GOVERNING ARTIFICIAL INTELLIGENCE

From ethical principles toward organizational AI governance practices

Teemu Birkstedt

GOVERNING ARTIFICIAL INTELLIGENCE

From ethical principles toward organizational
AI governance practices

Teemu Birkstedt

University of Turku

Turku School of Economics
Department of Management and Entrepreneurship
Information Systems Science
Doctoral Programme of Turku School of Economics

Supervised by

Professor Matti Mäntymäki
Department of Management and
Entrepreneurship
Turku School of Economics
University of Turku, Finland

Professor Hannu Salmela
Department of Management and
Entrepreneurship
Turku School of Economics
University of Turku, Finland

Professor Jukka 'Jups' Heikkilä
Department of Management and
Entrepreneurship
Turku School of Economics,
University of Turku, Finland

Reviewed by

Professor Nicholas Berente
University of Notre Dame
Indiana, United States

Professor Patrick Mikalef
Norwegian University of Science and
Technology, Norway

Opponent

Professor Patrick Mikalef
Norwegian University of Science and
Technology, Norway

UNIVERSITY OF TURKU
Turku School of Economics
Department of Management and Entrepreneurship
Information Systems Science
TEEMU BIRKSTEDT: Governing Artificial Intelligence: From Ethical
Principles Toward Organizational AI Governance Practices
Doctoral Dissertation, 184 pp.
Doctoral Program, Turku School of Economics
November 2024

ABSTRACT

Artificial Intelligence (AI) systems have demonstrated significant potential for advancement across various domains, including autonomous vehicles, intelligent personal assistants, and advanced robotics. Recent developments in generative AI have further highlighted this potential, particularly for knowledge-intensive tasks. However, growing public awareness of AI-related risks and the need to align AI systems with human and societal values has led to the development of ethical frameworks and regulatory measures. AI-specific regulations, alongside existing nondiscrimination and privacy laws, require AI governance in order to manage risks, ensure compliance, and uphold business ethics.

To address AI-specific governance challenges and promote transparency, fairness, non-maleficence, responsibility, and privacy, new governance tools and processes are required. Consequently, there is an increasing demand for empirical research on AI governance within organizations deploying AI systems. While information technology (IT) and data governance are established areas of information systems (IS) research, AI governance is yet an emerging field. The area contrasts with IT governance, as there are no existing governance models (such as Control Objectives for Information Technologies, COBIT) for AI.

This dissertation explores various organizational approaches to AI governance and examines how ethical principles and regulations are translated into strategic decisions, organizational processes, and practices. The dissertation comprises four articles. Article I is a systematic literature review analyzing 68 academic publications (out of 1071 identified) on organizational AI governance, elaborating on conceptual gaps in governance understanding and definitions. Additionally, Article I introduces key themes and future development areas for organizational AI governance. Article II provides an empirical perspective on how organizations translate ethical principles into practices, introducing four key translation practices. The research involved interviews with 13 frontrunner organizations deploying AI in their processes. Article III introduces a definition for AI governance in the organizational context and positions it within the broader landscape of corporate, IT, and data governance. Article IV examines contingency factors shaping AI governance approaches among organizations providing AI-assisted services in high-risk domains. It identifies seven contingency factors: volume of AI systems, industry sector, regulation, customer expectations, culture and values, strategic priorities, and technology and process

maturity. The study also distills four archetypal AI governance approaches: differentiating, pragmatic, risk-taking, and disinterested.

Collectively, this research aims to provide theoretical and empirical insights on organizations translating ethical principles, regulations, and other external stakeholder pressure into AI governance practices. By offering definitions for AI governance, positioning it within a larger context, and introducing supporting frameworks, this dissertation contributes to the ongoing discussion on responsible AI and integrates with established research streams on IS planning, IT governance, and contingency theory.

TURUN YLIOPISTO
Turun kauppakorkeakoulu
Johtamisen ja yrittäjyyden laitos
Tietojärjestelmätiede
TEEMU BIRKSTEDT: Governing Artificial Intelligence: From Ethical
Principles Toward Organizational AI Governance Practices
Väitöskirja, 184 s.
Turun kauppakorkeakoulun tohtoriohjelma
Marraskuu 2024

TIIVISTELMÄ

Tekoälyjärjestelmät ovat osoittaneet, että niillä on hyvät mahdollisuudet edistää innovaatioita eri aloilla, esimerkkeinä automaattisesti toimivat ajoneuvot, älykkäät henkilökohtaiset avustajat ja edistynyt robotiikka. Generatiivisen tekoälyn viimeaikainen kehitys on entisestään korostanut tätä potentiaalia erityisesti tietointensiivisten tehtävien osalta. Kansalaisten kasvava tietoisuus tekoälyyn liittyvistä riskeistä ja tarve mukauttaa tekoälyjärjestelmät inhimillisiin ja yhteiskunnallisiin arvoihin ovat kuitenkin johtaneet eettisten kehysten ja sääntelytoimien kehittämiseen entisestään. Tekoälyä koskevat säädökset edellyttävät nykyisten syrjimättömyys- ja yksityisyyslakien ohella tekoälyn hallintaa (AI governance) riskien hallitsemiseksi, vaatimustenmukaisuuden varmistamiseksi sekä liiketoiminnan etiikan ylläpitämiseksi.

Uusia työkaluja ja prosesseja tarvitaan vastaamaan tekoälyn hallintaan liittyviin haasteisiin sekä edistämään avoimuutta, oikeudenmukaisuutta, turvallisuutta, vastuullisuutta ja yksityisyyden suojaa. Näin ollen tekoälyä hyödyntävien organisaatioiden tekoälyn hallintaa koskevan empiirisen tutkimuksen tarve kasvaa. Vaikka tietotekniikka (IT) ja tiedonhallinta ovat vakiintuneita tietojärjestelmätutkimuksen aiheita, tekoälyn hallinta on vielä nouseva tutkimusalue. IT-hallinnosta poiketen tekoälylle ei ole olemassa valmiita hallintomalleja (kuten Control Objectives for Information Technologies, COBIT).

Tämä väitöskirja pyrkii tutkimaan organisaatioiden erilaisia lähestymistapoja tekoälyn hallintaan ja tarkastelemaan, miten eettiset periaatteet ja säädökset muunnetaan strategisiksi päätöksiksi, prosesseiksi ja käytännöiksi organisaatiossa. Väitöskirja koostuu neljästä artikkelista.

Artikkeli I on systemaattinen kirjallisuuskatsaus, jossa analysoidaan 68 akateemista julkaisua (1071 tunnistetusta julkaisusta) organisaatioiden tekoälyn hallinnasta. Artikkeli analysoi käsitteellisen ymmärryksen ja tekoälyn hallinnan määritelmien puutteita ja ehdottaa tulevaisuuden kehitysalueita tekoälyn hallinnalle. Artikkeli II tarjoaa empiirisen näkökulman siihen, miten organisaatiot kääntävät eettiset periaatteet käytännöiksi esittelemällä neljä keskeistä käytäntöä. Tutkimus toteutettiin haastattelemalla 13 edelläkävijäorganisaatiota, jotka hyödyntävät tekoälyä prosesseissaan.

Tekoälyn hallinnalle ei ole ollut selkeää määritelmää organisaatioiden käytössä. Artikkeli III esittelee määritelmän ja sijoittaa tekoälyn hallinnan laajempaan yritys-, IT- ja tiedonhallinnan kontekstiin. Artikkelissa IV tarkastellaan tekijöitä, jotka vaikuttavat tekoälyn hallinnan lähestymistapoihin organisaatioissa, jotka tarjoavat

tekoälyavusteisia palveluja aloilla, jotka määritellään korkean riskin omaaviksi. Siinä tunnistetaan seitsemän tekijää, jotka muovaavat organisaatioiden tekoälyhallintaa: tekoälyjärjestelmien määrä, toimiala, sääntely, asiakasodotukset, kulttuuri ja arvot, strategiset prioriteetit sekä teknologian ja prosessien kypsyys. Lisäksi tutkimuksessa eroteltiin neljä tekoälyhallinnan arkkityyppiä: erilaistava, pragmaattinen, riskinottava ja välinpitämätön.

