive in a community or society, they must acknowledge that their legitimacy stems from the approval of that society (Levitov, 2015). Consequently, they have an obligation to fulfill the needs and expectations of the society in which they operate.

The literature (Colleoni, 2013; Dai et al., 2018; Francés-Gómez, 2020; Zheng et al., 2015) focuses on the connection between RAIG and legitimacy theory, highlighting how they are related and what it means for organizations. We define responsible AI governance as:

> *A set of practices that documents the involved process of developing, applying and monitoring AI applications and products while addressing all challenges that surround AI with a set of rules and authorities for (1) managing the appropriate functionality of AI, (2) assuring the trustworthiness of AI, and (3) overseeing the whole life cycle of data and algorithms within and between organizations and firms (Papagiannidis Emmanouil et al., 2023).*

The foundation of RAIG is the ethical and responsible use of AI technology, which includes values and procedures that guarantee fairness, accountability, transparency responsibility, and social well-being (de Laat, 2021). By upholding societal norms, values, and expectations, organizations are able to gain and maintain legitimacy in the eyes of stakeholders (Mikalef et al., 2022). Several studies (Burlea & Popa, 2013; Dai et al., 2018; Mark Van Rijmenam & Schweitzer, 2018; Patten, 2020; Rodrigues, 2020; Suddaby et al., 2017) stress the critical role that ethical AI governance plays in enhancing corporate legitimacy. Organizations may show their dedication to social interests and conform to stakeholder expectations by implementing ethical AI practices and open decision-making procedures. This strengthens their credibility as accountable corporate players. Additionally, enterprises may control the social, ethical, and legal concerns connected to the use of AI through responsible AI governance (Enholm, 2021). Organizations may address issues with biases, privacy, and algorithmic transparency by implementing ethical AI practices, potentially minimizing harm and eliminating negative externalities (Papagiannidis, Enholm, et al., 2022). By ensuring that AI technologies are employed in a way that is consistent with social norms and expectations, such proactive efforts will support and strengthen the legitimacy of the business. Furthermore, research emphasizes the significance of communication and stakeholder participation in RAIG and legitimacy (Beyers & Arras, 2021; Colleoni, 2013; Khuong et al., 2021). Effective communication about AI practices, such as public disclosure of AI algorithms, data utilization, and decision-making processes, builds stakeholder trust and understanding. This message allows firms to demonstrate their commitment to RAIG, bolstering their credibility (Mikalef et al., 2022). Overall, the research indicates that RAIG is critical to organizational legitimacy. Organizations may also boost their legitimacy, win stakeholder confidence, and portray themselves as responsible AI-driven entities in the eyes of the public and key stakeholders by controlling risks, and engaging stakeholders through effective communication.

There are still specific gaps, nevertheless, that require more research and inquiry. More specifically, despite the fact that there is a growing body of literature that discusses the significance of RAIG for organizational legitimacy, empirical studies that study and establish this link are still few. The precise mechanisms by which RAIG practices affect organizations' legitimacy and how this affects diverse organizational outcomes require further empirical investigation. Empirically examining them is critical to developing a comprehensive understanding of the role played by RAIG in fostering legitimacy and subsequent firms' performance. An assessment of communication practice would inform firms of the benefits of communicating RAIG. Consequently, this paper seeks to answer the following research questions:

RQ1

What is the effect of RAIG on legitimacy and, subsequently, on firm performance?

RQ2

What is the impact of communicating RAIG practices on firm performance?

To address this research gap, a conceptual model was developed, and an instrument of RAIG was constructed based on the guidelines provided by MacKenzie et al. (2011). This study aims to explain how RAIG practices, aligned with the utilization of responsible AI (RAI) communication, contribute to increased legitimacy, thereby leading to competitive performance gains. Accordingly, this study aims to investigate whether RAIG enhances organizational legitimacy, the extent to which RAI communication moderates the relationship between RAIG and firm performance, and whether legitimacy confers significant competitive advantages.

The remainder of this paper is organized as follows. The subsequent section provides the background of the study, presenting an overview of legitimacy and RAIG. In Section 3, we articulate our hypotheses, wherein we examine the impact of robust RAIG on competitive performance. Section 4 delineates the methodology employed in this study, including the survey-based design, data collection procedures, and measurement approaches for each concept under investigation. Subsequently, we present the results derived from our empirical analysis in Section 5, followed by a comprehensive discussion of their theoretical and practical implications in Section 6, along with a discussion of the notable limitations encountered in this research endeavor.

2. Theoretical Background

2.1 Legitimacy Theory

According to Suchman (1995) "Legitimacy is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions." Burlea and Popa (2013) add to this, suggesting that legitimacy theory serves the purpose of explaining how organizations behave when it comes to adopting and advancing voluntary disclosure of social and environmental information. This behavior is driven by their desire to fulfill the social contract they have with society, which allows them to achieve their goals and remain viable in an unpredictable and tumultuous environment (Fisher et al., 2016). That means that legitimacy could be seen as a process, focusing on the processual aspects that lead to the emergence of legitimacy (Suddaby et al., 2017; Thomas & Ritala, 2022). In society, there exist established norms and shared perceptions of what constitutes good or bad behavior. When a company takes an action, it can be evaluated as either positive or negative based on these societal standards (Fisher et al., 2017). Suppose the company's actions align with and are considered favorable according to the prevailing societal norms. In that case, it is seen as a diligent and commendable organization (Nagy et al., 2017). Essentially, the legitimacy of an organization is determined by how well it conforms to the standards set by society. Hence, legitimacy establishes a connection between the actions of an entity or a business and what is considered acceptable or permissible within society, where society evaluates these actions based on societal standards and determines whether they are desirable, appropriate, or morally right (Taeuscher et al., 2021).

Just like the social contract theory, legitimacy theory is founded on the idea that a social contract exists between society and an organization (Francés-Gómez, 2020; Mäkelä & Näsi, 2010). This contract entails mutual benefits for both parties involved. On the one hand, the organization gains profits and promotes its own interests, while on the other hand, it produces goods or services that cater to customer needs, generate employment opportunities, and contribute to overall prosperity in the community or region (Lee, 2011). There are two sides at play, and both stand to gain from this social contract (Fisher, 2020). A firm receives permission from society to operate and it is ultimately accountable to the society for what it does and how it operates because society provides corporations with the authority to own and use natural resources, and to hire employees (Bes et al., 2019).

2.2. Importance of legitimacy theory

The importance of legitimacy theory can be recognized when the organization's actions are reported in line with societal expectations and perceptions (Deegan, 2002). If the organization's activities fail to align with social and moral values, it faces significant repercussions and sanctions from society, which could potentially result in its downfall (Haniffa & Cooke, 2005). The organization must establish its legitimacy through economically and socially responsible actions that safeguard the well-being of both the society it operates in and the natural environment (Anh & Velencei, 2019). The reputation of the organization is shaped throughout its whole life cycle, and legitimacy cannot be given to any institution as a state or situation.

Legitimacy is completely based on how trustworthy and moral the organization is (Koh et al., 2023). Stakeholders apply internal and external pressure to shape the institutionalized environment in which the firm functions. Although legitimacy theory is used to encourage voluntary disclosure of social and environmental standards, it should not be thought of as a panacea for all of a corporation's social and environmental issues. Its goal is to show stakeholders that the organization's actions are moral and consistent with established standards and principles (Chan et al., 2014). In fact, stakeholders are thought to need a sense of legitimacy before they tolerate corporate behavior (Yim, 2021). For example, it is not difficult to identify the motivations behind corporations revealing environmental information in the media and corporate annual reports, as these motivations are connected with the organization's life cycle, reputation, and pressures from internal and external stakeholders (Sharif & Rashid, 2014). The fundamental issue is to connect the actions, norms, values, and culture of the business with those of the society in which it functions (Patten, 2020).

2.2. The role of legitimacy in the age of RAIG

RAIG is key to legitimacy. Failure to comply with societal norms can result in a variety of consequences, including restrictions on the organization's operations, access to resources, and product demand. When a firm fails to follow societal norms due to the lack of RAIG, its legitimacy suffers, and it loses the community's respect and support (Dai et al., 2018). As a result, it attracts criticism and scrutiny, with society requesting that local or national governments establish controls and limitations on its operations. Society may turn against an organization that has not included ethics in its AI practices, considering it as illegitimate and operating against society's larger interests (Dellmuth & Schlipphak, 2020). Customers may boycott the company's products if they believe its conduct to be ethically reprehensible, thus diminishing its legitimacy. Hence, misalignment with responsible principles, for instance, not having explainable AI or not being accountable for AI actions, may intersect with the overall corporate governance and public opinion, as the company should provide disclosures regarding community concerns, fears, or issues that affect the well-being of the community (Stupak et al., 2021). Failure to address these aspects can result in the loss of legitimacy and the company's exclusion from the community, meaning that if there is a divergence between what the organization wants or does and what the society expects, the organization will lose legitimacy and encounter difficulties within that society (Giacomini et al., 2021).

According to the literature (Al-Abrrow et al., 2022; Islam et al., 2021; Khuong et al., 2021; Patten, 2020; Silva, 2021; Stupak et al., 2021; Tian & Tian, 2022), an organization's sustainability is dependent on its legitimation procedures and its ability to successfully handling of continuing pressures and difficulties. The major purpose of responsible AI processes and practices is to get and maintain consent from stakeholders. Legitimacy is viewed as a necessary prerequisite for stakeholders to accept corporate activity (Beyers & Arras, 2021). The incentives for corporations revealing environmental information in the media and corporate annual reports are reasonably straightforward to determine, as these motivations are connected with the organization's life cycle, reputation, and the pressures applied by both external and internal stakeholders (Tian & Tian, 2022). Given that legitimacy is the pinnacle of societal approval, we may infer that it becomes a vital goal for the organization (Silva, 2021). Therefore, failing to meet this goal, i.e. applying RAIG practices, reflects poorly on the company as a whole.

3. Research Model

The research model is presented in Figure 1, along with the hypotheses surrounding it. Our contention is that RAIG will impact the overall credibility of a company. Additionally, we argue that effective communication regarding AI and technology will amplify the influence of RAIG on the company's performance. Therefore, implementing a sound AI governance model that appropriately communicates the use of AI systems to the public and internal stakeholders will enhance the company's credibility, leading to improved performance. Legitimacy is a crucial aspect of any organization, multinational or otherwise, as it serves as a valuable resource for acquiring other resources (Zimmerman & Zeitz, 2002). Like all organizations, firms require resources and social support from their environment in order to thrive and survive (Crossley et al., 2021). For instance, the external legitimacy of a foreign subsidiary refers to its acceptance and approval by institutions in the host country, while internal legitimacy pertains to its acceptance and approval by the parent company and other subunits (Kostova & Zaheer, 1999; Thomas & Ritala, 2022). Additionally, external actors such as regulators, media, influencers, and analysts confer legitimacy by acknowledging the existence of the ecosystem (King et al., 2007).

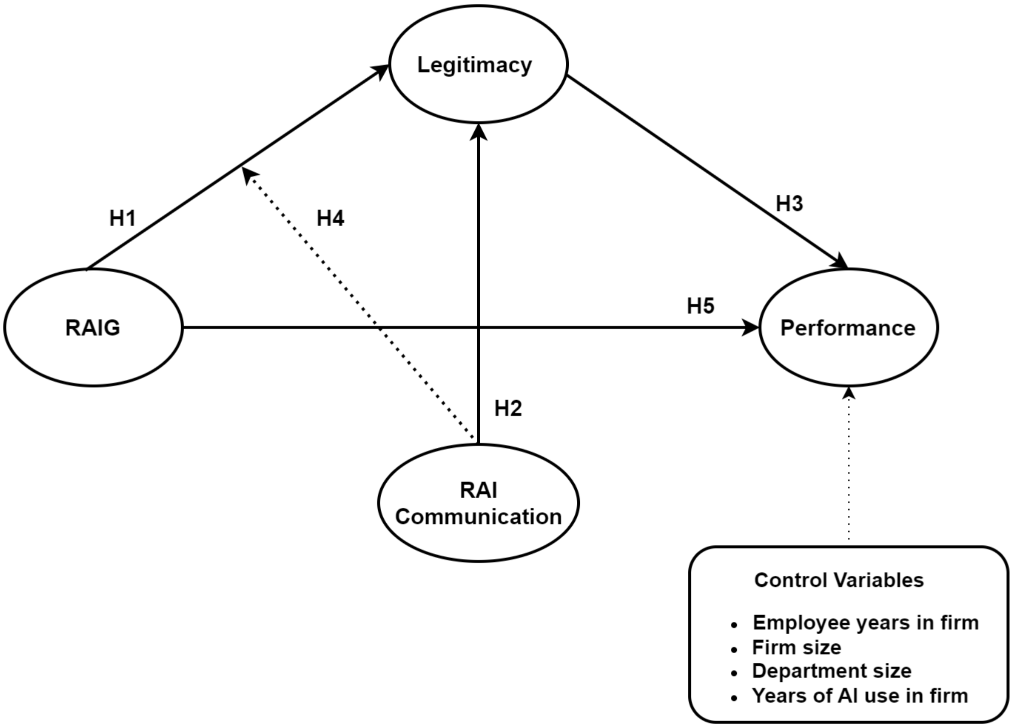

Figure 1. Research model.

RAIG is conceptualized as a higher-order construct, with each dimension comprising more than one sub-dimension. In the Table 1, we provide definitions for each construct.

Table 1. Constructs and definitions.

Construct	Definition	Sources
Responsible AI Governance	Refers to the set of policies that outlines the various approaches to ethically governing AI, encompassing every phase of AI projects' lifecycles, while adhering to the principles of responsible utilization	(Amershi et al., 2019; de Laat, 2021; Papagiannidis, Mikalef, et al., 2022; Singapore Government, 2020)
RAI Communication Use	Refers to methods and mechanisms through which AI systems communicate information internally or to the public regarding the ethical, legal, and societal consequences	(Guttal, 2023; Laato et al., 2022; Steyn, 2004)

	that emerge from the integration of AI within human communication contexts.	
External Legitimacy	Refers to the perception of how society/public and other external entities perceive an organization based on their ethical and societal expectations, rather than solely on the organization's actual performance.	(Caspersen, 2015; Drori & Honig, 2013; Levitov, 2015; Zhang et al., 2023; Zubek & Góra, 2021)
Internal Legitimacy	Refers to the acceptance and approval of internal stakeholders with the organization's domestic governance.	(Breit, 2014; Gammelgaard & Kumar, 2023; Kostova & Zaheer, 1999; Lu & Xu, 2006; Park et al., 2012)
Firm Performance	Refers to the extent to which a company outperforms its rivals in terms of its environmental practices, financial results, competitive strength, and overall corporate image.	(Khan et al., 2020; Kristoffersen et al., 2021; Rai et al., 2006)

3.1. Hypotheses

The adoption of AI technologies by organizations has witnessed a steady increase, and the manner in which they handle and implement responsible AI practices holds significant influence over their internal legitimacy. When organizations embrace ethical frameworks to govern AI, employees perceive them as more legitimate (Carter, 2018). These frameworks establish guidelines that promote fairness, transparency, accountability, and privacy, ensuring that organizational practices align with ethical standards (Du & Xie, 2021). Moreover, RAIG practices assume a crucial role in establishing and maintaining trust between corporations and their stakeholders (Morsing & Schultz, 2006). Transparent AI systems, which explain the decision-making process, enable stakeholders to comprehend and assess the ethical implications associated with the technology. By ensuring transparency, organizations can demonstrate their dedication to responsible AI practices, thereby augmenting their reputation as reliable and trustworthy entities (Sharma, 2022). For instance, it should be noted that the fairness of AI algorithms is contingent upon the quality and impartiality of the data upon which they are trained. Neglecting to address biases and discrimination can result in adverse consequences and harm the reputation and loyalty of a company (Ucbasaran et al., 2013). However, when employees recognize that an organization prioritizes RAIG, it strengthens their trust in both the technology itself and the overall legitimacy of the organization.

The world prefers controls on AI to ensure responsible management and protection of the interests of organizations, stakeholders, and society as a whole. While ethical frameworks provide some guidance, a more effective approach involves applying adapted forms of risk assessment and risk management techniques (Clarke, 2019). It is in the best interest of all organizations to avoid causing harm to their stakeholders, as this can lead to a loss of trust in AI and opposition to its use. To achieve this, organizations need to adopt responsible approaches to AI that boost its legitimacy (Colleoni, 2013). This requires fostering an appropriate organizational culture and establishing business processes dedicated to identifying and managing risks (Clarke, 2019). Companies that employ responsible practices are better positioned to justify the use of emerging technologies and adapt their operations accordingly (Malsch, 2013). Risk assessment in the context of RAIG helps organizations expand their social responsibility by implementing previously unavailable practices and methods. According to Cowls et al. (2019), organizations that effectively manage risks associated with AI enhance their internal legitimacy by creating a sense of security and minimizing negative consequences. By taking proactive steps to address issues such as bias, discrimination, and privacy concerns, these organizations demonstrate their dedication to responsible AI practices, thereby strengthening their internal legitimacy. Another research by Jobin et al. (2019) highlights that the adoption of RAIG practices helps organizations align with emerging regulations and guidelines on AI ethics. Organizations that uphold ethical and legal norms are perceived as more legitimate and trustworthy. Hence, by proactively implementing RAIG measures, organizations show their commitment to meeting regulatory requirements, which further enhances their internal legitimacy.