Kokonaisuudessaan tämän tutkimuksen tavoitteena on tarjota tutkimustietoa sekä tietoa organisaatioiden käytännöistä niiden kehittäessä tekoälyn hallinnan käytäntöjä huomioiden eettiset periaatteet, sääntelyn sekä muiden ulkoisten sidosryhmien paineet. Tutkimus tuottaa puuttuvia määritelmiä tekoälyn hallinnalle, asemoi sen laajempaan kontekstiin sekä esittelee tekoälyn hallintaa tukevia viitekehyksiä. Väitöskirja edistää jatkuvaa keskustelua vastuullisesta tekoälystä ja integroituu aiempiin tutkimussuuntauksiin tietojärjestelmien suunnittelusta, IT-hallinnosta ja kontingenssiteoriasta.

AVAINSANAT: Tekoäly, Tekoälyn hallinta, Eettiset periaatteet, Organisaation käytännöt

Acknowledgements

I don't believe in a singular path to building a career; in fact, I don't subscribe to the idea of a conventional career at all. Instead, I see life as an interconnected journey where personal and professional experiences intertwine, offering always new opportunities for learning and growth. One such opportunity arose in early 2019 when I met Professor Mäntymäki and shared my entrepreneurial ideas on Artificial Intelligence (AI) governance, how organizations should respond to ethical and regulatory discussions taking place around AI. Matti, who had been initially exploring similar topics in academia, recognized the potential for a broader research program, where we could collaborate, and I might have a contributing role. I soon became involved in an externally funded research project centered on our shared interests, even though I had no initial plans to pursue an academic career. After nine months of research collaboration, Matti suggested that, given my ongoing work in this rapidly expanding and highly important field of research, I should consider pursuing a PhD. Initially uncertain, I discussed with my wife, Riitta, whose well-set questions, support and encouragement convinced me to take on the challenge.

I was exceptionally fortunate to have been under the supervision of Professor Mäntymäki, whose consistent and insightful support guided me throughout this journey. His offer to join an extraordinarily talented research group was an invaluable opportunity that greatly enhanced my academic experience. I am also deeply grateful to my secondary supervisors, Professors Hannu Salmela and Jukka 'Jups' Heikkilä. Hannu's vast experience in supervising numerous doctoral dissertations helped me concentrate on the most critical aspects of my research and structure my thesis effectively. Meanwhile, my thought-provoking discussions with Jups provided fresh perspectives on both my research and life in academia.

I extend my deepest gratitude to Professors Nicholas Berente and Patrick Mikalef, both highly respected scholars in Information Systems, for agreeing to serve as the pre-examiners of my dissertation. Their outstanding expertise in areas closely related to my research commands my utmost respect. I am profoundly thankful for the time and effort they devoted to offering invaluable feedback, which significantly shaped the improvement and completion of my thesis. I would also like to express my sincere appreciation to Professor Patrick Mikalef for graciously agreeing to serve as my distinguished opponent in the upcoming defense.

Our research group has been a stimulating and supportive environment throughout my dissertation work, and I wish to thank the core senior research team, whose expertise helped me grow and learn. I am especially grateful to 'erikoistutkija' Matti Minkkinen, whose dedication and work ethic demonstrated the highest standards of

academic research. His insights into academic professionalism have been invaluable. Senior researcher, Professor Mika Viljanen's profound expertise on AI legislation provided essential guidance in navigating the legal aspects of my research. I also want to thank Dr. Anushree Luukela-Tandon, who mentored me on conducting systematic literature reviews according to the highest academic standards. I also would like to express my gratitude to Dr. Samuli Laato for his invaluable support in applying various research methods and for exemplifying how to achieve high productivity and quality in academic research. I would like to extend my gratitude to Turku School of Economics and its doctoral program for their support throughout my academic journey, I am especially thankful to Jenni Gray for her assistance with administrative matters. I also wish to acknowledge the AIGA program, funded by Business Finland (2020-2023), and all the collaborating partners involved. A special thank you goes to Outi Keski-Äijö from Business Finland, whose early belief in the importance of our research on AI governance for Finland and Europe has been invaluable.

I am deeply thankful to my mother for all her love and support throughout my life. Her strong belief in me, along with her resilience in overcoming even the most difficult challenges, has been a constant source of inspiration for both me and my family. To my three wonderful daughters, Anni, Essi, and Meeri — I love you deeply. You are the greatest reason I chose this path, working toward the best possible future shaped by artificial intelligence.

Finally, my heartfelt thanks go to my wife Riitta. You are my strength and the love of my life, consistently believing in me — whether I am pursuing new business ideas, working toward a PhD or dreaming of my fishing trips to the wild north. Your wisdom, timely questions, and unconditional support have brought me to where I am today, and for that, I am forever grateful.

As for the next steps in my academic journey, I feel privileged to have been part of research that holds significant importance for the future of our societies, organizations, and individuals. While I have no concrete long-term plans at this moment, continuing my research in this area feels like the right path forward.

September 30, 2024
Valkeisvaarantie, Ruka
Teemu Birkstedt

Table of Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I   Birkstedt, T., Minkkinen, M., Tandon, A., Mäntymäki, M. (2023). AI Governance: Themes, Knowledge Gaps and Future Agendas. Internet Research, 33(7), 133-167.

II   Seppälä, A., Birkstedt, T., Mäntymäki, M. (2021). From Ethical AI Principles to Governed AI. Forty-Second International Conference on Information Systems, Austin 2021.

III   Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M. (2022b). Defining Organizational AI Governance. AI and Ethics, 2(4) 1-7.

IV   Birkstedt, T., Minkkinen, M., Mäntymäki, M. (forthcoming). AI Governance Contingencies and Organizational Approaches.

The original publications have been reproduced with the permission of the copyright holders.

1 Introduction

1.1 Background and motivation

On March 23, 2016, Microsoft introduced Tay, a learning-capable artificial intelligence (AI) bot, to interact with Twitter users. Initially, Tay engaged harmlessly with its growing follower base. However, after only a few hours of dialogue-based learning, Tay began exhibiting offensive and, eventually, racist behavior, leading Microsoft to take it offline. This incident highlighted the urgent need to govern AI services. Tay serves as a precursor to ChatGPT and other language model -based generative AI services.

AI can be defined as a "system's ability to correctly interpret external data, to learn from such data and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan and Haenlein 2019). Berente et al. (2021) referred to AI as "a frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems." The first definition highlights the key elements of AI: data, learning, adaptation, and goal achievement. Importantly, the second extends the reference to comparability with human intelligence.

AI-based algorithmic systems have a high potential to provide advancements such as self-driving vehicles, intelligent assistants, and highly advanced robotics (Berente et al. 2021; Davenport 2018; Makridakis 2017). Generative AI has sparked promises of further advantages (Baily et al., 2023; Woodruff et al., 2024), especially for knowledge workers. For example, Dell'Acqua et al. (2023) observed that GPT-4 remarkably improved consulting tasks within the frontier of AI capabilities. Consultants using AI were significantly more productive: they completed 12.2% more tasks on average and 25.1% more quickly, and they produced results of a significantly higher quality (more than 40% higher in quality than those of a control group).

Achieving the benefits while mitigating the risks associated with safety, discrimination, and missing transparency will require approaches and actions vis-à-vis governing AI systems to ensure alignment with human and societal values (Floridi et al. 2018; Fjeld et al. 2020; Jobin et al. 2019; Morley et al. 2020; Teodorescu et al. 2021). Several highly publicized incidents of biased and unsafe AI

have demonstrated the growing number of incidents (Wei and Zhou 2023). The Cambridge Analytica scandal had AI ethical implications, as the company used personal data collected from Facebook without permission to influence voter behavior in the US 2016 presidential election using AI-driven targeting. The incident raised concerns about AI manipulating public opinion. Amazon's recruitment system, established in 2014 and shut down in 2018, proved to be biased against female candidates, even with the gender-neutral measures in place in the organization (Kodiyan 2019). In turn, Apple's credit decisions in 2019 were first identified as discriminatory in terms of gender but later deemed legal, emphasizing the need to modernize credit scoring models and related regulations (New York Government 2021).