The RAIG practices can encompass safeguarding user privacy and data security, as the protection of user data privacy rights not only fosters customer trust but shields businesses from data breaches and the subsequent damage to their reputation too. At the same time, the reputation of corporations is increasingly entangled with their social and environmental impact (Babiak & Trendafilova, 2011). By leveraging AI to tackle societal challenges, companies can bolster their reputation by showcasing their commitment to making a positive impact on the world. For example, AI-driven solutions that contribute to sustainability, social equality, or healthcare accessibility not only enhance reputation but also attract conscientious consumers and investors. However, despite the prevalence of ethical issues surrounding AI, most companies lack effective preparedness to address public concerns (Wen & Holweg, 2023). Cases such as Amazon Rekognition and IBM Watson (facial recognition), where both products exhibited biases based on skin color and gender, illustrate this point (Wen & Holweg, 2023). Companies need to integrate socially responsible activities with their core business strategies and consider initiatives focused on stakeholders. These actions can result in enhanced customer satisfaction, which in turn can boost corporate reputation (Sharma, 2022). Hence, trust forms the foundation of corporate reputation, consistently emerging as a key attribute when measuring brand equity, as any interaction between a brand and its customers entails an exchange of value.

Although the concept of CSR is grounded in ethics, which can encompass environmental aspects (Carroll, 1999), it can extend beyond an organization’s direct interests to include philanthropic contributions to individuals, communities, society, or the environment. Its primary focus in practice often revolves around leveraging strategic advantages or enhancing public relations through regulatory compliance and philanthropic activities (Porter & Kramer, 2006). Recognizing the potential of responsible principles in addressing critical societal challenges, an increasing number of studies have highlighted their positive correlation with legitimacy and corporate reputation. Research delves into CSR antecedents and practices that influence credibility and sustainability (Abdullah & Abdul Aziz, 2013; Al-Abrrow et al., 2022; Goldsmith et al., 2000; Maas & Liket, 2011; Zheng et al., 2015). They emphasize the profound relationship where strategic decisions can shape internal and external acceptance and impact an organization's overall legitimacy. Therefore, RAIG is not merely a regulatory requirement but a strategic imperative for corporations committed to long-term success and ethical innovation. From the preceding discussion, it is hypothesized that:

H1 - RAIG has a positive effect on legitimacy, both internal and external.

Effective AI technology communication enables firms to build external stakeholder transparency, trust, and understanding too (Wang et al., 2020). Organizations can address the worries and skepticism of stakeholders, including consumers, investors, and the general public, by providing accurate and understandable information regarding the responsible application of AI technologies and the governance systems in place, typically through corporate websites (Hickman & Petrin, 2021). This openness makes it easier for stakeholders to grasp the organization's dedication to ethical AI practices and comprehend the issues at hand (Schultz & Seele, 2023). Additionally, communication about AI technology enables firms to harmonize their values and objectives with the ethical application of AI, improving external legitimacy. What is more, it helps an organization's reputation and legitimacy in the eyes of clients, investors, and the general public when external stakeholders view it as open, dependable, and ethical in its AI operations (Khuong et al., 2021); thus effective communication enables businesses to address any potential risks, biases, or ethical issues related to AI in a proactive manner, fostering confidence and trust among external stakeholders (Bedué & Fritzsche, 2022).

Through framing and recognizing, communication helps external legitimacy. In order to organize experience and direct behaviour, framing entails the formation of meaning that draws attention to a few key conspicuous elements (Battilana et al., 2009; Benford & Snow, 2000; Cornelissen & Werner, 2014; Thomas & Ritala, 2022). Ecosystem orchestrators engage in framing to influence how an ecosystem is described (Lindgren et al., 2015; Snihur et al., 2018). External recognition is especially important for the legitimation of an emerging ecosystem because it makes legitimation possible (Drori & Honig, 2013; Thomas & Ritala, 2022). The success of a social movement or institutional entrepreneur can be influenced by the discursive acts of the media, analysts, regulators, and other actors in society (Lounsbury & Crumley, 2007), and these actors can also influence the establishment of legitimacy. An organization, for instance, will be perceived as genuine if it transacts business with other reputable, well-known companies (Snihur et al., 2018). From the preceding discussion, it is hypothesized that:

H2: RAI communication use has a positive effect on legitimacy.

Legitimacy plays a vital role in shaping the success and performance of organizations. Legitimacy is closely linked to stakeholder trust, which is crucial for the long-term success of any organization (Du et al., 2022). More specifically, internal legitimacy plays a critical role in firm performance. When an organization's members accept and support a firm, perceiving it as legitimate and credible, there are positive effects on performance (Suchman, 1995). Moreover, employees who view their organization as legitimate are more likely to demonstrate commitment, motivation, and engagement, ultimately leading to improved performance. Organizations that prioritize ethical practices, transparency, and legal compliance are better equipped to mitigate risks and avoid legal and reputational damage. By adhering to applicable laws, regulations, and industry standards, businesses can avoid penalties, litigation, and loss of public trust. Bitektine (2011), emphasizes the connection between internal legitimacy and organizational outcomes, including financial performance, innovation, and overall effectiveness. Employees who believe in their organization's legitimacy are more inclined to be motivated, loyal, and proud of their association with the company, ultimately contributing to enhanced organizational performance (Blanco-Gonzalez et al., 2020; Min et al., 2023). Legitimacy is also a crucial factor in attracting and retaining talented employees. Today's workforce seeks purpose and meaning in their work (Gabrielova & Buchko, 2021). Organizations that prioritize corporate legitimacy are better positioned to attract top talent, foster a positive work culture, and boost employee engagement and productivity. Ultimately, the pursuit of corporate legitimacy is not just a moral imperative; it is also a strategic advantage that drives organizational performance and sustainability.

When an organization demonstrates a commitment to ethical conduct and social responsibility, it enhances trust and confidence among stakeholders (Swift, 2001). This trust translates into positive perceptions, increased customer loyalty, higher employee morale, and improved relationships with investors (Papasolomou-Doukakis et al., 2005). Furthermore, in an era of heightened consumer awareness and social media influence, legitimacy significantly impacts brand reputation and customer loyalty (Islam et al., 2021). Consumers are increasingly inclined to support businesses that align with their values and exhibit responsible behaviour. Organizations with strong corporate legitimacy have a greater likelihood of building a positive brand reputation, which can translate into a competitive advantage. Customers perceive these organizations as trustworthy and are more likely to choose their products or services over those of competitors (Harrison et al., 2010). In addition, satisfied customers who identify with a company's values and ethical practices tend to develop strong brand loyalty, leading to repeat purchases and positive word-of-mouth recommendations (Tuškej et al., 2013). Lastly, organizations that proactively integrate responsible practices, like green innovation, into their operations are better prepared to anticipate and respond to emerging societal and regulatory expectations, thereby reducing the likelihood of disruptions and enhancing long-term performance (Yuan & Cao, 2022). From the preceding discussion, it is hypothesized that:

H3: Legitimacy has a positive effect on firm performance.

Effective communication about AI technology is essential in order to moderate the relationship between RAIG and internal legitimacy inside businesses. It primarily enables firms to promote openness, understanding, and trust among their workforce. Organizations can address any potential worries or misunderstandings by disseminating accurate and understandable information on the appropriate use of AI technologies and the governing frameworks in place (ÓhÉigeartaigh et al., 2020). Employees can comprehend the justification for AI-related actions and the ethical considerations involved thanks to this transparency (Ehsan et al., 2021). Additionally, communication about AI technology makes it easier to integrate company goals and values with ethical AI use, which improves internal legitimacy (Men, 2014). Employees are more likely to believe that an organization is trustworthy and ethical when they are aware of and understand its commitment to ethical AI practices, which strengthens the organization's internal legitimacy (Shneiderman, 2020). Additionally, good communication enables businesses to address employee worries, misunderstandings, and concerns about AI, which lowers resistance and promotes acceptance. The improved moderation of the relationship between RAIG and internal legitimacy is a result of the increased awareness of and support for AI technology. In the end, businesses can use AI-technology communication as a powerful tool to build and maintain trust, legitimacy, and openness when deploying AI systems. From the preceding discussion, it is hypothesized that:

H4: RAI communication use has a positive moderating effect on the relationship between RAIG and legitimacy.

RAIG plays a significant role when it comes to ethical principles and practices, such as transparency, and accountability, in order to successfully integrate AI within organizations (Felzmann et al., 2020). By implementing responsible AI governance practices, firms can ensure that AI technologies are developed, deployed, and managed in a manner that aligns with legal regulations, societal expectations, and corporate values (Rakova et al., 2021). This not only mitigates the risks associated with unethical or biased AI applications but also boosts customer trust, stakeholder confidence, and brand reputation. To be more specific, RAIG encourages innovation while minimizing potential negative impacts, thereby optimizing operational efficiency and driving sustainable long-term growth (Kulkov et al., 2023). A commitment to responsible AI governance is not only about ethics but also about planning ahead, contributing to firm performance in an increasingly AI-driven environment; thus, RIAG is essential for ensuring that AI technologies are deployed ethically and transparently within organizations (Di Vaio et al., 2020).

This approach of implementing RAIG practices creates a positive relationship with firm performance in several ways. Firstly, it instills trust and confidence among customers, stakeholders, and regulators by demonstrating a commitment to ethical practices and compliance with relevant regulations. This, in turn, enhances the organization's reputation and reduces the likelihood of reputational damage due to AI-related controversies (Vlaeminck, 2023). Secondly, responsible AI governance promotes innovation by providing clear guidelines and standards for the development and implementation of AI solutions. By encouraging ethical experimentation and responsible risk-taking, firms can leverage AI to drive competitive advantage and unlock new business opportunities. Additionally, by prioritizing fairness, accountability, and transparency in AI decision-making processes, firms can minimize the potential for discrimination and ensure that AI systems align with corporate values and societal expectations (Khakurel et al., 2018). Therefore, a proactive approach to RAIG not only mitigates risks but also maximizes the positive impact of AI on firm performance, driving sustainable growth and creating value for all stakeholders. Thirdly, through new products and innovation, the return on investment may be increased, creating new sources of revenue (Babina et al., 2021). By leveraging AI technologies ethically and transparently, firms can introduce groundbreaking solutions that address emerging market needs and surpass competitors' offerings (Wamba-Taguimdje et al., 2020). This innovation-driven approach not only attracts new customers but also creates loyalty among existing ones, leading to sustained sales growth. With a diverse portfolio of intelligent products and services, the organization becomes more resilient to market fluctuations, contributing to financial stability and long-term success (Pwc, 2019). This financial stability not only instills confidence among investors and stakeholders but also provides the necessary foundation for strategic expansion into new markets and opportunities. Hence, RAIG not only drives innovation but also fortifies the organization's financial position, making the way for sustainable growth and prosperity. From the preceding discussion, it is hypothesized that:

H5: RAIG has a direct positive effect on firm performance.

4. Empirical study

In this work, we conducted a quantitative study to examine the suggested research paradigm. We gathered uniform data from a diverse set of pivotal respondents, employing a survey methodology, including data scientists, managers and CEOs. This approach allows diverse responses, enabling broader conclusions to be drawn. The study focused on Western Europe and the USA because these regions lead in frameworks for RAIG and have common ethical considerations and practices when it comes to fairness, transparency, and accountability in AI systems. Another reason for our sample choice is that many AI companies are based in these regions and their approaches to RAIG can significantly impact global practices.

Prior to evaluating the impacts of the suggested proposed research model, we conducted a statistical power analysis to ascertain the necessary minimum sample size (Kock & Hadaya, 2018). Based on the inverse square root method to determine the minimum required sample size, our model exhibited no bias in estimators (Kock & Hadaya, 2018). With a sample size of 329, our study possesses sufficient statistical power to assess the statistical significance of the proposed research model concerning path coefficients of at least 0.2, at a significance level of 0.05.

We invested considerable time in the data collection process to uphold reliability and validity by mitigating common method variance (Maier et al., 2023). Initially, measures were taken to ensure respondent anonymity, maintain the random presentation of variables and items and restrict respondents from revisiting survey sections (Podsakoff et al., 2003). Additionally, the survey was administered by a company with a well-developed protocol. We utilized 87 questions to measure our items and we used a 7-point Likert scale, where a value of 1 means disagrees entirely, and 7 means agree entirely.

4.1. Sample characteristics

To put the study model to the test, we sent an internet questionnaire-based survey using a panel service. The survey was administered across six countries, namely France, Germany, Italy, the Netherlands, Spain, and the USA, with data collected as follows: 39 responses (12%) from France, 39 responses (12%) from Germany, 38 responses (11%) from Italy, 11 responses (4%) from the Netherlands, 44 responses (13%) from Spain, and 158 responses (48%) from the USA. Respondents represented a diverse range of company sizes, with 56 (17%) reporting employment in companies with 1000 to 2500 employees, 20 (6%) not specifying company size, 54 (16%) in companies with 250 to 500 employees, 58 (18%) in companies with over 2500 employees, 62 (19%) in companies with 50 to 250 employees, and 79 (24%) in companies with 500 to 1000 employees.

The survey revealed that the majority of respondents reported the integration of one or more AI applications within their respective organizations. Specifically, the most commonly reported applications were as follows: 250 respondents (75%) reported utilizing chatbots, 146 (44%) reported virtual agents, 140 (42%) reported real-time translation for meetings, 202 (62%) reported cybersecurity applications, 142 (43%) reported using AI for decision management, 105 (32%) reported robotic process automation applications, 163 (49%) reported speech analytics applications, 128 (39%) reported AI expert systems, 165 (50%) reported planning scheduling and optimization techniques applications, 89 (27%) reported machine vision applications, and 201 (60%) reported other types of machine learning applications.

Furthermore, a significant portion of respondents indicated the duration of AI usage within their organizations, with 194 (59%) reporting usage for 1-2 years, 99 (30%) for 3-4 years, and only 36 (11%) for less than one year. Additionally, the majority of participants reported substantial investments in AI technologies by their companies. Regarding respondents' roles within their organizations, 56 (17%) identified as IT directors, 49 (15%) as business managers, 74 (22%) as chief officers, 40 (12%) as IT project managers, 30 (9%) as operation managers, and 80 (25%) with other occupational roles.

4.2. Conceptualization and measurements of constructs

In line with the RAIG definition that we provided earlier, the RAIG construct is conceptualized as a multidimensional third-order formative construct, which is comprised of the following AI-specific dimensions: transparency, accountability, fairness, environmental and societal well-being, technical robustness and reproducibility, data governance, and human agency and oversight. These dimensions are, in turn, conceptualized as second-order formative constructs comprising eighteen first-order constructs (Table 2). As for legitimacy and RAI communications, these are formative second-order constructs, and each one has two first-order reflective constructs. These are external and internal legitimacy for legitimacy and explicit and implicit communications for RAI communications (Table 2).

Table 2. Latent constructs and sub-dimensions.

Third-order	**Type**	**Second-order (sub-dimensions)**	**First-order (sub-dimensions)**	**Type**
RAIG	Formative	Human agency and oversight	Human agency	Reflective
			Human oversight	Reflective
		Accountability	Auditability	Reflective
			Minimization and reporting negative impacts	Reflective
		Transparency	Communication	Reflective
			Traceability	Reflective
			Explainability	Reflective
		Fairness	Avoidance of unfair bias	Reflective
			Accessibility and universal design	Reflective
		Environmental and societal well-being	Social impact	Reflective
			Sustainable and environmental friendly AI	Reflective
		Technical robustness and safety	Accuracy	Reflective
			Reliability and reproducibility	Reflective
			Fallback plan and general safety	Reflective
			Resilience to attack and security	Reflective
		Data governance	Quality and data protection	Reflective
			Privacy and data governance	Reflective
			Access to data	Reflective
		Legitimacy	External legitimacy	Reflective
			Internal legitimacy	Reflective
		RAI Communications	Explicit communications	Reflective
			Implicit communications	Reflective

Performance	Customer relationship	Reflective
	Finances	Reflective
	Competition	Reflective

In the process of constructing the first-order constructs for RAIG, the measures employed were derived from existing literature or expert opinions on RAIG. This involved an examination and analysis of similar conceptual models related to the governance of AI systems. By drawing upon this extensive body of literature, the chosen measures in the development of the first-order constructs were purposefully selected and adapted to ensure their relevance, validity, and applicability within the specific context of the study. This approach not only enabled the identification of suitable dimensions and variables for RAIG but also enabled the establishment of a theoretical foundation upon which subsequent empirical investigations could be grounded. Furthermore, by using existing literature, the measures have theoretical rigor and methodological robustness, thereby ensuring the overall credibility and trustworthiness of the research findings.

Similarly, we derived existing literature for legitimacy and RAI communications (Burlea & Popa, 2013; Colleoni, 2013; Kostova & Zaheer, 1999; Steyn, 2004). When RAIG practices are in place, firms are able to increase their legitimacy and through communication, they can boost even more their internal and external acceptance from stakeholders and society (Morsing & Schultz, 2006). This idea is reflected in the proposed theoretical framework (Fig. 1) and in the items used to capture the first-order constructs, which are related to the AI-given value within organizations.

Firm performance is shaped by multiple interconnected factors, among which competition, customer relationships, and firm finances play crucial roles (Bergek et al., 2013; Lei & Slocum Jr, 2005; Naradda Gamage et al., 2020). In a highly competitive environment, firms must be able to challenge their rivals, handle technological disruptions, and adapt to consumer preferences to secure their survivability and drive growth (Rivard et al., 2006). The ability to effectively rival your competition, whether through product differentiation, cost leadership, or market segmentation, significantly influences a firm's market share, profitability, and long-term viability (Santos-Vijande et al., 2012). Moreover, the quality of customer relationships is a necessary step for sustained success, as customer loyalty not only contributes to recurring revenues but also serve as brand awareness, leading to organic growth and mitigating the impact of market fluctuations (Liu et al., 2022). Beyond this, firm finances are necessary for enabling strategic initiatives, boosting operational efficiency, and allowing resilience in the face of economic uncertainties (Naidoo, 2010). It is essential to allocate resources appropriately, handle finances carefully, and have a strong financial structure to support future expansions, organizational stability, and maximize shareholder value (Mihajlović et al., 2020).

5. Analysis

5.1. Measurement model

For the analysis of our model, we used partial least squares (PLS) path modelling. There are two reasons for using structural equation modelling: (1) our model is complex since it has many constructs and indicators (Benitez et al., 2020), and (2) we have to calculate the latent variable scores through a second-step approach in order to estimate high-order constructs (Hair et al., 2019). We used SmartPLS4 to estimate both the measurement and structural models. Significance levels (i.e., weights, loadings, and path coefficients) were determined via bootstrapping with 10,000 subsamples. For more information about the correlation matrix, refer to Table A1 of the Appendix.

The analysis of variance-based Structural Equation Modeling (SEM) results was conducted using the repeated indicator approach. The measurement model was validated by testing the construct measurement quality and by examining the associations between constructs (Sarstedt et al., 2021). The initial step involved testing the measurement model's convergent validity. This was evaluated based on factor loadings, Composite Reliability (CR), and Average Variance Extracted (AVE). As presented in Table A2, all item loadings surpassed the recommended threshold of 0.6 (Chin et al., 2008). Additionally, the Composite Reliability values, indicating the extent to which construct indicators reflect the latent construct, exceeded the suggested threshold of 0.7, while the Average Variance Extracted values, reflecting the overall variance explained by the latent construct, surpassed the recommended threshold of 0.5 (Hair Jr et al., 2021).

The next thing was to assess the reliability and validity of the model. Reliability refers to the consistency and repeatability of the measurement, and validity refers to the accuracy and appropriateness of the measurement. Table A3 shows that all values are above the recommended value of 0.7 for both cronbach's alpha and Composite Reliability (CR), including rho_c and rho_c, and above. Furthermore, the values exceed the recommended value of 0.5 for Average Variance Extracted (AVE), indicating a solid foundation for the model's reliability. Discriminant validity values exceed the recommended values in all cases, with one small exception for internal and external legitimacy, where the values are 0.792 for external legitimacy and 0.808 for internal legitimacy (marked in red in Table A4).

Regarding collinearity statistics, the variance inflation factor (VIF) values seem, as shown in Table A5, to not exceed the of value, suggesting minimal multicollinearity concerns. Given the statistically acceptable values, the assessment process extended to the second order, reaffirming the model's reliability. Outer loadings, once again, surpassed the critical threshold of 0.707, indicating strong construct validity (refer to Table A6). While some VIF values exceeded 3, they remained below the threshold of 5 (refer to Table A7). Upon inspecting the inner model, however, some values surpassed the threshold of 5, which warrants further attention and examination. For detailed insights, please refer to Table A8.

5.2. Structural model

The structural model derived from the PLS analysis is outlined in Figure 2, where it encapsulates both the explained variance of endogenous variables (R2) and the standardized path coefficients (β). Validation of this model involves examining the coefficient of determination (R2) values, predictive relevance (Stone-Geisser Q2), and the effect size of path coefficients, with the significance of estimates (t-statistics) determined through a bootstrap analysis with 10,000 resamples. As depicted in Figure 2, four out of five hypotheses were empirically supported. Specifically, a firm's RAIG practices were found to significantly impact its performance ($\beta = 0.549$, $t = 6.646$, $p < 0.001$) and legitimacy ($\beta = 0.591$, $t = 9.016$, $p < 0.01$). Moreover, legitimacy was shown to have a significant effect on a firm's performance ($\beta = 0.306$, $t = 3.477$, $p < 0.001$). Additionally, RAI communications exhibited a positive association with legitimacy ($\beta = 0.336$, $t = 4.796$, $p < 0.001$), but no significant relationship was found for RAI communications as a mediator for RAIG practices and legitimacy ($\beta = -0.029$, $t = 1.593$, $p > 0.1$). The structural model explained 81.3% of the variance for legitimacy (R2 = 0.813) and 69.1% for competitive performance (R2 = 0.691), indicating moderate to substantial predictive power. Furthermore, the effect size f2 was evaluated, revealing moderate to high effect sizes for all direct values, except for legitimacy (0.07) and the mediator, both falling below the thresholds. Aligned with IS research, we similarly investigated the impact of control variables on performance. The results revealed that firm size and AI department size have a relationship with the dependent variable, but employes' years within the firm and the length of time that a company uses AI did not reveal any statistically significant relationship with the dependent variable.

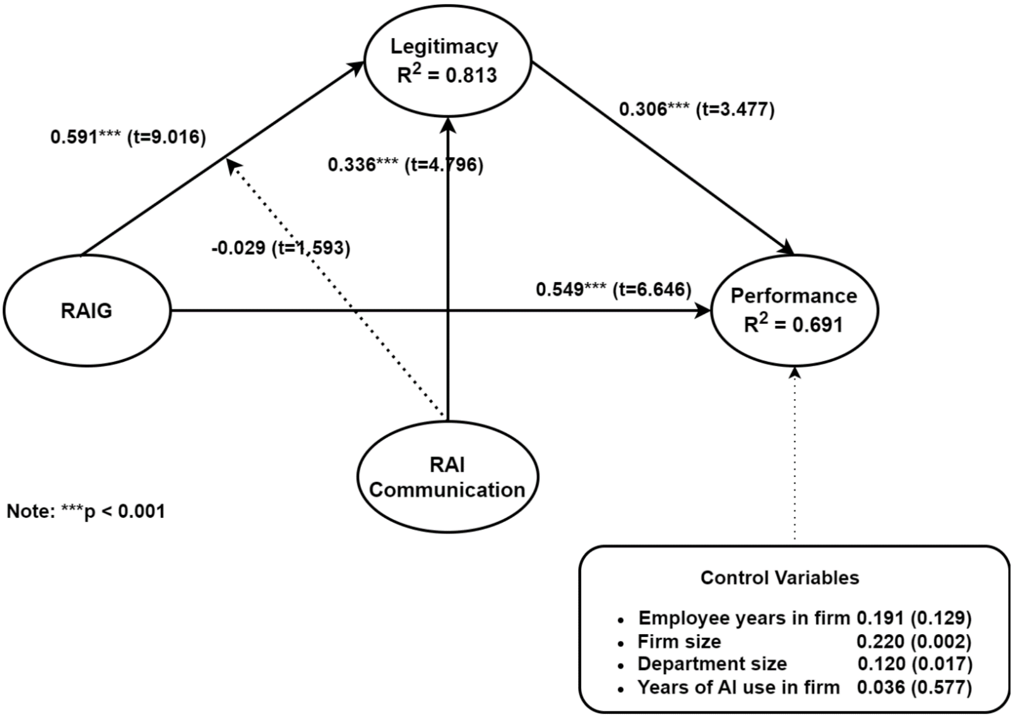

Figure 2. Estimated relationships of structural model.

To assess the precision and evaluate the statistical methods, we calculate intervals bias, where no zeros are present and values increase from 2.5% to 97.5%, so these paths are significant (Table 3).

Table 3. Path coefficient and intervals bias.

	Original sample (O)	**Sample mean (M)**	**Bias**	**2.5%**	**97.5 %**
RAI Communications -> Legitimacy	0.336	0.342	0.007	0.207	0.475
Legitimacy -> Performance	0.306	0.299	-0.007	0.141	0.485
RAIG -> Legitimacy	0.591	0.587	-0.004	0.457	0.709
RAIG -> Performance	0.549	0.559	0.009	0.374	0.698
RAI Communications x RAIG -> Legitimacy	-0.029	-0.028	0.001	-0.067	0.005

For the outer loadings, which are the estimated relationships in reflective measurement models and determine an item's absolute contribution to its assigned construct, we see that the values p values are statistically significant.

Table 4. Outer loadings.

	Original sample (O)	**Sample mean (M)**	**Standard deviation (STDEV)**	**T statistics (|O/STDEV|)**	**P values**
Accountability -> RAIG	0.839	0.836	0.026	32.518	*
Competition -> Performance	0.95	0.947	0.018	52.42	*
Customer relationship -> Performance	0.91	0.909	0.023	40.309	*
Data governance -> RAIG	0.945	0.941	0.014	65.639	*
Environmental and societal well-being -> RAIG	0.874	0.871	0.037	23.806	*
Explicit RAI Communication -> RAI Communications	0.973	0.971	0.011	86.833	*
External legitimacy -> Legitimacy	0.916	0.915	0.018	49.898	*
Fairness -> RAIG	0.869	0.866	0.025	34.87	*
Finances -> Performance	0.919	0.918	0.021	44.302	*
Human agency and oversight -> RAIG	0.934	0.931	0.016	60.161	*
Implicit RAI Communication -> RAI Communications	0.954	0.953	0.016	58.498	*
Internal Legitimacy -> Legitimacy	0.977	0.976	0.012	80.939	*
Technical robustness and safety -> RAIG	0.942	0.939	0.015	62.589	*
Transparency -> RAIG	0.859	0.856	0.03	28.368	*
RAI Communications x RAIG -> Communications x RAIG	1	1	0	n/a	n/a

Note: * < 0.001

We also check outer weights for validity, which involves assessing the measurement model's validity by examining the relationships between observed variables (indicators) and latent constructs. These weights are obtained through the PLS regression, which aims to maximize the covariance between the indicators and the latent construct. Outer weights did have some p-values that exceed 0.05, as shown in Table 5.

Table 5. Outer weight.

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
Accountability -> RAIG	0	0.002	0.061	0.002	NS
Competition -> Performance	0.467	0.461	0.088	5.297	*
Customer relationship -> Performance	0.335	0.335	0.078	4.278	*
Data governance -> RAIG	0.31	0.307	0.076	4.071	*
Environmental and societal well-being -> RAIG	0.151	0.163	0.072	2.096	**
Explicit RAI Communication -> Communications	0.584	0.579	0.086	6.769	*
External legitimacy -> Legitimacy	0.365	0.361	0.078	4.697	*
Fairness -> RAIG	0.021	0.016	0.077	0.268	NS
Finances -> Performance	0.273	0.277	0.078	3.508	*
Human agency and oversight -> RAIG	0.25	0.247	0.068	3.67	*
Implicit RAI Communication -> RAI Communications	0.452	0.457	0.087	5.184	*
Internal Legitimacy -> Legitimacy	0.682	0.685	0.07	9.753	*
Technical robustness and safety -> RAIG	0.268	0.26	0.09	2.988	*
Transparency -> RAIG	0.083	0.084	0.062	1.324	NS
RAI Communications x RAIG -> Communications x RAIG	1	1	0	n/a	n/a

Note: * < 0.001, ** < 0.05, NS > 0.05

5.3 Test for mediation

Mediation test allows researchers to check specific hypotheses about how indirect connections between independent and dependent variables work through mediators. It lets researchers see if these indirect relationships are important and how strong they are in their models. To test for mediation, we assess the total effect of the independent variable on the dependent variable without considering the mediator; thus, this step involves estimating the direct effect of the dependent variable on the independent variable. Then, we evaluate the indirect effect of the dependent variable on the independent variable through the mediator. This step allows us to estimate the effect of the dependent variable on the mediator, the effect of mediator on the independent variable, and the product of paths between the effect of dependent variable and the effect of the mediator on the independent variable, which is the indirect effect. We used bootstrap with 10000 subsamples, and we checked whether the indirect effect (mediation) was statistically significant. When the indirect effect is significant, it means that the mediator mediates the relationship between the dependent variable and the independent variable.

For our model, mediation analysis was performed to assess the mediating role of RAI communications in the relationship between legitimacy and firm performance. The results (see Table 6) revealed a significant indirect effect of legitimacy on performance (β=0.103, t=2.654, p<0.01). The total effect of legitimacy on performance was significant; with the inclusion of the mediator the effect of legitimacy on performance was still significant (β=0.306, t=3.477, p<0.001). The mediating role of RAIG in the relationship between legitimacy and performance revealed a significant indirect effect of legitimacy on performance (β=0.181, t=3.345, p<0.01). The total effect of legitimacy on performance was significant, with the inclusion of the mediator, the effect of legitimacy on performance was still significant (b=0.591, t=9.016, p<0.001). As for the mediating role of RAI communications in the relationship between RAIG, legitimacy and performance did not reveal a significant indirect effect of legitimacy on performance (β=-0.009, t=1.479, p>0.1). The total effect of RAI communications as a mediator for RAI practices and legitimacy on performance was not significant (b=-0.029, t=1.593, p>0.1). For the first two, we see that there is no zero between the intervals (last two columns in Table 6), which means there a mediation, but there is a zero for the last one, meaning there is no mediation between them; thus, H4 is not supported.

Table 6. Indirect effects.

	Original sample (O)	**Sample mean (M)**	**Standard deviation (STDEV)**	**T statistics (|O/STDEV|)**	**P values**	**2.5%**	**97.5 %**
RAI Communications -> Legitimacy -> Performance	0.103	0.103	0.039	2.654	*	0.038	0.188
RAIG -> Legitimacy -> Performance	0.181	0.175	0.054	3.345	*	0.073	0.288
RAI Communications x RAIG -> Legitimacy -> Performance	-0.009	-0.008	0.006	1.479	NS	-0.021	0.002

Note: * < 0.001, ** < 0.05, NS > 0.05

5.4 Post-hoc analysis

As a final step in our analysis, we conducted a post-hoc analysis to evaluate whether there exists a statistical difference between companies originating from the USA and Europe. Multigroup analysis is used to assess predefined data groups in order to ascertain significant differences across group-specific parameter estimates. MGA was chosen because it allows researchers to examine variations between different groups within two identical models, provided that the groups are known. This analysis was conducted using a non-parametric test, specifically PLS-

MGA. The outcomes are displayed in Table A9. The findings indicate that there is no significant difference between the two groups, implying that there is no distinction in the comparison between European firms and those from the USA. The lack of differences might be due to the fact that RAIG is a relatively new concept and has not been mature enough, hence finding no significant differences should not be a surprise. Besides that, Europe and the USA have similar perspectives on ethical aspects in comparison with other countries that develop AI on massive scale, like China, which could be an interpretation of the results.

5.5 Finding summary

This study examined the proposed research model through data collection via an online survey. The results indicate that RAIG practices directly boost performance and legitimacy, supporting hypotheses 1 and 5. Moreover, the findings demonstrate that firms can enhance their performance by building legitimacy through stakeholder trust in their practices, leading to an increased willingness to engage with the firm's offerings or adopt its practices, thus confirming hypothesis 3. Additionally, the results underscore the importance of RAIG communications in enhancing firm legitimacy, highlighting the significance of communicating RAIG practices to both internal and external stakeholders for overall acceptance, as posited in hypothesis 2. However, the findings did not reveal any significant moderating effect of RAIG communications between RAIG practices and legitimacy, failing to confirm hypothesis 4. Our post-hoc analysis showed no significant differences between the USA and Europe. In summary, the findings emphasize the critical role of RAIG practices in garnering internal and external acceptance while also positively impacting firm performance.