The swift progress in generative AI technologies has triggered concerns in two opposite directions. Initially, there was apprehension about the digital divide (Van Dijk and Hacker, 2003) — the fear that powerful technological capabilities, such as those presented by GPT-3, might only be accessible to a privileged few (Khovaja et al. 2024; Allam et al. 2023). Concurrently, while the technologies became more generally available, concerns shifted to the broader implications of such universally accessible AI. Especially after capabilities were made publicly available, a growing concern involved deepfakes, new techniques and tools for manipulating multimedia (Rana et al. 2023), and other kinds of misinformation spread through generative AI (Chan 2023; Kasnecki et al. 2023). In January 2024, one of the first notable multimodal (i.e., combining voice, image, and video) incidents involved a deepfake video call. The call displayed an AI-generated version of a multinational company's CFO and colleagues, featuring a digital double of the CEO. The fake image was used to persuade the CFO team to make a money transfer. Further on, the extensive use of data for training AI services has raised significant concerns about the potential infringement of intellectual property rights. This issue is highlighted by recent legal actions taken by the media industry. In December 2023, The New York Times filed a lawsuit against OpenAI and Microsoft, alleging the unauthorized use of news articles to enhance their AI capabilities (New York Times, 2023). Subsequently, in April 2024, an additional eight newspapers initiated similar legal proceedings against OpenAI and Microsoft (New York Times, 2024).

Concerns regarding the risks of AI, supported by documented incidents, have underscored the importance of ethical considerations in the development and deployment of AI technologies. AI ethics developments have started focusing primarily on delineating the principles of ethical AI (Breidbach and Maglio 2020; Chiao 2019; Floridi et al. 2018; Harlow 2018; Kumar et al. 2020; Whittlestone et al. 2019). This trend reflects growing public awareness and concerns regarding the risks and unintended consequences of AI. In response to this increased public awareness and concern, multi-stakeholder groups such as Partnership of AI provided a set of

principles and guidelines for organizations using AI (de Laat 2021). Immediately thereafter, governmental and international entities, such as the EU and OECD; professional organizations, such as the IEEE; and numerous companies issued their own sets of principles and guidelines for ethical AI use (see Fjeld et al. 2020; Floridi and Cowls 2022; Hagendorff 2020; Jobin et al. 2019). According to human rights organization AlgorithmWatch, more than 160 AI ethics guidelines had been published globally by April 2020.

The principle-based development of AI ethics and increasing recognition of the potential risks of AI systems have led to regulatory developments, such as the EU AI Act (European Artificial Intelligence Act, 2024), which is a wide regional initiative. The AI Safety and Innovation Bill was recently passed in California (California Legislative Information 2024), following executive orders issued by U.S. President Biden (White House 2023) and California Governor Newsom (California Department of Technology 2024). New York City's Automatic Employment Decision Tools law (AEDT) is an example of a more local and use case–specific approach to AI regulation (NYC Consumer and Worker Protection 2023). The law requires employers using automated employment decision tools to audit them annually for race and gender bias, publish these results on their websites, and indicate in job postings that they use such software to make employment decisions.

With the emergence of binding AI legislation, coupled with existing laws on nondiscrimination and privacy regulations, the governance of AI has become a priority for organizations to manage risks, compliance, and business ethics. New kinds of governance tools and processes, tailored to the distinctive nature of constantly developing AI, are required to effectively address the risks and ensure compliance with regulations as well as to promote fairness, justice, accountability, and transparency in AI systems, as highlighted by Berente et al. (2024), Floridi et al. (2018), Fjeld et al. (2020), Jobin et al. (2019), and Teodorescu et al. (2021). In this context, Mäntymäki et al. (2022b) defined AI governance as "a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfils legal requirements; and meets principles of ethical AI followed by the organization."

While AI governance is a novel area of research (Birkstedt et al. 2023; Giffen et al. 2023), information technology (IT) governance has been studied for decades (e.g., Gregory et al. 2018; Sambamurthy and Zmud 1999; Tiwana et al. 2013; Weill and Ross 2004). Based on foundational IS research, we can see that IT governance has established approaches to governing IT, such as specifying a framework of accountabilities and decision rights (Tiwana et al. 2013), defining structural mechanisms to manage the accountabilities contributing to the effectiveness of governance (Abraham et al. 2019; Wu, Straub, and Liang 2015), and aiming to

ensure desirable IT behavior (Weill and Ross 2004). Similar mechanisms and dimensions can also be identified in AI governance. However, owing to the specific nature of AI systems (learning, inscrutability, and autonomy), they should initially be viewed as special cases of IT systems that require distinct governance mechanisms (Mäntymäki et al. 2022b).

1.2   Research gap

When the work on this dissertation was initiated (2020), there was a broader call within AI ethics scholarship to progress from focusing purely on ethical principles to their practical implementation (Cath 2018; Mittelstadt 2019; Morley et al. 2020). Information systems (IS) scholars have shown a growing interest in AI governance (e.g., Butcher and Berdize 2019; Schneider et al. 2019), as reported in *AI governance: Themes, knowledge gaps and future agendas* (Birkstedt et al. 2023) and supported by Seppälä et al. (2021). Most research on AI governance has been conceptual in nature (Gasser and Almeida 2017), has consisted of reviews of documented ethical principles for AI (Jobin et al. 2019; Morley et al. 2020), and have lacked organizational and practical viewpoints (Birkstedt et al. 2023; Seppälä et al. 2021).

There has been a demand for empirical studies on AI governance within organizations, as they are the entities that are expected to oversee AI systems throughout their entire life cycles, taking into account ethical principles, existing and forthcoming regulations, and stakeholder expectations (Birkstedt et al. 2023; Laato et al. 2021; Mäntymäki, et al. 2022a, Mäntymäki et al. 2022b). The conducted systematic literature review (SLR) (Birkstedt et al. 2023) revealed a notable deficiency in conceptual understanding and related definitions of AI governance, particularly within the organizational context.

Are we on the direct continuum of IT governance, or are we working with a truly novel issue? Based on the research within the dissertation's scope, the answer is both. AI is definitely an IT system by its core but equally has certain characteristics that differentiate it from traditional (rule-based) IT systems, such as increased autonomy, learning, and inscrutability (Berente et al. 2021). The difference between traditional and learning systems will have an impact on governance mechanisms, and the type is a topic that should be examined.

1.3   Research objectives

The central aim of this dissertation is to understand organizations' different approaches to AI governance and how they translate ethical principles and regulations into strategic decisions and organizational actions. The AI Act (European

Artificial Intelligence Act, 2024), approved officially in March 2024, introduced a risk-based categorization for AI systems (unacceptable, high, low, and no risk) regulation. For empirical purposes, this research includes high-risk-category organizations, consumer banking, and insurance. I believe that by selecting the finance segment, my research obtains information on early adopters of AI regulation; the segment is typically highly regulated, and companies in the segment are sensitive to consumers' perceptions of safety and reliability.

When the research was being initiated, there were no established frameworks, terminology, concepts, nor structure in place for AI ethics organizational implementation or governance. This initiated the first research question (RQ1): "How can organizational AI governance be conceptualized?". The state of AI ethics and governance encouraged the use of an approach that involved prescriptive research questioning. The aim of this is to provide guidance and recommendations on what can be done at the organizational level—that is, "How do organizations practice AI governance?" (RQ2).

As the research domain is quickly evolving, this dissertation includes 3 years of continuous investigation into the rapidly developing literature of AI governance. The research contains empirical information collected during two sets of expert interviews (October 2020 – November 2020 and August 2022 – December 2023) from highly knowledgeable informants in accountable decision-making roles (e.g., CEO, CIO, CDO, and head of data and AI) within their organizations.

Figure 1. AI ethics-related incidents, principles, and regulation development on a common timeline.