6. Discussion

6.1. Research implications

The outcomes of this investigation significantly contribute to the extant scholarly literature by empirically testing the significance of RAIG and legitimacy in attaining a certain level of organizational performance. While the literature has begun to explore RAIG (Clarke, 2019; Dignum, 2019; European Commission, 2019), there remains a scarcity of research on the legitimacy granted by RAIG upon a firm. Thus, this study represents one of the initial attempts to explain the relationship between RAIG, legitimacy, and RAI communications. In alignment with the study's objectives, five hypotheses were formulated and tested. The SEM results confirmed all hypotheses except for the anticipated moderate effect of RAIG. Consequently, the findings affirm that RAIG exerts a positive direct influence on performance and indirectly through legitimacy. Furthermore, RAI communications are found to boost legitimacy, even though without a moderating effect between RAIG and legitimacy. This study contributes to the literature around RAIG by emphasizing its role in not only governing AI systems but also in bestowing legitimacy to organizations that have systems with RAIG practices (Grimmelikhuijsen & Meijer, 2022). This expands the conceptualization of RAIG beyond its traditional operational and ethical dimensions. By uncovering the mediating role of legitimacy, this research underscores the importance of organizational legitimacy in the context of AI governance. It highlights how RAIG practices can enhance the perceived legitimacy of AI implementations within firms, thereby influencing organizational performance. Furthermore, The study's findings encourage a critical evaluation of RAI communication strategies within organizations. While these communications are observed to boost legitimacy, the absence of a moderating effect between RAIG and legitimacy suggests a lack of understanding of how such communications interact with broader AI governance practices. Understanding the direct and indirect pathways through which RAIG and legitimacy influence performance provides insights for organizational leaders and policymakers (Clark et al., 2014). It underscores the importance of integrating responsible AI practices not only for ethical reasons but also for enhancing organizational effectiveness and competitiveness.

Future research could explore the role of contextual factors and industry-specific dynamics in shaping the relationship between RAIG, legitimacy, and performance (Wang & Wu, 2024). This suggests a need for further theoretical exploration into the mechanisms through which RAIG practices contribute to organizational legitimacy, as well as the implications of legitimacy for organizational performance (Leonard, 2023). Furthermore, researchers could go deeper into the underlying processes through which legitimacy influences performance outcomes, exploring factors such as stakeholder perceptions, organizational reputation, and regulatory environments. Alternatively, investigating the potential moderating effects of contextual factors, such as industry dynamics and organizational culture, on the relationship between RAIG, legitimacy, and performance could provide valuable insights into the practices of AI governance in different organizational contexts (Zhang et al., 2023). Since the study highlighted the importance of RAI communications researchers could focus on the effectiveness of different communication strategies in conveying organizational commitment to responsible AI practices and their impact on stakeholders' perceptions of legitimacy. This could involve examining the role of communication channels, message framing, and organizational transparency in shaping stakeholder attitudes towards AI governance practices. Additionally, the study suggests that legitimacy theory could be extended to other similar or related disciplines, such as organizational theory, institutional theory, and stakeholder theory, to develop a better understanding of the factors influencing AI governance practices and their implications on performance (Jan et al., 2021).

6.2. Practical implications

The practical implications of this study are profound for managers navigating RAIG, legitimacy, and RAI communications. Firstly, the study emphasizes the critical role of RAIG in boosting organizational performance (Rakova et al., 2021). It underscores the importance of implementing robust RAIG practices not only for ethical reasons but also for improving organizational effectiveness and competitiveness (Lin et al., 2020). Therefore, organizational leaders are encouraged to prioritize investments in RAIG frameworks to drive performance improvements and maintain a competitive edge, especially in AI-driven business environments (Mikalef et al., 2023). Moreover, the study highlights the mediating role of legitimacy in the relationship between RAIG and organizational performance. This suggests that organizations can enhance their performance outcomes by strategically managing their legitimacy through the adoption of responsible AI practices (Rakova et al., 2021). By prioritizing transparency and accountability in their AI governance practices, organizations can bolster their perceived legitimacy, thereby positively influencing their performance metrics.

Furthermore, the study emphasizes the importance of Responsible AI communications in shaping stakeholder perceptions of legitimacy (Buhmann & Fieseler, 2021). Organizational leaders are encouraged to critically evaluate their communication strategies to ensure alignment with RAIG practices and organizational goals. Effective communication of AI-related practices and outcomes is essential for building trust and credibility among stakeholders, thereby enhancing organizational legitimacy and performance (Hohma & Lütge, 2023). Additionally, the study highlights the need for organizational leaders to consider contextual factors and industry-specific dynamics in their AI governance practices. Understanding how these factors influence the relationship between RAIG, legitimacy, and performance is crucial for tailoring AI governance strategies to specific organizational contexts. Therefore, organizational leaders are encouraged to conduct thorough assessments of their organizational context and industry landscape to inform their AI governance practices effectively.

Future research in this domain could explore several avenues to further advance our understanding of Responsible AI Governance (RAIG), legitimacy, and Responsible AI (RAI) communications. Firstly, researchers could delve deeper into the mechanisms through which RAIG practices contribute to organizational legitimacy (Guerreiro et al., 2021). This could involve exploring the specific RAIG practices that are most effective in enhancing legitimacy and identifying the underlying processes through which these practices influence stakeholder perceptions (Fisher et al., 2020). Additionally, future research could examine the implications of legitimacy for organizational performance in greater detail. This could involve investigating how different dimensions of legitimacy (e.g., cognitive, moral, pragmatic) influence various aspects of organizational performance, such as financial performance, innovation, and employee satisfaction (Soomro et al., 2021). Understanding the nuanced relationship between legitimacy and performance could provide valuable insights for organizational leaders seeking to leverage legitimacy as a strategic asset. Furthermore, researchers could explore the role of contextual factors and industry-specific dynamics in shaping the relationship between RAIG, legitimacy, and performance. This could involve conducting comparative studies across different industries to identify industry-specific challenges and opportunities related to AI governance and legitimacy (Abioye et al., 2021).

Additionally, researchers could examine how factors such as organizational culture, regulatory environments, and stakeholder expectations influence the effectiveness of RAIG practices and their impact on organizational performance (Akram et al., 2018). Another area for future research is to investigate the effectiveness of Responsible AI communications in enhancing organizational legitimacy further. This could involve conducting experimental studies to test different communication strategies and message framings to determine their impact on stakeholders' perceptions of legitimacy (Kim, 2016). Additionally, researchers could explore the role of emerging communication technologies, such as artificial intelligence-driven chatbots and virtual assistants, in facilitating transparent and effective communication of AI-related practices and outcomes (Kannan et al., 2023). Finally, future research could explore interdisciplinary approaches to understanding AI governance, legitimacy, and performance. This could involve integrating insights from disciplines such as organizational theory, institutional theory, stakeholder theory, and ethics to develop a holistic understanding of the factors influencing AI governance practices and their implications for organizational performance (Du & Xie, 2021). By adopting interdisciplinary perspectives, researchers can gain a more comprehensive understanding of the complex dynamics at play in the intersection of AI governance, legitimacy, and organizational performance.

6.3. Limitations

While online surveys are a valuable research tool, they come with a set of potential limitations. Our sample consists of responses from specific geographical areas and does not include countries like China and India, meaning that our results might not be valid for other countries with different customs and norms. Response bias may occur as respondents tailor their answers to what they perceive the researcher expects or misrepresent their genuine opinions. Selection bias is a concern too, since online surveys potentially attract specific demographics; thus, limiting the generalizability of findings. Our survey might have limited depth, making it challenging to collect in-depth information. Self-selection bias is another issue, as participants may possess distinct characteristics or motivations. Non-response bias might arise from differences between those who initiate but do not complete the survey and those who finish it. Social desirability bias can influence responses as individuals may provide socially acceptable answers or align with their perceived image rather than reveal true beliefs or behaviors. Question-wording, phrasing, and ordering of questions can unintentionally affect responses, potentially impacting result validity. Researchers had less control over the survey environment and interpretation of questions compared to other data collection methods. Data security and privacy concerns can affect participation rates, and the inability to clarify responses is a limitation, as online surveys lack the capacity for follow-up questions. Our survey may provide a limited understanding of context for some participants, as we were not present to explain our questions if needed further. Recall bias is also a concern, with respondents potentially struggling to accurately recall past experiences, leading to potential inaccuracies. Demographics are also noteworthy, as shifts in technology and internet usage patterns could impact the representativeness of the sample, although we tried to be homogenous to some extent by collecting data from people who are located in Western Europe and the USA.

7. Conclusion

In conclusion, the business performance and hence survival is occurring upon the implementation of RAIG practices. Consequently, organizations must cultivate the capability to effectively employ RAIG in order to boost their legitimacy, increase their corporate reputation, and attain superior performance outcomes. Driven by the noticeable association between RAIG and performance, the current research employed PLS-SEM to analyze survey data obtained from a sample of 329 Western European and USA companies, employing legitimacy theory as a theoretical lens. The study proceeded to develop and empirically validate various higher-order constructs, along with a conceptual model highlighting the interplay between RAIG, legitimacy, AI communication, and firm performance. The empirical findings underscore the significance of adopting a comprehensive perspective when considering the development of RAIG practices. Such an approach empowers organizations to navigate their daily operations and manage their organizational structures more effectively, ultimately culminating in improved overall performance and heightened profitability.

A. Survey instrument.

Measure	Item
Transparency	When designing and building a AI applications, interpretability and explainability are a high priority
Transparency	We design AI applications with explainability and interpretability in mind from the start
Transparency	We assess to what extent the decisions and hence the outcome made by the AI application can be understood
Transparency	We communicate to users that they are interacting with an AI application and not with another human
Transparency	We have established mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI application
Transparency	Users can provide feedback of their experience with the AI application(s)
Transparency	Processes and mechanisms for data collection, data labelling, data transformation and data use are well documented.:
Transparency	We have established well-documented processes and mechanisms for AI development
Transparency	We have adopted measures that can ensure traceability of our AI models
Fairness	We have ensured that our AI applications are accessible to all users and accommodate individual preferences and abilities
Fairness	We have involved and consulted different stakeholders (e.g. users of assistive technologies) in the AI system's development and use
Fairness	We have ensured that the information about the AI system is accessible also to users in need of assistive technologies
Fairness	We have established a process to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design
Fairness	The datasets we use for AI applications are assessed in terms of diversity and representativeness of the population
Fairness	We have put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system
Accountability	We have established an "ethical AI review board" or similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas
Accountability	We have established an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact from our AI applications
Accountability	We communicated company policies to design and development teams so there is clarity over responsibility of AI
Accountability	We have established processes that facilitate the assessment of algorithms, data and design processes
Accountability	We have established mechanisms that facilitate the system's auditability, such as logging of the AI system's processes and outcomes
Accountability	Third parties (e.g. suppliers, consumers, distributors/vendors) or workers can easily report potential vulnerabilities, risks or biases in the AI system?
Robustness and Safety	We assess if our AI applications are making unacceptable amount of inaccurate predictions
Robustness and Safety	We have processes in place to increase the AI applications' accuracy
Robustness and	We have processes in place to figure out if there is a need for additional data to

Safety	improve accuracy
Robustness and Safety	We have put in place verification methods to measure and ensure different aspects of the system's reliability
Robustness and Safety	We have tested whether specific contexts or particular conditions need to be taken into account to ensure AI reproducibility
Robustness and Safety	We have processes in place for describing when an AI system fails in certain types of settings
Robustness and Safety	We have verified how our AI system (models) behaves in unexpected situations and environments
Robustness and Safety	We have considered the level of risk raised by the AI system in specific use cases
Robustness and Safety	We are Identifying, assessing, documenting and minimizing the potential negative impacts of AI systems
Robustness and Safety	We have assessed potential forms of attacks to which AI systems could be vulnerable (E.g. data pollution, physical infrastructure, cyber-attacks)
Robustness and Safety	We have measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks
Data governance	We continuously monitor our AI applications to know that the models/datasets have not been compromised or hacked
Data governance	We continuously assess the quality and integrity of our data
Data governance	We do periodic reviewing and updating of our AI datasets
Data governance	The data follows relevant standards (ISO, IEEE) or protocols for data management and governance
Data governance	We always enhance privacy by e.g. encrypting, anonymizing and aggregating our data where it is needed
Data governance	We consider ways of training AI models without, or with minimal, use of potentially sensitive or personal data
Data governance	We have ensured that our products and services that use anonymized data pose no unreasonable risk of re- identification
Data governance	We ensure that people who access data are qualified, and that they have the necessary competence to understand the details of data protection policy
Data governance	We always log data on when, why, and by whom data is accessed
Data governance	We have established access rights and policies to the relevant datasets
Human agency and oversight	We have safeguards to prevent overconfidence and overreliance on AI applications
Human agency and oversight	We have considered the appropriate level of human control for particular AI systems and use cases
Human agency and oversight	We ensure that an AI system does not undermine human autonomy or causes other adverse effects
Human agency and oversight	We have assessed whether there is a probable chance that the AI system may cause damage or harm to users or third parties
Human agency and oversight	We have assessed the possible negative impacts of our AI products and services on human rights
Human agency and oversight	We ensure that an AI system does not undermine human autonomy or causes other adverse effects
Environmental and societal well-being	We monitor and consider the effects that our AI system have on the environment
Environmental	We have established mechanisms to measure and reduce the environmental impact of

and societal well-being	the AI system's development, deployment and use
Environmental and societal well-being	Our AI systems are designed so that they minimize negative impacts on the environment
Environmental and societal well-being	We have ensured that the social impacts of the AI system are well understood
Environmental and societal well-being	We clarify the purpose of the AI applications and who or what may benefit from its use
Environmental and societal well-being	We you take action to minimize potential societal harm that may be caused by our AI systems
External legitimacy	Our company has won social recognition and praise
External legitimacy	Our company has established good relationships with non-governmental organizations
External legitimacy	Our company has strengthened its relationships with suppliers
External legitimacy	Our company has strengthened its relationships with customers
External legitimacy	Our company always comply with the newest legal principles as soon as possible
Internal legitimacy	Our activities can strengthen the internal cohesion in the company
Internal legitimacy	Our activities can increase employee satisfaction in the company
Internal legitimacy	Our activities can improve operational efficiency in the company
Internal legitimacy	Our company works with regulators to adapt, formulate and develop the regulatory standards
Internal legitimacy	Our company's management is seen as accountable for its actions and decisions
Explicit (laws and regulations of society) communication	We communicate to stakeholders how we comply with laws and regulations concerning responsible AI
Explicit (laws and regulations of society) communication	We communicate to stakeholders that we integrate ethical language into our procedures responsible AI
Explicit (laws and regulations of society) communication	We communicate to stakeholders that we collaborate exclusively with organizations complying to laws and regulations when integrating our AI systems
Explicit (laws and regulations of society) communication	We communicate to stakeholders about instances where our practices did not align with legal expectations around AI development

Explicit (laws and regulations of society) communication	We communicate to stakeholders about our diversity teams of AI development
Implicit (expectations and values of society) communication	We communicate to stakeholders about our company's initiatives, policies, and ethical guidelines on responsible AI
Implicit (expectations and values of society) communication	We communicate to stakeholders how our mission and values reflect a commitment to ethical use of AI
Implicit (expectations and values of society) communication	We communicate to stakeholders about the training provided to our employees to ensure responsible AI practices are followed
Implicit (expectations and values of society) communication	We communicate to stakeholders about awards or recognitions received for our responsible AI practices
Performance	Our customer satisfaction has been increased: Compared with your key competitors
Performance	Our customer loyalty has been increased: Compared with your key competitors
Performance	Our employee satisfaction has been increased: Compared with your key competitors
Performance	Our company provided employment to the local economy: Compared with your key competitors
Performance	Our company created new products/services: Compared with your key competitors
Performance	Our profitability has increased: Compared with your key competitors
Performance	Our return on investment has increased: Compared with your key competitors
Performance	Our sales growth has increased: Compared with your key competitors
Performance	Our financial stability has increased: Compared with your key competitors
Performance	Our company often defeats the main competitors in the marketplace: Compared with your key competitors
Performance	Our company provides higher quality products/services to customers compared with the main competitors: Compared with your key competitors
Performance	Our company responds rapidly to market demands compared to the main competitors: Compared with your key competitors
Performance	Our company can respond more promptly to environmental changes as compared to the main competitors: Compared with your key competitors
Performance	Our company adapts rapidly in market demands: Compared with your key competitors

Declaration of competing interest

The authors declare no conflict of interests regarding the publication of this article.

References

Abdullah, Z., & Abdul Aziz, Y. (2013). Institutionalizing corporate social responsibility: effects on corporate reputation, culture, and legitimacy in Malaysia. *Social Responsibility Journal*, *9*(3), 344-361.

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*, 103299.

Akram, M. S., Goraya, M., Malik, A., & Aljarallah, A. M. (2018). Organizational performance and sustainability: exploring the roles of IT capabilities and knowledge management capabilities. *Sustainability*, *10*(10), 3816.

Al-Abrrow, H., Ali, J., & Alnoor, A. (2022). Multilevel influence of routine redesigning, legitimacy and functional affordance on sustainability accounting: Mediating role of organizational sense-making. *Global Business Review*, *23*(2), 287-312.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Montreal, QC, Canada.

Anh, D. N. P., & Velencei, J. (2019). Measuring corporate social performance. *Serbian Journal of Management*, *14*(1), 193-204.

Archel, P., Husillos, J., Larrinaga, C., & Spence, C. (2009). Social disclosure, legitimacy theory and the role of the state. *Accounting, Auditing & Accountability Journal*.

Babiak, K., & Trendafilova, S. (2011). CSR and environmental responsibility: Motives and pressures to adopt green management practices. *Corporate social responsibility and environmental management*, *18*(1), 11-24.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2021). Artificial intelligence, firm growth, and product innovation. *Firm Growth, and Product Innovation (November 9, 2021)*.

Battilana, J., Leca, B., & Boxenbaum, E. (2009). 2 how actors change institutions: towards a theory of institutional entrepreneurship. *The academy of management annals*, *3*(1), 65-107.

Bedué, P., & Fritzsche, A. (2022). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management*, *35*(2), 530-549.

Benford, R. D., & Snow, D. A. (2000). Framing processes and social movements: An overview and assessment. *Annual review of sociology*, *26*(1), 611-639.

Benitez, J., Henseler, J., Castillo, A., & Schuberth, F. (2020). How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory IS research. *Information & Management*, *57*(2), 103168.

Bergek, A., Berggren, C., Magnusson, T., & Hobday, M. (2013). Technological discontinuities and the challenge for incumbent firms: Destruction, disruption or creative accumulation? *Research Policy*, *42*(6-7), 1210-1224.

Bes, B. J., Sommerer, T., & Agné, H. (2019). On legitimacy crises and the resources of global governance institutions: A surprisingly weak relationship? *Global Policy*, *10*(3), 313-326.

Beyers, J., & Arras, S. (2021). Stakeholder consultations and the legitimacy of regulatory decision-making: A survey experiment in Belgium. *Regulation & Governance*, *15*(3), 877-893.

Bitektine, A. (2011). Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review*, *36*(1), 151-179.

Blanco-Gonzalez, A., Diéz-Martín, F., Cachón-Rodríguez, G., & Prado-Román, C. (2020). Contribution of social responsibility to the work involvement of employees. *Corporate social responsibility and environmental management*, *27*(6), 2588-2598.

Breit, E. (2014). Remedy through paradox? Constructions of internal legitimacy in a publicly discredited organization. *Management Communication Quarterly*, *28*(4), 585-608.

Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, *64*, 101475.
Burlea, A. S., & Popa, I. (2013). Legitimacy theory. *Encyclopedia of corporate social responsibility*, *21*, 1579-1584.
Carroll, A. B. (1999). Corporate social responsibility: Evolution of a definitional construct. *Business & society*, *38*(3), 268-295.
Carter, D. (2018). How real is the impact of artificial intelligence? The business information survey 2018. *Business Information Review*, *35*(3), 99-115.
Caspersen, N. (2015). Degrees of legitimacy: Ensuring internal and external support in the absence of recognition. *Geoforum*, *66*, 184-192.
Chan, M. C., Watson, J., & Woodliff, D. (2014). Corporate governance quality and CSR disclosures. *Journal of business ethics*, *125*, 59-73.
Chin, W. W., Peterson, R. A., & Brown, S. P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of marketing theory and practice*, *16*(4), 287-298.
Clark, J. R., Murphy, C., & Singer, S. J. (2014). When do leaders matter? Ownership, governance and the influence of CEOs on firm performance. *The Leadership Quarterly*, *25*(2), 358-372.
Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, *35*(4), 410-422.
Colleoni, E. (2013). CSR communication strategies for organizational legitimacy in social media. *Corporate Communications: an international journal*, *18*(2), 228-248.
Cornelissen, J. P., & Werner, M. D. (2014). Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. *Academy of Management Annals*, *8*(1), 181-235.
Cowls, J., King, T., Taddeo, M., & Floridi, L. (2019). Designing AI for social good: Seven essential factors. *Available at SSRN 3388669*.
Crossley, R. M., Elmagrhi, M. H., & Ntim, C. G. (2021). Sustainability and legitimacy theory: The case of sustainable social and environmental practices of small and medium-sized enterprises. *Business Strategy and the Environment*, *30*(8), 3740-3762.
Dai, N. T., Du, F., Young, S. M., & Tang, G. (2018). Seeking legitimacy through CSR reporting: Evidence from China. *Journal of Management Accounting Research*, *30*(1), 1-29.
de Laat, P. (2021). Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? *Philosophy & technology*, *34*(4), 1135-1193. https://doi.org/10.1007/s13347-021-00474-3
Deegan, C. (2002). Introduction: The legitimising effect of social and environmental disclosures–a theoretical foundation. *Accounting, Auditing & Accountability Journal*, *15*(3), 282-311.
Dellmuth, L., & Schlipphak, B. (2020). Legitimacy beliefs towards global governance institutions: a research agenda. *Journal of European Public Policy*, *27*(6), 931-943. https://doi.org/10.1080/13501763.2019.1604788
Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, *121*, 283-314.
Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (1 ed.). Springer Nature. https://doi.org/10.1007/978-3-030-30371-6
Drori, I., & Honig, B. (2013). A process model of internal and external legitimacy. *Organization studies*, *34*(3), 345-376.
Du, S., El Akremi, A., & Jia, M. (2022). Quantitative Research on Corporate Social Responsibility: A Quest for Relevance and Rigor in a Quickly Evolving, Turbulent World. *Journal of business ethics*, 1-15.
Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, *129*, 961-974.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in ai systems. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems,
Enholm, I. M. (2021). *Responsible AI Governance and Its*

Effect on Competitive Performance Norwegian University of Science and Technology].
European Commission. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, *26*(6), 3333-3361. https://doi.org/10.1007/s11948-020-00276-4
Fisher, G. (2020). The complexities of new venture legitimacy. *Organization Theory*, *1*(2), 2631787720913881.
Fisher, G., Kotha, S., & Lahiri, A. (2016). Changing with the times: An integrated view of identity, legitimacy, and new venture life cycles. *Academy of Management Review*, *41*(3), 383-409.
Fisher, G., Kuratko, D. F., Bloodgood, J. M., & Hornsby, J. S. (2017). Legitimate to whom? The challenge of audience diversity and new venture legitimacy. *Journal of Business Venturing*, *32*(1), 52-71.
Fisher, G., Stevenson, R., Neubert, E., Burnell, D., & Kuratko, D. F. (2020). Entrepreneurial hustle: Navigating uncertainty and enrolling venture stakeholders through urgent and unorthodox action. *Journal of Management Studies*, *57*(5), 1002-1036.
Francés-Gómez, P. (2020). Social contract theory and business legitimacy. *Handbook of Business Legitimacy: Responsibility, Ethics and Society*, 277-295.
Gabrielova, K., & Buchko, A. A. (2021). Here comes Generation Z: Millennials as managers. *Business Horizons*, *64*(4), 489-499.
Gammelgaard, J., & Kumar, R. (2023). Internal legitimacy crises: the roles of the regulatory focus and relational social capital of headquarters and subsidiaries. *Critical Perspectives on International Business*, *19*(3), 469-486.
Giacomini, D., Martini, M., Sancino, A., Zola, P., & Cavenago, D. (2021). Corporate social responsibility actions and organizational legitimacy at the peak of COVID-19: a sentiment analysis. *Corporate Governance: The international journal of business in society*, *21*(6), 1043-1058.
Goldsmith, R. E., Lafferty, B. A., & Newell, S. J. (2000). The impact of corporate credibility and celebrity credibility on consumer reaction to advertisements and brands. *Journal of advertising*, *29*(3), 43-54.
Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance*, *5*(3), 232-242.
Guerreiro, M. S., Lima Rodrigues, L., & Craig, R. (2021). Institutional theory and IFRS: an agenda for future research. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, *50*(1), 65-88.
Guttal, T. (2023). AI as a Communication Facilitator: Shared Decision-Making Inspired Strategies for Bipolar Disorder Diagnosis and Treatment. *arXiv preprint arXiv:2304.07878*.
Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, *31*(1), 2-24.
Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.
Haniffa, R. M., & Cooke, T. E. (2005). The impact of culture and governance on corporate social reporting. *Journal of accounting and public policy*, *24*(5), 391-430.
Harrison, J. S., Bosse, D. A., & Phillips, R. A. (2010). Managing for stakeholders, stakeholder utility functions, and competitive advantage. *Strategic Management Journal*, *31*(1), 58-74.

Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: the EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. *European Business Organization Law Review*, *22*, 593-625.
Hohma, E., & Lütge, C. (2023). From Trustworthy Principles to a Trustworthy Development Process: The Need and Elements of Trusted Development of AI Systems. *AI*, *4*(4), 904-925.
Islam, T., Islam, R., Pitafi, A. H., Xiaobei, L., Rehmani, M., Irfan, M., & Mubarak, M. S. (2021). The impact of corporate social responsibility on customer loyalty: The mediating role of corporate reputation, customer satisfaction, and trust. *Sustainable Production and Consumption*, *25*, 123-135.
Jan, A. A., Lai, F.-W., Draz, M. U., Tahir, M., Ali, S. E. A., Zahid, M., & Shad, M. K. (2021). Integrating sustainability practices into islamic corporate governance for sustainable firm performance: From the lens of agency and stakeholder theories. *Quality & Quantity*, 1-24.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.
Kannan, H., Rodriguez, R. V., Paprika, Z. Z., & Ade-Ibijola, A. (2023). *Exploring Ethical Dimensions of Environmental Sustainability and Use of AI*. IGI Global.
Khakurel, J., Penzenstadler, B., Porras, J., Knutas, A., & Zhang, W. (2018). The rise of artificial intelligence under the lens of sustainability. *Technologies*, *6*(4), 100.
Khan, O., Daddi, T., & Iraldo, F. (2020). The role of dynamic capabilities in circular economy implementation and performance of companies. *Corporate social responsibility and environmental management*, *27*(6), 3018-3033.
Khuong, M. N., Truong An, N. K., & Thanh Hang, T. T. (2021). Stakeholders and Corporate Social Responsibility (CSR) programme as key sustainable development strategies to promote corporate reputation—evidence from vietnam. *Cogent Business & Management*, *8*(1), 1917333.
Kim, Y. (2016). Understanding publics' perception and behaviors in crisis communication: Effects of crisis news framing and publics' acquisition, selection, and transmission of information in crisis situations. *Journal of Public Relations Research*, *28*(1), 35-50.
King, B. G., Bentele, K. G., & Soule, S. A. (2007). Protest and policymaking: Explaining fluctuation in congressional attention to rights issues, 1960–1986. *Social forces*, *86*(1), 137-163.
Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, *28*(1), 227-261.
Koh, K., Li, H., & Tong, Y. H. (2023). Corporate social responsibility (CSR) performance and stakeholder engagement: Evidence from the quantity and quality of CSR disclosures. *Corporate social responsibility and environmental management*, *30*(2), 504-517.
Kostova, T., & Zaheer, S. (1999). Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. *Academy of Management Review*, *24*(1), 64-81.
Kristoffersen, E., Mikalef, P., Blomsma, F., & Li, J. (2021). The effects of business analytics capability on circular economy implementation, resource orchestration capability, and firm performance. *International Journal of Production Economics*, *239*, 108205.
Kulkov, I., Kulkova, J., Rohrbeck, R., Menvielle, L., Kaartemo, V., & Makkonen, H. (2023). Artificial intelligence-driven sustainable development: Examining organizational, technical, and processing approaches to achieving global goals. *Sustainable Development*.
Laato, S., Tiainen, M., Najmul Islam, A., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. *Internet Research*, *32*(7), 1-31.
Lee, L. (2011). Business-community partnerships: understanding the nature of partnership. *Corporate Governance: The international journal of business in society*.
Lei, D., & Slocum Jr, J. W. (2005). Strategic and organizational requirements for competitive advantage. *Academy of Management Perspectives*, *19*(1), 31-45.
Leonard, K. A. K. (2023). *DEVELOPING A RESPONSIBLE AI INSTRUCTIONAL FRAMEWORK FOR ENHANCING AI LEGISLATIVE EFFICACY IN THE UNITED STATES* Purdue University Graduate School].

Levitov, A. (2015). Human rights, self-determination, and external legitimacy. *Politics, Philosophy & Economics, 14*(3), 291-315.
Lin, W. L., Yip, N., Ho, J. A., & Sambasivan, M. (2020). The adoption of technological innovations in a B2B context and its impact on firm performance: An ethical leadership perspective. *Industrial Marketing Management*, *89*, 61-71.
Lindgren, R., Eriksson, O., & Lyytinen, K. (2015). Managing identity tensions during mobile ecosystem evolution. *Journal of Information Technology*, *30*(3), 229-244.
Liu, Q., Qu, X., Wang, D., Abbas, J., & Mubeen, R. (2022). Product market competition and firm performance: business survival through innovation and entrepreneurial orientation amid COVID-19 financial crisis. *Frontiers in Psychology*, *12*, 790923.
Lounsbury, M., & Crumley, E. T. (2007). New practice creation: An institutional perspective on innovation. *Organization studies*, *28*(7), 993-1012.
Lu, J. W., & Xu, D. (2006). Growth and survival of international joint ventures: An external-internal legitimacy perspective. *Journal of management*, *32*(3), 426-448.
Maas, K., & Liket, K. (2011). Talk the walk: Measuring the impact of strategic philanthropy. *Journal of business ethics*, *100*, 445-464.
MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS quarterly*, 293-334.
Maier, C., Thatcher, J. B., Grover, V., & Dwivedi, Y. K. (2023). Cross-sectional research: A critical perspective, use cases, and recommendations for IS research. In (pp. 102625): Elsevier.
Mäkelä, H., & Näsi, S. (2010). Social responsibilities of MNCs in downsizing operations: A Finnish forest sector case analysed from the stakeholder, social contract and legitimacy theory point of view. *Accounting, Auditing & Accountability Journal*.
Malsch, B. (2013). Politicizing the expertise of the accounting industry in the realm of corporate social responsibility. *Accounting, Organizations and Society*, *38*(2), 149-168.
Mark Van Rijmenam, & Schweitzer, J. (2018). How to build responsible AI? Lessons for governance from a conversation with Tay. AOM Specialized Conference: Big Data and Managing in a Digital Economy, Surry, UK.
Men, L. R. (2014). Why leadership matters to internal communication: Linking transformational leadership, symmetrical communication, and employee outcomes. *Journal of Public Relations Research*, *26*(3), 256-279.
Mihajlović, M. B., Tadin, D. B., & Gordić, B. M. (2020). The role of financial management in the company. *Tehnika*, *75*(4), 498-503.
Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems*

31(3), 257-268.
Mikalef, P., Lemmer, K., Schaefer, C., Ylinen, M., Fjørtoft, S. O., Torvatn, H. Y., Gupta, M., & Niehaves, B. (2023). Examining how AI capabilities can foster organizational performance in public organizations. *Government Information Quarterly*, *40*(2), 101797.
Min, J., Kim, J., & Yang, K. (2023). CSR attributions and the moderating effect of perceived CSR fit on consumer trust, identification, and loyalty. *Journal of Retailing and Consumer Services*, *72*, 103274.
Morsing, M., & Schultz, M. (2006). Corporate social responsibility communication: stakeholder information, response and involvement strategies. *Business ethics: A European review*, *15*(4), 323-338.
Nagy, B. G., Rutherford, M. W., Truong, Y., & Pollack, J. M. (2017). Development of the legitimacy threshold scale. *Journal of Small Business Strategy*, *27*(3), 50-58.
Naidoo, V. (2010). Firm survival through a crisis: The influence of market orientation, marketing innovation and business strategy. *Industrial Marketing Management*, *39*(8), 1311-1320.

Naradda Gamage, S. K., Ekanayake, E., Abeyrathne, G., Prasanna, R., Jayasundara, J., & Rajapakshe, P. (2020). A review of global challenges and survival strategies of small and medium enterprises (SMEs). *Economies*, *8*(4), 79.

ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. *Philosophy & technology*, *33*, 571-593.