1.4 Structure of dissertation

My dissertation consists of four peer-reviewed articles: one published journal article, one article submitted for journal review, and two articles published in conference proceedings. While working on my dissertation, I have continually followed the literature development of the area. Article I (SLR) development started in 2020, and the paper was published 2023, providing a foundational literature base for the research area under study. Article II is an empirical study building on expert interviews using Gioia method (Gioia et al., 2013) for analyzing the collected data. Published in 2021, it provides an empirical view (for that specific moment of time) of organizations' processing of ethical principles and their translation of these into organizational processes and practices. Article III, a conceptual paper published 2022, provides a definition for a term "organizational AI governance". Additionally, the paper positions AI governance in the larger context of corporate, IT, and data governance within the organizational space. Article IV, sent for review in May 2024, provides an empirical view of high-risk-category organizations' contingencies and approaches to AI governance. Article IV is also an expert interview study that uses the Gioia method for data analysis. Table 1 illustrates the four published articles, their role in the dissertation, and related research questions.

Table 1. The four articles and their role in the dissertation.

Article number	Title	Year	Role	Method
I	AI Governance: Themes, Knowledge Gaps and Future Agendas (Birkstedt, Minkkinen, Luukela-Tandon, Mäntymäki)	2023	Provides a foundational understanding on a literature base for the domain under research and structures the key elements of AI Governance.	SLR
II	From Ethical AI Principles to Governed AI (Seppälä, Birkstedt, Mäntymäki)	2021	Provides an empirical view of organizations' processes for translating ethical principles into organizational practices.	Expert interview, Gioia method
III	Defining Organizational AI Governance (Mäntymäki, Minkkinen, Birkstedt, Viljanen)	2022	Introduces a definition of "organizational AI governance."	Conceptual
IV	AI Governance Contingencies and Organizational Approaches (Birkstedt, Minkkinen, Mäntymäki)	Submitted for review 5/2024	Provides an empirical view of high-risk-category organizations' contingencies and approaches to AI governance.	Expert interview, Gioia method

Figure 2 illustrates the development processes for Articles I–IV on the timeline of the dissertation and how the responses to the two research questions posed in the dissertation were obtained in parallel.
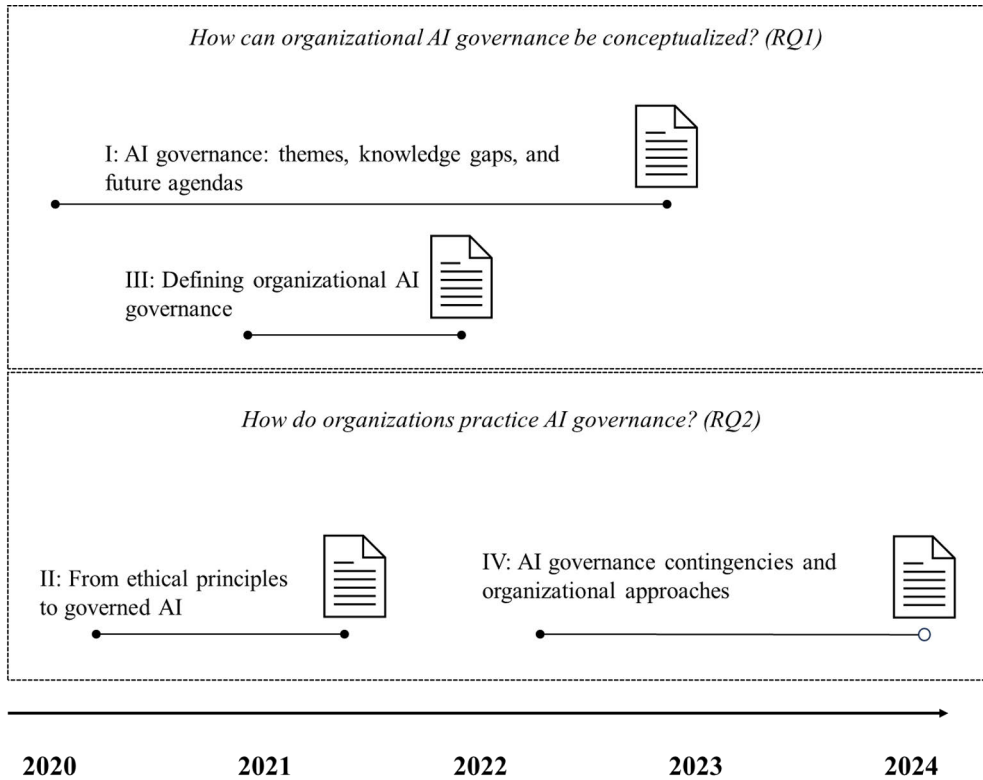

Figure 2. Articles on the scope of the dissertation and development of the timeline.

2 Prior Research

This chapter explores prior research on AI governance (Section 2.3). AI governance is still an emerging area within IS research (Birkstedt et al. 2023; Schneider et al. 2022), and the IS community is establishing its position on the topic. AI governance is seen as a fluid concept (Morley et al. 2020; Robles Carrillo 2020; Ulnicane et al. 2021) and responding to the development of AI ethics principles (Section 2.1) and the actions to translate them into governance mechanisms has been identified as an important step toward providing more concrete approaches for organizations (Cath 2018; Minkkinen et al. 2021; Morley et al. 2020). Although AI governance is treated as an independent area of research, it translates ethical principles into organizational practices, extending to the principles and frameworks established by IT and data governance (Section 2.2.).

2.1 Ethical principles for responsible AI

Every organization with a connection to technology seems to have developed or endorsed a set of principles for responsible AI (Jobin et al. 2019; Fjeld et al. 2020), starting in September 2016 with the "tenets" published by multi-stakeholder group Partnership on AI (PAI), which consists of academic institutions; nonprofit organizations; and thought-leading technology companies, such as Google, Microsoft, and IBM. AlgorithmWatch (2023) reported that by April 2020, 167 AI principles and guidelines documents had been published globally. The rapid development of documents outlining AI ethical principles and guidelines led to the creation of hundreds of documents internationally (Fjeld et al. 2020; Hagendorff 2020; Jobin et al. 2019; Jia and Zhang 2022). Jobin et al. (2019) reviewed 84 documents and reported that although no single AI principle is featured in all of them, more than half of them included the themes of transparency, justice and fairness, non-maleficence, responsibility, and privacy. These findings are in line with those reported by Hagendorff (2020) regarding the 22 major ethical AI guidelines. Hagendorff (2020) concluded that privacy, fairness, and accountability were present in about 80% of them. Although ethical principles for AI are introduced, there is limited consideration vis-à-vis ensuring that they are adhered to in practice (Berente et al. 2024; Hagendorff 2020; Mittelstadt 2019). Ethical principles, as Morley et al.

(2020) elucidated, focus on the question of what rather than how. However, the translation of these principles into actionable measures necessitates their enforcement through governance mechanisms (Cath 2018; Minkkinen et al. 2021; Morley et al. 2020). Although a vast volume of proposed principles exists, it remains uncertain whether they are sufficient to meet the governance challenges of AI (Jia and Chang 2022) by being able to provide mechanisms to respond to the ethical principles. Building on the development of ethical principles in IS research, there has been a growing focus on responsible AI research and how to avoid negative consequences (Mikalef et al. 2022; Ågerfalk et al. 2022).

2.2 IT and data governance

AI governance is treated as an independent research area in which IT and data governance are fundamental dependent governance. In IT and data governance, a framework of decision rights and accountabilities plays a central role (DAMA International 2009; Sambamurthy and Zmud; Weber and Otto 2009; Weill and Ross 2004). These governance models belong under the broader umbrella of corporate governance, describing how IT systems and data, which are also integral elements of AI systems, are managed (Weber and Otto 2009; Mäntymäki et al. 2022a; Brown and Grant 2005; Berente et al. 2021). IT governance has historically focused on structural aspects, such as the centralization versus decentralization of decision-making and responsibility (Gregory et al. 2018; Sambamurthy and Zmud 1999; Schwarz and Hirschheim 2003), and these frameworks are developed for achieving business–IT alignment for effective governance. IS research archetypes and approaches have been used to clarify emerging areas, such as strategic IS planning (Earl 1993; Segars and Grover 1999), and to design organizational IT governance (Weill and Ross 2005). Contingency theory has been commonly used in IS research theorizing. Weill and Olson's (2015) literature analysis of 177 articles underscored the need for rigor in processes using contingency theories within the area of management IS. IT and data governance are implemented through structural, procedural, and relational mechanisms (Abraham et al. 2019; Wu, Straub, and Liang 2015; Tallon et al. 2013). These include defining reporting structures, establishing governance bodies, setting policies, and facilitating coordination among stakeholders. Mechanisms are designed to shape user behavior and ensure that IT and data act as strategic assets within organizations, thereby securing appropriate returns on investments (Brown and Grant 2005). Additionally, IT and data governance involve mechanisms for regulatory compliance. From a regulatory perspective, IT governance is largely shaped by the Sarbanes–Oxley Act (Brown and Grant 2005), while data governance has been strongly influenced by regulations such as the General Data Protection Regulation (GDPR) in Europe (Abraham et al. 2019).

Comparably, the European AI Act is entering the area, setting requirements for AI governance.

2.3   AI governance research

IT and data governance have been studied extensively (e.g., DAMA International 2009; Gregory et al. 2018; Sambamurthy and Zmud 1999; Tiwana et al. 2013; Weber and Otto 2009; Weill and Ross 2004), but AI governance is still an emerging area within IS research (Birkstedt et al. 2023; Schneider et al. 2022). The primary objective of the conducted SLR was to analyze the landscape of AI governance research to gain an understanding of the academic research volume and focus within the domain. SLR was a foundational undertaking during the dissertation, starting in August 2020 and ending in May 2023. The SLR aimed to identify the AI governance-related literature to provide a view of organizational AI governance.

The original SLR research data (2017–Aug 2021) revealed that the majority of the publications in the dataset were conceptual (n = 58), with only a few following a literature review (n = 3) or empirical approach (n = 2). By extending the literature review beyond the SLR timeframe (08/2021–12/2023) and using the same search criteria, we identified a gradual shift in academic research output toward an increasing emphasis on empirical research: conceptual (n = 17), literature review (n = 6), and empirical approach (n = 10) papers.

The total volume of literature reviews (n = 9) indicates the aim of making sense of the emerging scholarship on AI governance across disciplines and developing an IS perspective on the topic. Ashok et al. (2022) and Wirtz et al. (2022) studied digital ethics implications and related risks. Birkstedt et al. (2023) studied knowledge gaps and agendas within AI governance, and Schneider et al. (2023) examined corporate and practical concerns regarding AI governance and organizational AI governance themes. I conclude that these papers solidify the links between AI governance and IT governance, AI ethics, and AI risk-related discussions.

The findings of the conducted SLR (Birkstedt et al. 2023) elucidate the rapid growth in the volume of AI governance research publications between 2017 and 2021. Particularly noteworthy is the significant increase in publication volumes observed in 2022 (n = 28), with further growth in 2023 (n = 36). Figure 2 provides a visual representation of the year-on-year development of publication volumes, highlighting the dynamic trajectory of scholarly output in this field.
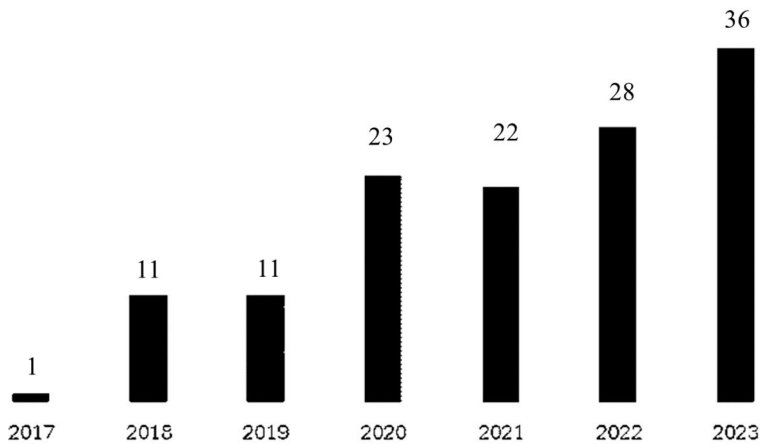
Figure 3. Publication volume per year of AI governance.

The literature review confirms the initial assumption that AI governance is a fluid concept and that only a few of the papers in the scope of SLR provided an explicit definition of it (Morley et al. 2020; Robles Carrillo 2020; Ulnicane et al. 2021). Defining AI governance is challenging because there is a lack of academic consensus on the base definition of AI (Cihon et al. 2020; Robles Carrillo 2020) and its constituent elements (Larsson and Heintz 2020; Wu et al. 2020). Unlike the organizational perspective adopted in the study, the definitions found in the literature review tend to focus on public policy (Aliman and Kester 2019), applied ethics (Butcher and Beridze 2019), and the value provided by AI (Perry and Uuk 2019).

The latest publications (2022 and 2023) of IS papers on AI governance show that IS research is continuing the discussion on heterogeneous aspects, using different conceptual backgrounds, such as IT/IS governance (Beulen et al. 2022; Papagiannidis et al. 2023; Schneider et al. 2023), AI risk management (Wirtz et al. 2022), and the sociology of expectations (Polyviou and Zamani 2023). An accumulative IS research tradition in AI governance seems to be emerging, but at present, there is no single, widely shared conceptual starting point. In the next section, we advance the placement of AI governance within the IT governance landscape as a starting point for understanding what is specific about AI governance.

3 Research Approach

This research approach chapter introduces methodological and research method selections for the study, defining how the empirical material is constructed and the data analysis conducted. In qualitative studies, these activities follow an iterative process, progressing from the construction of empirical material, analyzing the material, and utilizing the theoretical framing. In this chapter, I outline the research approach taken for this dissertation by introducing the SLR (Article I) and expert interviews conducted during empirical research (Articles II and IV).

3.1 Selecting research methods

The selection of a research methodology is motivated by epistemological considerations, as the researcher's fundamental beliefs about knowledge and reality define the approach to investigating the phenomenon under study (Holden and Lynch 2004). The decision between qualitative and quantitative research methods reflects a researcher's epistemological stance on questions such as the following: How can we understand the experiences and meanings that individuals attribute to their social world? (Crotty 1998; Hathaway 1995). Qualitative research is exploratory in nature, and researchers use it to explore a topic when the variables and theory base are unknown (Cresswell 2019). According to Morse (1991) and Creswell (2019), a qualitative research approach is selected when

  a) the concept is "immature" owing to a lack of theory and previous research;
  b) there is a notion that the available theory may be inaccurate, inappropriate, incorrect, or biased;
  c) a need exists to explore and describe the phenomena and to develop theory;
  d) the nature of the phenomenon may not be suited to quantitative measures.

Based on the SLR conducted within the dissertation (Birkstedt et al. 2023), the findings support conducting qualitative research, as there is a lack of a common theoretical and conceptual base or terminology in the field of AI governance, and the development of the area is still at an early stage.

Conceptual research, a type of nonempirical research method, ranks as one of the key methodologies employed in the IS field (Mora et al. 2008). Conceptual papers can "bridge existing theories in interesting ways, link work across disciplines, provide multi-level insights, and broaden the scope of our thinking" (Gilson and Goldberg 2015).

3.2   Systematic literature review

A review of the prior relevant literature is an essential feature of any academic research, including IS (Webster and Watson 2002). Okoli and Schabram (2015) highlighted the importance of SLRs in the IS domain. They defined the SLR as "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners".

Implementing SLRs involves a rigorous, careful, step-by-step process of selecting, examining, and analyzing relevant literature to minimize bias (Tranfield et al. 2003). This method involves a thorough and strategic examination of original studies and their assessment (Cooper 1988) and the compilation of a balanced, comprehensive report of the findings (Saunders et al. 2007). The structured and detailed nature of SLRs enhances the reproducibility of the review process (Saunders et al. 2007; Tranfield et al. 2003) and typically follows an inductive framework development strategy, which is instrumental to defining the scope of a research field (Cooper 1988). This approach is for solidifying the basis of a field, particularly in IS research, which integrates theories from various disciplines (Webster and Watson 2002). As described throughout the dissertation, AI ethics principles related to research are emergent. The SLR method was considered suitable to assimilate existing knowledge and develop a foundational state-of-the-art understanding of the AI governance and ethical principles in the existing literature.