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2022). Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. *Information Systems Frontiers*, *25*(1), 123–114. https://doi.org/10.1007/s10796-022-10251-y

Papagiannidis, E., Mikalef, P., Krogstie, J., & Conboy, K. (2022). From responsible AI governance to competitive performance: The mediating role of knowledge management capabilities. The Role of Digital Technologies in Shaping the Post-Pandemic World: 21st IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2022, Newcastle upon Tyne, UK, September 13–14, 2022, Proceedings,

Papagiannidis Emmanouil, Mikalef, P., & Conboy, K. (2023). Responsible AI governance: a systematic literature review. *Manuscript in preparation*.

Papasolomou-Doukakis, I., Krambia-Kapardis, M., & Katsioloudes, M. (2005). Corporate social responsibility: the way forward? Maybe not! A preliminary study in Cyprus. *European business review*.

Park, H.-S., Auh, S., Maher, A. A., & Singhapakdi, A. (2012). Marketing's accountability and internal legitimacy: Implications for firm performance. *Journal of Business Research*, *65*(11), 1576-1582.

Patten, D. M. (2020). Seeking legitimacy. *Sustainability Accounting, Management and Policy Journal*, *11*(6), 1009-1021.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, *84*(12), 78-92.

Pwc. (2019). *Sizing the prize: exploiting the AI Revolution, What's the real value of AI for your business and how can you capitalise?* PwC's Global Artificial Intelligence Study. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS quarterly*, 225-246.

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-23.

Riahi-Belkaoui, A. (1992). Executive compensation, organizational effectiveness, social performance and firm performance: An empirical investigation. *Journal of Business Finance and Accounting*, *19*(1), 25-38.

Rivard, S., Raymond, L., & Verreault, D. (2006). Resource-based view and competitive strategy: An integrated model of the contribution of information technology to firm performance. *The Journal of Strategic Information Systems*, *15*(1), 29-50.

Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, 100005.

Santos-Vijande, M. L., López-Sánchez, J. Á., & Trespalacios, J. A. (2012). How organizational learning affects a firm's flexibility, competitive strategy, and performance. *Journal of Business Research*, *65*(8), 1079-1089.

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2021). Partial least squares structural equation modeling. In *Handbook of market research* (pp. 587-632). Springer.

Schultz, M. D., & Seele, P. (2023). Towards AI ethics' institutionalization: Knowledge bridges from business ethics to advance organizational AI ethics. *AI and Ethics*, *3*(1), 99-111.
Sharif, M., & Rashid, K. (2014). Corporate governance and corporate social responsibility (CSR) reporting: an empirical evidence from commercial banks (CB) of Pakistan. *Quality & Quantity*, *48*, 2501-2521.
Sharma, R. (2022). Corporate Social Responsibility and Customer Satisfaction: Role of Artificial Intelligence. *Acta Universitatis Bohemiae Meridionalis*, *25*(2), 162–174. https://doi.org/doi:10.32725/acta.2022.016
Shneiderman, B. (2020). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1-31.
Silva, S. (2021). Corporate contributions to the Sustainable Development Goals: An empirical analysis informed by legitimacy theory. *Journal of Cleaner Production*, *292*, 125962.
Singapore Government. (2020). *Model Artificial Intelligence Governance Framework* (Second Edition: Singapore Digital (SG:D), Issue. https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf
Snihur, Y., Thomas, L. D., & Burgelman, R. A. (2018). An ecosystem-level process model of business model disruption: The disruptor's gambit. *Journal of Management Studies*, *55*(7), 1278-1316.
Soomro, B. A., Mangi, S., & Shah, N. (2021). Strategic factors and significance of organizational innovation and organizational learning in organizational performance. *European Journal of Innovation Management*, *24*(2), 481-506.
Steyn, B. (2004). From strategy to corporate communication strategy: A conceptualisation. *Journal of communication management*, *8*(2), 168-183.
Stupak, I., Mansoor, M., & Smith, C. T. (2021). Conceptual framework for increasing legitimacy and trust of sustainability governance. *Energy, sustainability and society*, *11*(1), 1-57.
Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, *20*(3), 571-610.
Suddaby, R., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academy of Management Annals*, *11*(1), 451-478.
Swift, T. (2001). Trust, reputation and corporate accountability to stakeholders. *Business ethics: A European review*, *10*(1), 16-26.
Taeuscher, K., Bouncken, R., & Pesch, R. (2021). Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Academy of Management Journal*, *64*(1), 149-179.
Terrien, M., Scelles, N., Morrow, S., Maltese, L., & Durand, C. (2017). The win/profit maximization debate: strategic adaptation as the answer? *Sport, Business and Management: An International Journal*, *7*(2), 121-140.
Thomas, L. D., & Ritala, P. (2022). Ecosystem legitimacy emergence: A collective action view. *Journal of management*, *48*(3), 515-541.
Tian, H., & Tian, G. (2022). Corporate sustainability and trade credit financing: Evidence from environmental, social, and governance ratings. *Corporate social responsibility and environmental management*, *29*(5), 1896-1908.
Tuškej, U., Golob, U., & Podnar, K. (2013). The role of consumer–brand identification in building brand relationships. *Journal of Business Research*, *66*(1), 53-59.
Ucbasaran, D., Shepherd, D. A., Lockett, A., & Lyon, S. J. (2013). Life after business failure: The process and consequences of business failure for entrepreneurs. *Journal of management*, *39*(1), 163-202.
Vlaeminck, E. (2023). Malicious Use of Artificial Intelligence and the Threats to Corporate Reputation in International Business. In *The Palgrave Handbook of Malicious Use of AI and Psychological Security* (pp. 273-293). Springer.
Wamba-Taguimdje, S.-L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation

projects. *Business Process Management Journal*, *26*(7), 1893-1924. https://doi.org/10.1108/bpmj-10-2019-0411
Wang, H., & Wu, W. (2024). The Effects of Digital Technology Strategy Configurations on New Venture Performance. *IEEE Transactions on Engineering Management*.
Wang, Y., Xiong, M., & Olya, H. (2020). Toward an understanding of responsible artificial intelligence practices. Proceedings of the 53rd hawaii international conference on system sciences,
Wen, Y., & Holweg, M. (2023). A phenomenological perspective on AI ethical failures: The case of facial recognition technology. *AI & SOCIETY*, 1-18.
Yim, M. C. (2021). Fake, faulty, and authentic stand-taking: what determines the legitimacy of corporate social advocacy? *International Journal of Strategic Communication*, *15*(1), 60-76.
Yuan, B., & Cao, X. (2022). Do corporate social responsibility practices contribute to green innovation? The mediating role of green dynamic capability. *Technology in Society*, *68*, 101868.
Zhang, L., Yang, X., Zhu, S., & Xia, Z. (2023). Business Model Innovation and Performance of Startups: The Moderating Role of External Legitimacy. *Sustainability*, *15*(6), 5351.
Zhang, Q. (2015). Research of the corporate performance evaluation indicators based on the stakeholder theory. *Open Journal of Social Sciences*, *3*(07), 111.
Zheng, Q., Luo, Y., & Maksimov, V. (2015). Achieving legitimacy through corporate social responsibility: The case of emerging economy firms. *Journal of world business*, *50*(3), 389-403.
Zimmerman, M. A., & Zeitz, G. J. (2002). Beyond survival: Achieving new venture growth by building legitimacy. *Academy of Management Review*, *27*(3), 414-431.
Zubek, M., & Góra, M. (2021). Stuck in ENP purgatory?: an assessment of the EU's external legitimacy mechanisms. In: Oslo: ARENA Centrum Studiów Europejskich.

Appendix

Results from step 1 of the analysis.

Table A1. Correlation matrix.

	1	**2**	**3**	**4**	**5**
1. Communications	1				
2. Legitimacy	0.829	1			
3. Performance	0.708	0.789	1		
4. RAIG	0.808	0.878	0.818	1	
5. Communications x RAIG	-0.382	-0.395	-0.335	-0.383	1

Table A2. Outer loadings.

	Value
interpretability and explainability:Transparency <- Explainability	0.878
accessible: Fairness <- Accessibility and universal design	0.87
avoid unfair bias: Fairness <- Avoidance of unfair bias	0.865
measure the environmental impact of AI:Environmental and societal well-being <- Sustainable and environmentally friendly AI	0.865
established rights and policies: Data governance <- Access to data	0.863
level of risk raised:Robustness and SafetyAssess <- Fallback plan and general safety	0.859
strengthened its relationships with suppliers:External legitimacy <- External legitimacy	0.857
improve accuracy.:Robustness and SafetyAssess <- Accuracy	0.856
prevent overconfidence and overreliance:Human agency and oversight <- Human agency	0.856
assess the quality data:Data governance <- Quality and data protection	0.855
standards (ISO, IEEE) for data:Data governance Assess <- Quality and data protection	0.852
minimize societal harms:Environmental and societal well-being <- Social impact	0.852
the effects AI system have on the environment. :Environmental and societal well-being <- Sustainable and environmentally friendly AI	0.852
increase the AI applications' accuracy:Robustness and SafetyAssess <- Accuracy	0.85
ensure that AI system does not undermine human autonomy:Human agency and oversight <- Human oversight	0.848
ethical AI review board: Accountability <- Minimisation and reporting of negative impacts	0.848
traceability:Transparency <- Traceability	0.848
Identifying negative impacts of AI:Robustness and SafetyAssess <- Fallback plan and general safety	0.847
Processes and mechanisms are well documented:Transparency <- Traceability	0.847
well-documented processes :Transparency <- Traceability	0.846
Our employee satisfaction has been increased :Compared <- Customer relationship	0.845
assessed that the AI system may cause damage:Human agency and oversight <- Human oversight	0.845
enhance privacy :Data governance Assess <- Privacy and data governance	0.845

people have the necessary competence of DPP :Data governance Assess <- Access to data	0.844
explainability and interpretability:Transparency <- Explainability	0.842
training AI models :Data governance <- Privacy and data governance	0.842
AI benefit from its use:Environmental and societal well-being <- Social impact	0.839
consulted different stakeholders: Fairness <- Accessibility and universal design	0.839
diversity and representativeness of the population :Fairness <- Avoidance of unfair bias	0.839
biases during the development:Fairness <- Avoidance of unfair bias	0.838
minimize negative impacts on the environment.:Environmental and societal well-being <- Sustainable and environmentally friendly AI	0.837
redress: Accountability <- Minimisation and reporting of negative impacts	0.837
established mechanisms :Transparency <- Communication	0.836
accessible to users in need of assistive technologies :Fairness <- Accessibility and universal design	0.835
avoid unfair bias: Fairness <- Fairness	0.83
communicate about policies, and ethical guidelines: Implicit <- Implicit Communication	0.829
measures resilience:Robustness and SafetyAssess <- Resilience to attack and security	0.829
verification methods to measure system's reliability:Robustness and SafetyAssess <- Reliability and reproducibility	0.829
integrate ethical language:Explicit <- Explicit Communication	0.828
accessible: Fairness <- Fairness	0.827
We communicated company policies :Accountability <- Auditability	0.825
collaborate with organizations complying to laws and regulations:Explicit <- Explicit Communication	0.825
strengthened its relationships with suppliers:External legitimacy <- Legitimacy	0.825
communicate about awards:Implicit <- Implicit Communication	0.823
communicate about the training:Implicit <- Implicit Communication	0.822
log data:Data governance <- Access to data	0.822
AI application can be understood :Transparency <- Explainability	0.821
provide feedback :Transparency <- Communication	0.821
AI does not undermine human autonomy:Human agency and oversight <- Human oversight	0.82
processes that faciltate the assessment of algorithms:Accountability <- Auditability	0.82
reviewing and updating AI datasets :Data governance Assess <- Quality and data protection	0.82
Our profitability has increased:Compared <- Finances	0.819
minimize negative impacts on the environment.:Environmental and societal well-being <- Environmental and societal well-being	0.819
collaborate with organizations complying to laws and regulations:Explicit <- Communications	0.816
describing when an AI system fails:Robustness and SafetyAssess <- Reliability and reproducibility	0.816
communicate :Transparency <- Communication	0.815
management is accountable for its actions:Internal legitimacy <- Internal Legitimacy	0.815
strengthened its relationships with customers:External legitimacy <- External legitimacy	0.814
social impacts of the AI system are well understood:Environmental and societal well-being <- Social impact	0.812
communicate mission and values:Implicit <- Implicit Communication	0.811
ensure AI reproducibility:Robustness and SafetyAssess <- Reliability and reproducibility	0.811
assessed forms of attacks:Robustness and SafetyAssess <- Resilience to attack and security	0.81

respond to environmental changes as compared to competitors:Compared <- Competition	0.81
AI does not undermine human autonomy:Human agency and oversight <- Human agency and oversight	0.806
provides quality products to customers compared with competitors:Compared <- Competition	0.806
communicate about awards:Implicit <- Communications	0.804
established rights and policies: Data governance <- Data governance	0.803
communicate how comply with laws and regulations:Explicit <- Explicit Communication	0.8
ensure that AI system does not undermine human autonomy:Human agency and oversight <- Human agency and oversight	0.8
accessible to users in need of assistive technologies :Fairness <- Fairness	0.798
traceability:Transparency <- Transparency	0.797
comply with the legal principles:External legitimacy <- External legitimacy	0.796
verified how AI system behaves:Robustness and SafetyAssess <- Fallback plan and general safety	0.796
Our company often defeats the main competitors in the marketplace:Compared <- Competition	0.795
integrate ethical language:Explicit <- Communications	0.795
accessible: Fairness <- RAIG	0.794
considered the level of human control:Human agency and oversight <- Human agency	0.794
Our customer satisfaction has been increased:Compared <- Customer relationship	0.793
measure the environmental impact of AI:Environmental and societal well-being <- Environmental and societal well-being	0.793
Our sales growth has increased:Compared <- Finances	0.792
Processes and mechanisms are well documented:Transparency <- Transparency	0.792
Our customer loyalty has been increased:Compared <- Customer relationship	0.791
anonymized data - identification :Data governance <- Privacy and data governance	0.791
diversity and representativeness of the population :Fairness <- Fairness	0.79
interpretability and explainability:Transparency <- Transparency	0.79
diversity teams of AI development:Explicit <- Explicit Communication	0.789
formulate and develop the regulatory standards:Internal legitimacy <- Internal Legitimacy	0.789
established mechanisms :Transparency <- Transparency	0.786
monitor AI:Robustness and SafetyAssess <- Resilience to attack and security	0.786
comply with the legal principles:External legitimacy <- Legitimacy	0.785
good relationships with non-governmental organizations:External legitimacy <- External legitimacy	0.785
improve accuracy.:Robustness and SafetyAssess <- Technical robustness and safety	0.785
processes that faciltate the assessment of algorithms:Accountability <- Accountability	0.785
communicate about policies, and ethical guidelines: Implicit <- Communications	0.784
communicate mission and values:Implicit <- Communications	0.784
management is accountable for its actions:Internal legitimacy <- Legitimacy	0.784
strengthen the internal cohesion:Internal legitimacy <- Internal Legitimacy	0.784
improve operational efficiency:Internal legitimacy <- Internal Legitimacy	0.783
minimize societal harms:Environmental and societal well-being <- Environmental and societal well-being	0.782
inaccurate predictions:Robustness and SafetyAssess <- Accuracy	0.781
ethical AI review board: Accountability <- Accountability	0.78

prevent overconfidence and overreliance:Human agency and oversight <- Human agency and oversight	0.778
standards (ISO, IEEE) for data:Data governance Assess <- Data governance	0.777
AI benefit from its use:Environmental and societal well-being <- Environmental and societal well-being	0.777
consulted different stakeholders: Fairness <- Fairness	0.777
biases during the development:Fairness <- Fairness	0.776
Our company provided employment to the local economy:Compared <- Customer relationship	0.775
increase satisfaction:Internal legitimacy <- Internal Legitimacy	0.775
strengthened its relationships with customers:External legitimacy <- Legitimacy	0.775
assessed that the AI system may cause damage:Human agency and oversight <- Human agency and oversight	0.774
established mechanisms that facilitate the system's auditability:Accountability <- Auditability	0.774
Our return on investment has increased:Compared <- Finances	0.773
social impacts of the AI system are well understood:Environmental and societal well-being <- Environmental and societal well-being	0.773
the effects AI system have on the environment. :Environmental and societal well-being <- Environmental and societal well-being	0.773
communicate about the training:Implicit <- Communications	0.772
communicate how comply with laws and regulations:Explicit <- Communications	0.772
We communicated company policies :Accountability <- Accountability	0.771
redress: Accountability <- Accountability	0.771
verification methods to measure system's reliability:Robustness and SafetyAssess <- Technical robustness and safety	0.771
diversity teams of AI development:Explicit <- Communications	0.77
Our employee satisfaction has been increased :Compared <- Performance	0.769
responds rapidly to market demands compared to competitors:Compared <- Competition	0.768
Identifying negative impacts of AI:Robustness and SafetyAssess <- Technical robustness and safety	0.765
Our financial stability has increased:Compared <- Finances	0.764
training AI models :Data governance <- Data governance	0.764
Our company created new products/services:Compared <- Finances	0.762
assessed impacts of AI products on human rights:Human agency and oversight <- Human agency	0.762
level of risk raised:Robustness and SafetyAssess <- Technical robustness and safety	0.762
assessed forms of attacks:Robustness and SafetyAssess <- Technical robustness and safety	0.761
increase the AI applications' accuracy:Robustness and SafetyAssess <- Technical robustness and safety	0.761
strengthen the internal cohesion:Internal legitimacy <- Legitimacy	0.761
adapts rapidly in market demands:Compared <- Competition	0.76
avoid unfair bias: Fairness <- RAIG	0.76
considered the level of human control:Human agency and oversight <- Human agency and oversight	0.758
enhance privacy :Data governance Assess <- Data governance	0.757
traceability:Transparency <- RAIG	0.757
well-documented processes :Transparency <- Transparency	0.757
explainability and interpretability:Transparency <- Transparency	0.756
assess the quality data:Data governance <- Data governance	0.753
AI does not undermine human autonomy:Human agency and oversight <- RAIG	0.752