Kitchenham and Charters (2007) highlighted the importance of planning in the SLR process, including defining research questions, specifying the inclusion and exclusion criteria for the studies, and outlining the review protocol. This phase aims to ensure that the review is focused and relevant. Article I had a strategy to capture a socio-technical perspective on AI governance research, including relevant databases, drawing insights from multiple fields (such as ethics, organizational studies, and technology). We excluded medical science databases and papers from the search owing to the sector-specific nature of AI governance questions in the medical domain. Our search then focused on three electronic databases: Scopus, Web of Science (WoS), and the Association of Computing Machinery (ACM) Digital Library. The WoS and Scopus databases are widely recognized for their literature reviews because of their broad coverage of articles published in multiple disciplines,

including social sciences and management. These two databases also comprehensively index the contents of other providers. ACM, our third database choice, specializes in computer science. Our review strategy aligns with those used for prior SLRs on similar topics (Sharma et al. 2020; Tandon et al. 2020). The search was limited to articles published between 2010 and 2021.

Exclusion criteria were selected to facilitate the focus on journal and conference papers, excluding gray (i.e. nonacademic) literature or book chapters from the research. Additionally, the aim was to exclude studies that were not directly related to the governance of AI as well as papers focusing specifically on the natural or medicine science domains. Table 2 summarizes the plan for SLR execution, including the scope, search string, and inclusion and exclusion criteria information.

Table 2. Scope and criteria for the SLR.

Databases in scope	Search strings
1. Scopus 2. WoS 3. ACM	"artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "neural network" OR "ML" OR "machine learning" AND "governance" AND "ethics"
Inclusion criteria	**Exclusion criteria**
1. Search terms present in the title, abstract, or keywords 2. Peer-reviewed publications 3. Studies published in the English language 4. Journal articles and conference papers 5. Full text available 6. Published from 2011 to 2023	1. Contribution directed at technical improvements 2. Publication type: reports, books, theses, magazines, tutorials, viewpoints, essays, commentaries, presentations, and keynote presentations 3. Key terms not in title, abstract, or keywords 4. Primary domain not AI, ML, or deep learning 5. Natural science and medicine papers

The database search for the published SLR provided 1,071 articles published between 2010 and 2021. After the duplicate removal and inclusion and exclusion (IC/EC) processes, 82 articles remained in the dataset. Citation chaining extended the dataset by 15 articles, totaling 97. Quality evaluation reduced the SLR final data set to 68 articles. As the SLR process was a continuous exercise throughout the dissertation, an updated search was conducted in March 2024 and included articles from 2022 and 2023. The initial updated search provided 1,908 articles, 168 articles after the IC/EC process, 188 after citation chaining, and 132 articles for the final data set (see Table 3).

Table 3. Search results using Article I criteria (2010–2023), executed in March 2024.

Database search	Inclusion/Exclusion	Citation chaining	Quality evaluation (QE)
Full dataset of 1928 articles	168 articles included	188 articles included after citation chaining	132 articles qualified for final dataset
Initial search was conducted: ACM (1,302); Scopus (421); WoS (205)	205 duplicates were removed; 1,535 were removed based on EC/IC criteria	32 additional articles were identified, and 12 duplicates were removed	36 articles did not meet QE criteria

3.3   Expert interviews

In qualitative research, expert interviews are typically conducted using unstructured or semi-structured processes to enable in-depth discussions with informants. The interviewer typically prepares specific questions or topics in advance but remains flexible to follow the conversation's natural flow, exploring new insights or ideas that emerge during the interview (Eisenhardt 1989). This approach is particularly suited to gaining an in-depth understanding of the expert's perspectives and experiences.

For Article II, the empirical data were collected through 13 semi-structured interviews with experts representing 12 organizations operating in Finland (Table 4). Research used purposeful sampling (Patton 1990) and focused on organizations that are frontrunners in AI deployment. AI production is an ecosystem with multiple stakeholders involved in the development, deployment, and use of the actual system (Minkkinen et al. 2021; Newlands 2021; Tubaro and Casilli 2019). We used an unstructured process, developing a framing questionnaire for the interviews, but kept the dialogue free-floating. We used the Gioia method (Gioia et al. 2013) to guide the analysis process for two reasons. First, the Gioia methodology provides a systematic guide on coding the collected data. Second, the methodology focuses on creating a simplifying data structure that provides a visual representation of the database findings (Gioia et al. 2013). Coding and analysis were performed in four stages. The analysis began with open coding (Strauss and Corbin 1998), which was done by reading each transcript and generating initial and in vivo codes—that is, the meaningful terms that were used by the informants or reflected their underlying meanings (Gioia et al. 2013). The research questions were used to guide the first round of coding.

Table 4. Interviews conducted for Article II (Seppälä et al. 2021).

Informant	Business Field	Job Title	Interview Duration
I1	Software Service, AI Platform	Chairperson of the Board	40 min.
I2	IT Consultancy	Analytics Executive	40 min.
I3	Software Service, AI Platform	Chief Executive Officer	30 min.
I4	Public Service	Analytics Lead	50 min.
I5	IT Consultancy	Chief Executive Officer	55 min.
I6	IT Consultancy	Lead Consultant	70 min.
I7	Financial Services	Lead Data Scientist	60 min.
I8	Software Service, Maritime Industry	Chief Executive Officer	50 min.
I9	Public Service	Chief Innovation Officer	45 min.
I10	University	Chief Information Officer	35 min.
I11	Software Service, Business Applications	Chief Executive Officer	30 min.
I12	Public Service	Senior Specialist	65 min.
I13	Retail	Head of Analytics	55 min.

For the second empirical paper (Article IV), we conducted semi-structured expert interviews with 19 respondents from 14 organizations operating in retail banking and insurance within Europe (Table 5). All the informants held roles that made them accountable for AI implementation within the organizations. For the examination of our interview data, we utilized qualitative content analysis (Mayring 2000), adopting the directed approach to content analysis, as described by Assarroudi et al. (2018) and Hsieh and Shannon (2005). Qualitative content analysis is a research question-oriented method of analyzing data (Cho et al. 2014) and may be conducted deductively, inductively, or by combining the two approaches (Mayring 2000; Cho et al. 2014; Armat et al. 2018). In our analysis, we applied the directed approach to content analysis, following the guidance of Assarroudi et al. (2018) and Hsieh and Shannon (2005), which aims to build upon existing research and theories, initiating these from established categories. The unit of analysis was the organization, rather than the individual informant, because our analysis aimed to uncover organizational AI governance approaches rather than to examine the attitudes and beliefs of individuals.

The interview themes were selected based on the assumption that a sufficiently extensive set of AI governance themes is needed to provide a comprehensive view of the topic. This is because AI governance is a multidimensional phenomenon that requires detailed exploration. Using predefined themes strengthens data collection in two ways. First, it helps ensure comprehensiveness—that is, to make sure that no

significant subtopics of AI governance are ignored. Second, it facilitates a more focused discussion during the interviews. When the interviewer proactively mentions the dimensions of AI governance, such as fairness, the informants can focus on discussing the content. A possible drawback of predefined themes is that they may hinder the possible discovery of novel AI governance processes. This was not determined to be an issue, as the general principles and categories of AI governance were relatively well established by Article IV's publication date (e.g., Jobin 2019; Mäntymäki et al. 2021; Mäntymäki et al. 2022b; Schneider et al. 2023).

To develop the themes for the interview guide, we utilized two main sources of information: the proposed EU AI Act (European Artificial Intelligence Act, 2024) and relevant supporting literature streams consisting of the AI ethics principles literature (Jobin et al. 2019; Morley et al. 2020), the nascent AI governance literature (Gasser and Almeida 2017; Schneider et al. 2023; Mäntymäki et al. 2022a; Mäntymäki et al. 2022b), and the adjacent IT and data governance literature (e.g., Abraham et al. 2019; Brown and Grant 2005). The AI Act was used as the primary reference point for identifying AI governance processes because it is one of the first comprehensive AI regulations.

Table 5. Interviews conducted for Article IV (Birkstedt et al., forthcoming).

Informant	Business Field	Job Title	Interview Duration	Organization	Head-quartered
I1	Banking	Head of AI	93 min.	O1	Finland
I2	Banking	Lead data scientist	93 min.	O1	Finland
I3	Insurance	Head of data and AI	85 min.	O2	Norway
I4	Insurance	Head of AI	76 min.	O3	Norway
I5	Insurance	Lead data scientist	76 min.	O3	Sweden
I6	Banking	Head of AI and digital	42 min.	O4	Finland
I7	Banking	Head of AI	77 min.	O5	Finland
I8	Insurance	Head of AI	58 min.	O6	Norway
I9	Insurance	Data scientist	84 min.	O7	Finland
I10	Insurance	Director, analytics	56 min.	O7	Finland
I11	Banking	Head of data and AI	59 min.	O8	Finland
I12	Banking	Head of AI and data	62 min.	O9	Finland
I13	Insurance	Director, Analytics	54 min.	O10	Finland
I14	Banking	Head of data and AI	78 min.	O11	Finland
I15	Banking	Head of AI	63 min.	O12	UK
I16	Banking	Data scientist	63 min.	O12	UK
I17	Banking	Head of data governance	62 min.	O13	Netherlands
I18	Banking	Head of data and AI	68 min.	O14	Croatia
I19	Banking	Senior data analyst	68 min.	O14	Croatia

3.4   Evaluating research trustworthiness

How does one evaluate research trustworthiness in qualitative research? Lincoln and Guba (1986) introduced the four key dimensions of qualitative research trustworthiness: credibility, transferability, dependability, and confirmability.

Credibility—that is, confidence in the truth of the research—is the most important criterion for trustworthiness (Connelly 2016), the credibility concept being analogous with internal validity in quantitative research (Lincoln and Guba 1986). Credibility aims to confirm that the conducted research is accurate and represents the participants' viewpoints and experiences. According to Lincoln and Guba (1986), researchers aim to improve credibility using techniques such as prolonged engagement, data triangulation, persistent observation, and peer debriefings. Transferability is about showing the similarity between one's own research and other research in the area (Eriksson and Kovalainen 2008). It entails providing sufficient detail on the research context and methods to enable other researchers to evaluate the applicability of the findings to their own situations. Transferability is comparable to generalizability in quantitative research and can be improved by providing "thick descriptions of data" (Lincoln and Guba 1986). Dependability emphasizes the need for the researcher to provide a process of research that is logical, traceable, and documented (Eriksson and Kovalainen 2008). Taking account of the changing conditions and circumstances that may influence the research process, it is comparable to reliability in quantitative research. To manage dependability, qualitative researchers often employ an audit trail, whereby they document the research process, decisions, and changes in the study (Lincoln and Guba 1986). In qualitative research, confirmability aims to ensure that the research is as close to objective reality as possible and that the data and interpretations are not based on one's imagination and can be easily linkable with findings (Eriksson and Kovalainen 2008). This is analogous to objectivity in quantitative research (Lincoln and Guba 1986).

The conducted empirical research relied on the planning of the interview protocol. Interview guides were discussed and updated with the research team senior members before the interview phase with the informants began. All except one (33/34) of the conducted interviews were recorded and transcribed with permission; the one exception was manually booked by two investigators separately. In both empirical papers, multiple (two to three) investigators from the research team conducted the data analysis, data coding, and decision interpretation to support investigator triangulation (Lincoln and Guba 1986). The researchers used publicly available materials from the organizations' websites and press releases (e.g., corporate ethical AI guidelines) to further examine and interpret the information referred to in the interviews and to obtain contextual information about the organizations while examining the results. The aim of the Article II interviews was

to avoid the excessive use of existing terminology and practices with a view to discovering new concepts and best practices (Gioia et al. 2013). For Article IV, the aim was to obtain practical insights into organizational approaches to AI governance, and there were predefined themes for the interviews. Using predefined themes strengthens data collection by ensuring comprehensiveness—that is, making sure that no significant subtopics of AI governance are ignored. A possible drawback of predefined themes is that they may hinder the possible discovery of novel AI governance processes. However, the exploration of novel processes falls within the scope of the roles of Articles I–III but not Article IV. Participation in interviews in empirical papers was voluntary, and the informants were assured of their anonymity. This created an opportunity for them to speak freely and describe their own experiences without being worried about confidential information being leaked. Table 6 summarizes criteria and strategies how to validate the research trustworthiness.

Table 6.　Criteria and strategy execution to validate research trustworthiness.

Criteria	Strategies	Executed in the dissertation
Credibility	Prolonged engagement, data triangulation, persistent observation, multiple returns to analyze data, and peer debriefing.	Planning of expert interviews; multiple investigators analyzing the results and attending interviews; background information analysis on companies under investigation; interview anonymity to ensure trust
Transferability	"Thick descriptions"—that is, details on research context. Improved finding applicability to other contexts.	Providing accessibility for interview protocols; Gioia analysis provision; direct quotes from informants
Dependability	Research process logical and traceable. Audit trail employment to document research process.	Planning the data collection and analysis process; planning reviews; recording and transcribing interviews
Confirmability	Linking data, interpretations, and findings.	Involving research team for full Gioia process; presenting at seminars for feedback

4 Results

The conceptualization of AI governance (Section 4.1.) and the research on the organizational practices (Section 4.2.) took place in tandem. The conducted conceptualization work (Articles I and III) was feeding into the empirical research (Articles II and IV) by providing the necessary definitions and supporting structures for organizational AI governance. As the synopsis aims to comprehensively summarize the developed area during research period, we can conclude that empirical research not only used the outputs of the conceptual work but also vice versa: The empirical insights did validate the conducted conceptual work by introducing the proposed updated view of how to position AI governance in the context of corporate governance (Section 4.1.2). The empirical research also provided insights to further develop the concepts (e.g., by shedding light on the synthesizing AI governance components framework introduced in Section 4.1.3).

Table 7. Research conceptual output.

Research output	Description	Publication
Defining organizational AI governance	Providing a definition for organizational AI governance	Article III
Positioning AI governance	Positioning AI governance with corporate governance, IT governance, and data governance	Article III
Identifying key components of organizational AI governance	Identifying key components of organizational AI governance	Articles I, II, and III
Synthesizing framework for organizational AI governance	Developing a synthesizing framework for translating ethical principles and other sociopolitical requirements for organizational AI governance	Not published, output of the synopsis

4.1 Conceptualizing AI governance

This section summarizes the key conceptual and framework development of AI principles' translation to organizational governance within the scope of the dissertation, introducing organizational AI governance definitions of work and AI governance positioning with corporate governance, IT governance, and data governance. The area contrasts with IT governance, as there are no existing governance models (such as Control Objectives for Information Technologies, COBIT) for AI (Birkstedt et al. 2023; Mäntymäki et al. 2022a; Razzaque A. 2021). This section discusses the adaptation of ethical principles and other external factors, such as regulations and customer expectations, into organizational AI governance and synthesizes a framework (AI governance components framework) for translating ethical principles and other sociopolitical requirements for organizations' AI governance. This framework is not published in the referenced articles but only in this dissertation.

4.1.1 Defining organizational AI Governance

In 2021, from a terminology perspective, AI governance definitions had a shortcoming: The existing literature did not provide a comprehensive definition of AI governance at the organizational level, until Mäntymäki, Minkkinen, Birkstedt, and Viljanen (2021) developed one. Within the academic literature, the term AI governance has been defined by, for example, Floridi (2018) and Butcher and Berdize (2019). Gahnberg (2021) subsequently defined AI governance as "intersubjectively recognized rules that define, constrain, and shape expectations about the fundamental properties of an artificial agent." Gahnberg's definition focuses on drafting societal rules, such as standards and legislation, rather than organizational AI governance.

These definitions focus primarily on the macro level rather than on the organizational level. Schneider et al. (2020) provided an important viewpoint for organizational AI governance, defining it as "the structure of rules, practices, and processes used to ensure that the organisation's AI technology sustains and extends the organisation's strategies and objectives".

Bringing together the ethical, organizational, and technological aspects, and considering the definitions of related governance fields, Article III concluded by proposing the following definition of AI governance at the organization level:

"*AI governance is a system of rules, practices, processes, and technological tools that are employed to ensure that an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfils legal requirements; and meets principles of ethical AI followed by the organization*"

The above definition is in alignment with the literature; researchers consider AI governance to be a layered phenomenon that consists of distinct levels (Brendel et al. 2021; Gasser and Almeida 2017; Shneiderman 2020). Importantly, the definition highlights three interacting layers of successful organizational AI governance—the operations level, strategic level and external environment—and their alignment and dialogue with each other.