Our company created new products/services:Compared <- Performance	0.752
communicate practices did not align with legal expectations:Explicit <- Explicit Communication	0.752
provides quality products to customers compared with competitors:Compared <- Performance	0.752
report potential vulnerabilities:Accountability <- Minimisation and reporting of negative impacts	0.751
accessible to users in need of assistive technologies :Fairness <- RAIG	0.749
adapts rapidly in market demands:Compared <- Performance	0.749
verification methods to measure system's reliability:Robustness and SafetyAssess <- RAIG	0.748
respond to environmental changes as compared to competitors:Compared <- Performance	0.746
Our profitability has increased:Compared <- Performance	0.744
communicate :Transparency <- Transparency	0.744
increase satisfaction:Internal legitimacy <- Legitimacy	0.744
AI application can be understood :Transparency <- Transparency	0.743
people have the necessary competence of DPP :Data governance Assess <- Data governance	0.742
formulate and develop the regulatory standards:Internal legitimacy <- Legitimacy	0.74
provide feedback :Transparency <- Transparency	0.74
improve accuracy.:Robustness and SafetyAssess <- RAIG	0.739
good relationships with non-governmental organizations:External legitimacy <- Legitimacy	0.736
Our return on investment has increased:Compared <- Performance	0.735
report potential vulnerabilities:Accountability <- Accountability	0.735
Our sales growth has increased:Compared <- Performance	0.734
describing when an AI system fails:Robustness and SafetyAssess <- Technical robustness and safety	0.732
well-documented processes :Transparency <- RAIG	0.728
assessed impacts of AI products on human rights:Human agency and oversight <- Human agency and oversight	0.724
improve operational efficiency:Internal legitimacy <- Legitimacy	0.724
increase the AI applications' accuracy:Robustness and SafetyAssess <- RAIG	0.724
standards (ISO, IEEE) for data:Data governance Assess <- RAIG	0.723
Our company often defeats the main competitors in the marketplace:Compared <- Performance	0.723
communicate practices did not align with legal expectations:Explicit <- Communications	0.723
ensure that AI system does not undermine human autonomy:Human agency and oversight <- RAIG	0.722
prevent overconfidence and overreliance:Human agency and oversight <- RAIG	0.722
responds rapidly to market demands compared to competitors:Compared <- Performance	0.722
Identifying negative impacts of AI:Robustness and SafetyAssess <- RAIG	0.716
ensure AI reproducibility:Robustness and SafetyAssess <- Technical robustness and safety	0.716
Our company provided employment to the local economy:Compared <- Performance	0.715
established rights and policies: Data governance <- RAIG	0.715
log data:Data governance <- Data governance	0.713
measures resilience:Robustness and SafetyAssess <- Technical robustness and safety	0.713
biases during the development:Fairness <- RAIG	0.712
Our customer satisfaction has been increased:Compared <- Performance	0.711

anonymized data - identification :Data governance <- Data governance	0.709
training AI models :Data governance <- RAIG	0.709
Our financial stability has increased:Compared <- Performance	0.708
AI benefit from its use:Environmental and societal well-being <- RAIG	0.707
Our customer loyalty has been increased:Compared <- Performance	0.706
established mechanisms that facilitate the system's auditability:Accountability <- Accountability	0.706
measures resilience:Robustness and SafetyAssess <- RAIG	0.702
interpretability and explainability:Transparency <- RAIG	0.701
won social recognition and praise:External legitimacy <- External legitimacy	0.699

Table A3. Cronbach's alpha and Composite Reliability.

	Cronbach's alpha	**Composite reliability (rho_a)**	**Composite reliability (rho_c)**	**(AVE)**
Access to data	0.797	0.8	0.881	0.711
Accessibility and universal design	0.805	0.806	0.885	0.719
Accountability	0.852	0.853	0.89	0.575
Accuracy	0.773	0.78	0.869	0.689
Auditability	0.731	0.734	0.848	0.651
Avoidance of unfair bias	0.804	0.805	0.884	0.718
Communication	0.764	0.765	0.864	0.679
Communications	0.919	0.92	0.933	0.609
Competition	0.847	0.848	0.891	0.621
Customer relationship	0.814	0.815	0.878	0.642
Data governance	0.9	0.901	0.918	0.556
Environmental and societal well-being	0.876	0.877	0.907	0.618
Explainability	0.803	0.804	0.884	0.717
Explicit Communication	0.859	0.86	0.898	0.639
External legitimacy	0.85	0.858	0.893	0.627
Fairness	0.887	0.888	0.914	0.64
Fallback plan and general safety	0.782	0.787	0.873	0.696
Finances	0.841	0.841	0.887	0.612
Human agency	0.726	0.729	0.846	0.647
Human agency and oversight	0.866	0.867	0.899	0.599
Human oversight	0.787	0.787	0.876	0.701
Implicit Communication	0.839	0.839	0.892	0.675
Internal Legitimacy	0.849	0.849	0.892	0.623
Legitimacy	0.913	0.916	0.928	0.565
Minimisation and reporting of negative impacts	0.742	0.745	0.854	0.662
Performance	0.934	0.934	0.942	0.538
Privacy and data governance	0.767	0.769	0.866	0.683

Quality and data protection	0.796	0.799	0.88	0.71
RAIG	0.98	0.98	0.98	0.481
Reliability and reproducibility	0.754	0.755	0.859	0.671
Resilience to attack and security	0.735	0.737	0.85	0.654
Social impact	0.782	0.782	0.873	0.697
Sustainable and environmentally friendly AI	0.81	0.81	0.888	0.725
Technical robustness and safety	0.922	0.923	0.933	0.538
Traceability	0.803	0.804	0.884	0.717
Transparency	0.913	0.913	0.928	0.589

Table A4. Discriminant validity values.

Access to data	0.843																								
Accessibility and universal design	0.647	0.848																							
Accuracy	0.658	0.762	0.83																						
Auditability	0.639	0.74	0.727	0.807																					
Avoidance of unfair bias	0.714	0.78	0.657	0.708	0.848																				
Communication	0.607	0.759	0.695	0.682	0.75	0.824																			
Competition	0.659	0.641	0.66	0.581	0.62	0.52	0.788																		
Customer relationship	0.577	0.619	0.668	0.595	0.529	0.502	0.772	0.801																	
Explainability	0.6	0.7	0.7	0.6	0.6	0.7	0.6	0.5	0.8																

	29	7	08	52	69	38	05	9	47																
Explicit Communication	0.64	0.711	0.681	0.683	0.687	0.628	0.637	0.638	0.626	0.8															
External legitimacy	0.675	0.689	0.694	0.644	0.623	0.588	0.675	0.676	0.667	0.739	0.792														
Fallback plan and general safety	0.674	0.67	0.691	0.612	0.671	0.624	0.624	0.524	0.585	0.662	0.655	0.835													
Finances	0.667	0.621	0.615	0.576	0.595	0.542	0.82	0.784	0.591	0.605	0.675	0.616	0.782												
Human agency	0.681	0.705	0.73	0.684	0.649	0.636	0.696	0.634	0.642	0.71	0.774	0.687	0.669	0.805											
Human oversight	0.684	0.707	0.709	0.689	0.696	0.638	0.674	0.632	0.647	0.664	0.701	0.697	0.65	0.777	0.837										
Implicit Communication	0.649	0.686	0.668	0.646	0.642	0.602	0.657	0.638	0.621	0.858	0.716	0.654	0.625	0.681	0.645	0.821									
Internal Legitimacy	0.708	0.743	0.746	0.704	0.693	0.699	0.709	0.734	0.685	0.787	0.808	0.683	0.698	0.741	0.756	0.777	0.789								
Minimisation and reporting of negative impacts	0.558	0.732	0.659	0.755	0.71	0.677	0.612	0.559	0.635	0.68	0.629	0.596	0.553	0.687	0.688	0.695	0.683	0.813							
Privacy and data governance	0.724	0.679	0.644	0.642	0.714	0.633	0.658	0.6	0.614	0.621	0.645	0.669	0.654	0.681	0.755	0.617	0.722	0.657	0.826						
Quality and data protection	0.664	0.695	0.714	0.673	0.589	0.623	0.665	0.677	0.622	0.651	0.661	0.592	0.601	0.704	0.706	0.669	0.744	0.644	0.686	0.843					
Reliability and reproduci	0.6	0.7	0.7	0.6	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.6	0.7	0.7	0.6	0.7	0.6	0.7	0.6	0.8				

bility	89	36	43	92	15	6	89	15	25	67	72	4	85	12	38	6	43	87	09	52	19				
Resilience to attack and security	0.659	0.715	0.747	0.706	0.682	0.646	0.661	0.62	0.643	0.66	0.656	0.705	0.604	0.715	0.708	0.647	0.718	0.688	0.702	0.67	0.74	0.809			
Social impact	0.699	0.651	0.612	0.61	0.69	0.602	0.696	0.57	0.606	0.677	0.638	0.661	0.64	0.695	0.707	0.679	0.714	0.649	0.671	0.634	0.725	0.662	0.835		
Sustainable and environmentally friendly AI	0.585	0.62	0.604	0.596	0.635	0.582	0.665	0.613	0.51	0.625	0.615	0.603	0.597	0.707	0.664	0.637	0.66	0.691	0.662	0.657	0.647	0.659	0.741	0.851	
Traceability	0.663	0.781	0.688	0.681	0.756	0.79	0.614	0.575	0.735	0.635	0.625	0.676	0.601	0.658	0.685	0.628	0.71	0.693	0.689	0.67	0.709	0.664	0.675	0.633	0.847

Table A5. VIF values.

	VIF
communicate about policies, and ethical guidelines: Implicit	2.154
communicate about policies, and ethical guidelines: Implicit	1.91
standards (ISO, IEEE) for data:Data governance Assess	2.755
standards (ISO, IEEE) for data:Data governance Assess	2.1
standards (ISO, IEEE) for data:Data governance Assess	1.701
AI does not undermine human autonomy:Human agency and oversight	2.019
AI does not undermine human autonomy:Human agency and oversight	1.526
AI does not undermine human autonomy:Human agency and oversight	2.816
AI application can be understood :Transparency	1.916
AI application can be understood :Transparency	2.604
AI application can be understood :Transparency	1.604
AI benefit from its use:Environmental and societal well-being	2.915
AI benefit from its use:Environmental and societal well-being	1.875
AI benefit from its use:Environmental and societal well-being	1.676
Identifying negative impacts of AI:Robustness and SafetyAssess	2.079
Identifying negative impacts of AI:Robustness and SafetyAssess	2.627
Identifying negative impacts of AI:Robustness and SafetyAssess	1.666
Our company created new products/services:Compared	2.1
Our company created new products/services:Compared	1.614
Our company often defeats the main competitors in the marketplace:Compared	2.108
Our company often defeats the main competitors in the marketplace:Compared	1.877
Our company provided employment to the local economy:Compared	1.952

Our company provided employment to the local economy:Compared	1.62
Our customer loyalty has been increased:Compared	1.667
Our customer loyalty has been increased:Compared	1.958
Our customer satisfaction has been increased:Compared	1.934
Our customer satisfaction has been increased:Compared	1.691
Our employee satisfaction has been increased :Compared	1.951
Our employee satisfaction has been increased :Compared	2.325
Our financial stability has increased:Compared	1.656
Our financial stability has increased:Compared	1.863
Our profitability has increased:Compared	2.096
Our profitability has increased:Compared	1.953
Our return on investment has increased:Compared	2.023
Our return on investment has increased:Compared	1.676
Our sales growth has increased:Compared	2.034
Our sales growth has increased:Compared	1.782
Processes and mechanisms are well documented:Transparency	2.781
Processes and mechanisms are well documented:Transparency	2.197
Processes and mechanisms are well documented:Transparency	1.714
We communicated company policies :Accountability	2.511
We communicated company policies :Accountability	1.513
We communicated company policies :Accountability	1.74
accessible to users in need of assistive technologies :Fairness	1.955
accessible to users in need of assistive technologies :Fairness	3.1
accessible to users in need of assistive technologies :Fairness	1.657
accessible: Fairness	3.7
accessible: Fairness	2.208
accessible: Fairness	1.887
adapts rapidly in market demands:Compared	1.622
adapts rapidly in market demands:Compared	2.125
anonymized data - identification :Data governance	1.698
anonymized data - identification :Data governance	2.468
anonymized data - identification :Data governance	1.44
assess the quality data:Data governance	2.019
assess the quality data:Data governance	3.061
assess the quality data:Data governance	1.76
assessed forms of attacks:Robustness and SafetyAssess	2.137
assessed forms of attacks:Robustness and SafetyAssess	1.419
assessed forms of attacks:Robustness and SafetyAssess	2.708
assessed impacts of AI products on human rights:Human agency and oversight	2.465
assessed impacts of AI products on human rights:Human agency and oversight	1.354
assessed impacts of AI products on human rights:Human agency and oversight	1.713

assessed that the AI system may cause damage:Human agency and oversight	2.443
assessed that the AI system may cause damage:Human agency and oversight	1.746
assessed that the AI system may cause damage:Human agency and oversight	1.871
avoid unfair bias: Fairness	3.296
avoid unfair bias: Fairness	2.235
avoid unfair bias: Fairness	1.825
biases during the development:Fairness	1.845
biases during the development:Fairness	2.705
biases during the development:Fairness	1.706
collaborate with organizations complying to laws and regulations:Explicit	2.037
collaborate with organizations complying to laws and regulations:Explicit	2.564
communicate :Transparency	2.435
communicate :Transparency	1.883
communicate :Transparency	1.519
communicate about awards:Implicit	1.823
communicate about awards:Implicit	2.396
communicate about the training:Implicit	2.064
communicate about the training:Implicit	1.87
communicate how comply with laws and regulations:Explicit	1.866
communicate how comply with laws and regulations:Explicit	2.002
communicate mission and values:Implicit	2.105
communicate mission and values:Implicit	1.763
communicate practices did not align with legal expectations:Explicit	1.627
communicate practices did not align with legal expectations:Explicit	1.762
comply with the legal principles:External legitimacy	1.911
comply with the legal principles:External legitimacy	2.219
considered the level of human control:Human agency and oversight	1.448
considered the level of human control:Human agency and oversight	2.468
considered the level of human control:Human agency and oversight	1.845
consulted different stakeholders: Fairness	1.877
consulted different stakeholders: Fairness	2.481
consulted different stakeholders: Fairness	1.727
describing when an AI system fails:Robustness and SafetyAssess	1.931
describing when an AI system fails:Robustness and SafetyAssess	1.512
describing when an AI system fails:Robustness and SafetyAssess	2.452
diversity and representativeness of the population :Fairness	2.881
diversity and representativeness of the population :Fairness	1.921
diversity and representativeness of the population :Fairness	1.69
diversity teams of AI development:Explicit	1.79
diversity teams of AI development:Explicit	2.047
enhance privacy :Data governance Assess	2.484

enhance privacy :Data governance Assess	1.676
enhance privacy :Data governance Assess	1.92
ensure AI reproducibility:Robustness and SafetyAssess	2.412
ensure AI reproducibility:Robustness and SafetyAssess	1.806
ensure AI reproducibility:Robustness and SafetyAssess	1.508
ensure that AI system does not undermine human autonomy:Human agency and oversight	1.993
ensure that AI system does not undermine human autonomy:Human agency and oversight	2.655
ensure that AI system does not undermine human autonomy:Human agency and oversight	1.731
established mechanisms :Transparency	2.738
established mechanisms :Transparency	2.111
established mechanisms :Transparency	1.578
established mechanisms that facilitate the system's auditability:Accountability	2.493
established mechanisms that facilitate the system's auditability:Accountability	1.522
established mechanisms that facilitate the system's auditability:Accountability	1.375
established rights and policies: Data governance	2.274
established rights and policies: Data governance	1.767
established rights and policies: Data governance	3.181
ethical AI review board: Accountability	1.722
ethical AI review board: Accountability	1.906
ethical AI review board: Accountability	2.682
explainability and interpretability:Transparency	2.735
explainability and interpretability:Transparency	1.744
explainability and interpretability:Transparency	2.012
formulate and develop the regulatory standards:Internal legitimacy	1.953
formulate and develop the regulatory standards:Internal legitimacy	1.819
good relationships with non-governmental organizations:External legitimacy	2.108
good relationships with non-governmental organizations:External legitimacy	1.891
improve accuracy.:Robustness and SafetyAssess	3.045
improve accuracy.:Robustness and SafetyAssess	1.71
improve accuracy.:Robustness and SafetyAssess	2.42
improve operational efficiency:Internal legitimacy	1.757
improve operational efficiency:Internal legitimacy	1.869
inaccurate predictions:Robustness and SafetyAssess	1.618
inaccurate predictions:Robustness and SafetyAssess	2.365
inaccurate predictions:Robustness and SafetyAssess	1.446
increase satisfaction:Internal legitimacy	1.878
increase satisfaction:Internal legitimacy	1.702
increase the AI applications' accuracy:Robustness and SafetyAssess	2.089
increase the AI applications' accuracy:Robustness and SafetyAssess	2.984
increase the AI applications' accuracy:Robustness and SafetyAssess	1.708
integrate ethical language:Explicit	2.237

integrate ethical language:Explicit	2.077
interpretability and explainability:Transparency	1.968
interpretability and explainability:Transparency	2.424
interpretability and explainability:Transparency	3.179
level of risk raised:Robustness and SafetyAssess	1.757
level of risk raised:Robustness and SafetyAssess	2.18
level of risk raised:Robustness and SafetyAssess	2.521
log data:Data governance	1.615
log data:Data governance	1.758
log data:Data governance	2.223
management is accountable for its actions:Internal legitimacy	1.953
management is accountable for its actions:Internal legitimacy	2.214
measure the environmental impact of AI:Environmental and societal well-being	1.912
measure the environmental impact of AI:Environmental and societal well-being	2.069
measure the environmental impact of AI:Environmental and societal well-being	2.942
measures resilience:Robustness and SafetyAssess	1.849
measures resilience:Robustness and SafetyAssess	1.568
measures resilience:Robustness and SafetyAssess	2.923
minimize societal harms:Environmental and societal well-being	2.477
minimize societal harms:Environmental and societal well-being	1.755
minimize societal harms:Environmental and societal well-being	1.915
minimize negative impacts on the environment.:Environmental and societal well-being	2.887
minimize negative impacts on the environment.:Environmental and societal well-being	2.127
minimize negative impacts on the environment.:Environmental and societal well-being	1.632
monitor AI:Robustness and SafetyAssess	1.726
monitor AI:Robustness and SafetyAssess	1.423
monitor AI:Robustness and SafetyAssess	2.362
people have the necessary competence of DPP :Data governance Assess	1.898
people have the necessary competence of DPP :Data governance Assess	1.724
people have the necessary competence of DPP :Data governance Assess	2.506
prevent overconfidence and overreliance:Human agency and oversight	1.686
prevent overconfidence and overreliance:Human agency and oversight	2.975
prevent overconfidence and overreliance:Human agency and oversight	1.852
processes that faciltate the assessment of algorithms:Accountability	1.843
processes that faciltate the assessment of algorithms:Accountability	1.471
processes that faciltate the assessment of algorithms:Accountability	2.465
provide feedback :Transparency	2.352
provide feedback :Transparency	1.554
provide feedback :Transparency	1.908
provides quality products to customers compared with competitors:Compared	1.94
provides quality products to customers compared with competitors:Compared	2.22

redress: Accountability	1.874
redress: Accountability	1.683
redress: Accountability	2.959
report potential vulnerabilities:Accountability	1.293
report potential vulnerabilities:Accountability	1.66
report potential vulnerabilities:Accountability	2.562
respond to environmental changes as compared to competitors:Compared	2.113
respond to environmental changes as compared to competitors:Compared	1.896
responds rapidly to market demands compared to competitors:Compared	1.953
responds rapidly to market demands compared to competitors:Compared	1.666
reviewing and updating AI datasets :Data governance Assess	2.301
reviewing and updating AI datasets :Data governance Assess	1.7
reviewing and updating AI datasets :Data governance Assess	1.626
social impacts of the AI system are well understood:Environmental and societal well-being	1.879
social impacts of the AI system are well understood:Environmental and societal well-being	1.51
social impacts of the AI system are well understood:Environmental and societal well-being	2.743
strengthen the internal cohesion:Internal legitimacy	2.027
strengthen the internal cohesion:Internal legitimacy	1.761
strengthened its relationships with customers:External legitimacy	2.103
strengthened its relationships with customers:External legitimacy	1.961
strengthened its relationships with suppliers:External legitimacy	2.621
strengthened its relationships with suppliers:External legitimacy	2.364
the effects AI system have on the environment. :Environmental and societal well-being	1.977
the effects AI system have on the environment. :Environmental and societal well-being	1.842
the effects AI system have on the environment. :Environmental and societal well-being	2.946
traceability:Transparency	2.18
traceability:Transparency	3.262
traceability:Transparency	1.711
training AI models :Data governance	2.872
training AI models :Data governance	1.969
training AI models :Data governance	1.648
verification methods to measure system's reliability:Robustness and SafetyAssess	2.09
verification methods to measure system's reliability:Robustness and SafetyAssess	2.778
verification methods to measure system's reliability:Robustness and SafetyAssess	1.526
verified how AI system behaves:Robustness and SafetyAssess	1.706
verified how AI system behaves:Robustness and SafetyAssess	1.515
verified how AI system behaves:Robustness and SafetyAssess	2.162
well-documented processes :Transparency	1.993
well-documented processes :Transparency	1.762
well-documented processes :Transparency	2.86
won social recognition and praise:External legitimacy	1.614

won social recognition and praise:External legitimacy	1.552

Results from step 2 of the analysis.

Table A6. Outer loadings

Outer loadings	**Values**
Access to data -> RAIG	0.831
Access to data -> Data governance	0.881
Accessibility and universal design -> Fairness	0.968
Accessibility and universal design -> RAIG	0.839
Accuracy -> RAIG	0.854
Accuracy -> Technical robustness and safety	0.908
Auditability -> RAIG	0.79
Auditability -> Accountability	0.944
Avoidance of unfair bias -> RAIG	0.791
Avoidance of unfair bias -> Fairness	0.912
Communication -> RAIG	0.748
Communication -> Transparency	0.874
Competition -> Performance	0.951
Customer relationship -> Performance	0.911
Explainability -> RAIG	0.782
Explainability -> Transparency	0.913
Explicit Communication -> Communications	0.973
External legitimacy -> Legitimacy	0.918
Fallback plan and general safety -> Technical robustness and safety	0.846
Fallback plan and general safety -> RAIG	0.795
Finances -> Performance	0.916
Human agency -> Human agency and oversight	0.947
Human agency -> RAIG	0.883
Human oversight -> RAIG	0.873
Human oversight -> Human agency and oversight	0.937
Implicit Communication -> Communications	0.954
Internal Legitimacy -> Legitimacy	0.975
Minimisation and reporting of negative impacts -> Accountability	0.929
Minimisation and reporting of negative impacts -> RAIG	0.778
Privacy and data governance -> RAIG	0.839
Privacy and data governance -> Data governance	0.89
Quality and data protection -> Data governance	0.899
Quality and data protection -> RAIG	0.847
Reliability and reproducibility -> Technical robustness and safety	0.917
Reliability and reproducibility -> RAIG	0.862

Resilience to attack and security -> RAIG	0.834
Resilience to attack and security -> Technical robustness and safety	0.887
Social impact -> RAIG	0.83
Social impact -> Environmental and societal well-being	0.952
Sustainable and environmentally friendly AI -> RAIG	0.795
Sustainable and environmentally friendly AI -> Environmental and societal well-being	0.911
Traceability -> RAIG	0.805
Traceability -> Transparency	0.94

Table A7. VIF values.

	Values
Access to data	3.326
Access to data	2.364
Accessibility and universal design	2.553
Accessibility and universal design	4.808
Accuracy	3.987
Accuracy	2.871
Auditability	3.425
Auditability	2.322
Avoidance of unfair bias	4.3
Avoidance of unfair bias	2.553
Communication	3.657
Communication	3.105
Competition	3.514
Customer relationship	2.991
Explainability	3.325
Explainability	2.543
Explicit Communication	3.796
External legitimacy	2.881
Fallback plan and general safety	2.608
Fallback plan and general safety	2.992
Finances	3.679
Human agency	2.52
Human agency	3.746
Human oversight	3.886
Human oversight	2.52
Implicit Communication	3.796
Internal Legitimacy	2.881
Minimisation and reporting of negative impacts	2.322
Minimisation and reporting of negative impacts	3.528

Privacy and data governance	3.435
Privacy and data governance	2.497
Quality and data protection	2.125
Quality and data protection	3.14
Reliability and reproducibility	3.111
Reliability and reproducibility	3.879
Resilience to attack and security	3.423
Resilience to attack and security	2.914
Social impact	3.46
Social impact	2.219
Sustainable and environmentally friendly AI	3.187
Sustainable and environmentally friendly AI	2.219
Traceability	4.093
Traceability	3.075

Table A8. Inner model.

	Values
Accountability -> RAIG	3.859
RAI Communications -> Legitimacy	2.917
Data governance -> RAIG	4.817
Environmental and societal well-being -> RAIG	3.26
Human agency and oversight -> RAIG	4.81
Legitimacy -> Performance	4.441
RAIG -> Legitimacy	2.917
RAIG -> Performance	4.441
Technical robustness and safety -> RAIG	5.844
Transparency -> RAIG	4.668

Table A9. Path coefficients – Boostrap MGA.

	Difference (Europe - USA)	**1-tailed (Europe vs USA) p value**	**2-tailed (Europe vs USA) p value**
Access to data -> Data governance	-0.024	0.848	0.305
Accessibility and universal design -> Fairness	-0.017	0.77	0.461
Accountability -> RAIG	0.002	0.423	0.845
Accuracy -> Technical robustness and safety	-0.026	0.898	0.204
Auditability -> Accountability	-0.107	1	0

Avoidance of unfair bias -> Fairness	0.02	0.165	0.33
Communication -> Transparency	0.006	0.362	0.725
Communications -> Legitimacy	-0.002	0.826	0.348
Competition -> Performance	0.001	0.48	0.96
Customer relationship -> Performance	-0.012	0.739	0.521
Data governance -> RAIG	0.004	0.335	0.67
Environmental and societal well-being -> RAIG	0.002	0.396	0.791
Explainability -> Transparency	0.017	0.173	0.346
Explicit Communication -> Communications	-0.02	0.886	0.228
External legitimacy -> Legitimacy	-0.013	0.678	0.643
Fairness -> RAIG	0	0.509	0.981
Fallback plan and general safety -> Technical robustness and safety	0.017	0.21	0.419
Finances -> Performance	-0.003	0.552	0.896
Human agency -> Human agency and oversight	-0.04	0.966	0.069
Human agency and oversight -> RAIG	0.014	0.078	0.157
Human oversight -> Human agency and oversight	0.024	0.115	0.231
Implicit Communication -> Communications	0.016	0.157	0.313
Internal Legitimacy -> Legitimacy	0.008	0.371	0.741
Legitimacy -> Performance	0.001	0.286	0.572
Minimisation and reporting of negative impacts -> Accountability	0.096	0	0
Privacy and data governance -> Data governance	0.012	0.303	0.606
Quality and data protection -> Data governance	-0.025	0.885	0.229
RAIG -> Legitimacy	0	0.439	0.878
RAIG -> Performance	-0.001	0.648	0.704
Reliability and reproducibility -> Technical robustness and safety	0.001	0.464	0.928
Resilience to attack and security -> Technical robustness and safety	0.004	0.425	0.85
Social impact -> Environmental and societal well-being	-0.014	0.744	0.512
Sustainable and environmentally friendly AI -> Environmental and societal well-being	0.008	0.384	0.768
Technical robustness and safety -> RAIG	-0.004	0.637	0.726
Traceability -> Transparency	-0.019	0.815	0.37
Transparency -> RAIG	-0.005	0.679	0.642

Structuring AI resources to build an AI capability: A conceptual framework

Papagiannidis, Emmanouil, Norwegian University of Science and Technology, Trondheim, Norway, emmanouil.papagiannidis@ntnu.no

Merete Enholm, Ida, Norwegian University of Science and Technology, Trondheim, Norway,

idamen@stud.ntnu.no

Mikalef, Patrick, Norwegian University of Science and Technology, Trondheim, Norway,

patrick.mikalef@ntnu.no

Krogstie, John, Norwegian University of Science and Technology, Trondheim, Norway,

john.krogstie@ntnu.no

Abstract. Artificial Intelligence (AI) has gained traction over the past few years as the new frontier for gaining acompetitive advantage. While firms have started investing heavily in AI, there is a growing disillusionment around the value that can be generated and the process through which that can be obtained. Building on this gap, we develop a conceptual framework that builds on resource orchestration theory. The framework distinguishes between the ideation of AI capabilities and the implementation of AI capabilities and presents how activities related to resource orchestration theory are relevant in the context of AI deployments. We develop a set of propositions on the activities that underlie the main processes around resource orchestration of AI, and present a research design to actualize the research plan.

Keywords: Artificial Intelligence, Resource Orchestration, AI Capabilities, Conceptual Framework

Deploying AI Governance practices: A revelatory case study

Papagiannidis, Emmanouil, Norwegian University of Science and Technology, Trondheim, Norway, emmanouil.papagiannidis@ntnu.no

Merete Enholm, Ida, Norwegian University of Science and Technology, Trondheim, Norway,

idamen@stud.ntnu.no

Mikalef, Patrick, Norwegian University of Science and Technology, Trondheim, Norway,

patrick.mikalef@ntnu.no

Krogstie, John, Norwegian University of Science and Technology, Trondheim, Norway,

john.krogstie@ntnu.no

Abstract. In recent years artificial intelligence (AI) has been seen as a technology with the potential for significant impact in enabling firms to get an operational and competitive advantage. However, despite the use of AI, companies still face challenges and cannot quickly realize performance gains. Adding to the above, firms need to introduce robust AI systems and minimize AI risks, which places a strong emphasis on establishing appropriate AI governance practices. In this paper, we build on a single case study approach and examine how AI governance is implemented in order to facilitate the development of AI applications that are robust and do not introduce negative impacts to companies. The study contributes by exploring the main dimensions relevant to AI's governance in organizations and by uncovering the practices that underpin them.

Keywords: AI governance, Case study, Performance gains, IT governance

The dark side of AI-based decision-making: A study of B2B trading

Papagiannidis, Emmanouil, Norwegian University of Science and Technology, Trondheim, Norway, emmanouil.papagiannidis@ntnu.no

Mikalef, Patrick, Norwegian University of Science and Technology, Trondheim, Norway,

patrick.mikalef@ntnu.no

Krogstie, John, Norwegian University of Science and Technology, Trondheim, Norway,

john.krogstie@ntnu.no

Rogier van de Wetering, Open University of the Netherlands, Heerlen, Netherlands

rogier.vandewetering@ou.nl

Abstract. Artificial intelligence (AI) has recently been viewed as a technology with tremendous promise for helping businesses to gain an organizational and competitive advantage. At the same time, the deskilling of skill-intensive, in-sensitive domains at the level of society could create dangerous vulnerabilities if AI malfunctions or an adversarial attack take place. Despite the efforts to mitigate negative consequences of AI, businesses and employees continue to con-front negative dilemmas of AI, so it is essential to explore in detail the rising concerns around the negative and unintended consequences of such technologies. In this paper, we use a single case study method to investigate the dark aspects of AI in a Norwegian energy trading firm. As a contribution to the literature, the paper examines the key characteristics of AI trading in B2B organizations and suggests ways to mitigate the negative aspects of AI trading. This paper also discusses the theoretical and practical implications of the findings.

Keywords: AI dark side, AI trading, B2B, Artificial Intelligence, AI competitive advantage, AI challenges, AI decision making, Case study, Interviews

“I go wherever He leads me.”

\- Constantine the Great

ISBN 978-82-326-7960-7 (printed ver.)
ISBN 978-82-326-7959-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)