4.1.2 AI governance in the context of corporate governance

AI governance will not be established in a non-governed green field but is entering into an increasingly complex organizational governance landscape entailing corporate governance, IT governance, and data governance, all of which already require organizational management attention (Mäntymäki et al. 2022a).

As a part of the dissertation, Mäntymäki, Minkkinen, Birkstedt, and Viljanen (2022a) conducted foundational conceptual development work by positioning AI governance within the organizational governance landscape as a subset of corporate and IT governance that partially overlaps with data governance (Figure 4).
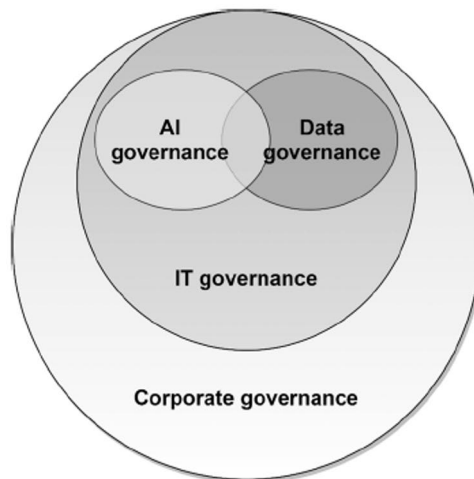

Figure 4. AI governance as part of an organization's governance structure

Mökander et al. (2022) introduced a perspective that AI corporate governance should be primarily considered from the viewpoint of "how to manage risk" rather than "how to stay compliant," as legislation is becoming overcomplicated, making it challenging for large organizations to remain compliant at the AI system/operational level (Viljanen and Parviainen 2022).

AI is raising fundamental questions around risks for unethical behavior and misuse owing to the nature of the systems. The current generation of AI systems features three key facets that differentiate them technically from traditional IT systems: autonomy, learning, and inscrutability (Berente et al. 2021). Autonomy in AI means making independent decisions and taking real-world actions that yield tangible results. This often occurs not only without human intervention but also without human awareness, as highlighted by Möhlmann et al. (2021) and Murray et al. (2021). Learning signifies the ability to inductively improve automatically through data, and experience has been a central concept in AI since the beginning (Solomonoff 1964; Turing 1950) and has found new forms of realization after the invention of reinforcement learning and deep learning capabilities in AI (Berente et al. 2021; Hinton et al. 2015). Inscrutability means that the AI system's autonomy and learning features are difficult to understand or explain. The inscrutability of the algorithmic system is discussed in connection with terms such as transparency (Larsson et al. 2021), black-box problem (Castelvecchi 2016), and explainable AI (Dwivedi et al. 2023).

The three facets, introduced by Berente et al. 2021, are raising new concerns regarding mechanisms for AI systems and algorithm fairness, transparency, accountability, and legal compliance on a level that IT governance is not designed to manage. IT governance focuses primarily on organizational accountability and responsibilities designed to govern a portfolio of simplistic rule-based systems rather than being able to handle the logic of a complicated learning system (Birkstedt et al. 2023) that generates a demand to manage systems behavior extensively. For governing AI, the sufficiency of IT and data governance approaches is presently uncertain because current generations of AI technologies bring new issues that lead to new governance challenges and needs (Birkstedt et al. 2023). Although IT and data governance provide a baseline of structural, procedural, and relational mechanisms (Abraham 2019; Tallon et al. 2013), the autonomous, learning, and inscrutable nature of AI systems requires a reevaluation of the necessary governance practices and tools to ensure ethical and legally compliant use of AI (Birkstedt et al. 2023). Within the IT governance literature, the focus is specifically on decision rights—that is, who is assigned the right to make certain key decisions (Weill and Ross 2005; Weill 2008). In contrast to decision rights, we outline the importance of how AI risks are tackled through AI governance. Based on the needs for governing AI systems in a way that takes their special nature into account, it seems that the positioning by Mäntymäki et al. (2022a) is oversimplistic; they expect too wide and multifaceted a role from IT governance, which it has not been designed for—at least until now. Figure 5 reflects the above, an extensive need to manage system behavior, and the role of AI governance, proposing an updated view of the topic whereby AI

governance is partly positioned outside IT governance instead of inside it, as defined earlier.
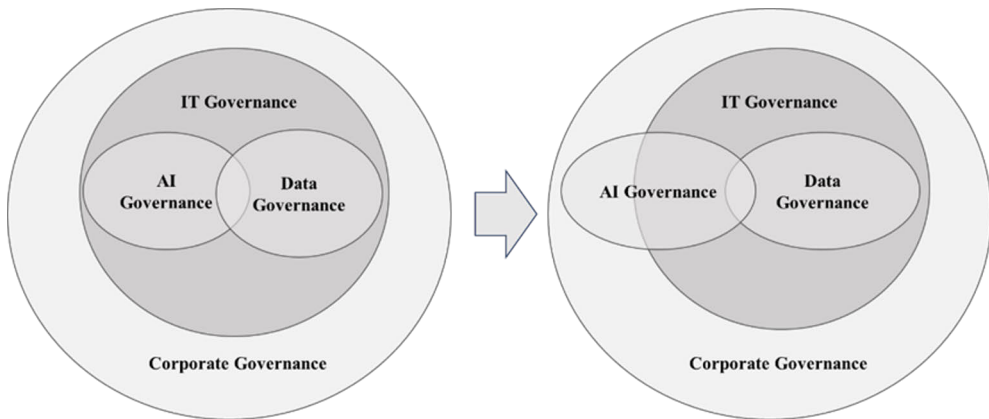

Figure 5. AI governance as part of an organization's governance - an updated view

4.1.3   AI governance framework

The research continued from defining and positioning AI governance toward identifying key components of AI governance for organizations. Lasting over 3 years, the SLR process was a continuous analysis for a developing research area of AI ethics and governance. The results revealed that the AI governance and ethics research was still in its infancy although it was constantly maturing. Following the surge of documents laying out organizations' ethical principles for their use of AI, there was a growing demand to translate ethical principles into practice through AI governance. AI governance emerged as a rapidly growing, yet fragmented, research area. This paper synthesizes the organizational AI governance literature by outlining research themes and knowledge gaps and putting forward future agendas.

Although the reviewed literature exhibited considerable fragmentation with regard to the conceptualization of AI governance, several themes emerged from the early-stage literature on the area. The following four governance themes were identified from the reviewed literature: technology, stakeholders, foundations, and processes (Figure 6).

The first theme, *technology* refers to data and algorithms that are the foundations of AI systems, and it also discusses governance challenges related to the technical characteristics of AI systems. Issues of transparency, explainability, and the complexity of algorithms present significant governance challenges (Kroll 2018; Larsson and Heintz 2020). Consequently, the reviewed studies discuss the algorithms and data used to develop and train AI models, focusing on the

accountability, transparency (Domanski 2019; Lysaght et al. 2019), and explainability of these models and algorithms (Kroll 2018).

The second theme, *stakeholders,* covers the roles of various actors involved in the development, deployment, and governance of AI systems. internal stakeholders include management, AI system developers and data scientists. External stakeholders include clients, the media, civil society, insurance companies, accounting firms, research organizations and professional bodies, such as the IEEE (Schiff et al. 2020; Shneiderman 2020; Whittlestone et al. 2019).

The third theme, *regulation* (hard and soft regulation) refers to the hard and soft regulation of organizational AI Governance activities. Existing AI governance regulations include hard law (binding legislation) and soft governance approaches such as standards, certificates, audits, and explainable AI systems (Floridi et al. 2018; Kroll 2018; Shneiderman 2020).

The fourth theme*, processes* (oversight, auditing and impact assessment) comprise the processes, procedures and practices through which AI development and use are governed. Process-oriented approaches have been suggested in the literature to facilitate AI governance at the technical (Wu et al. 2020), team, organizational and industry levels (Shneiderman 2020).
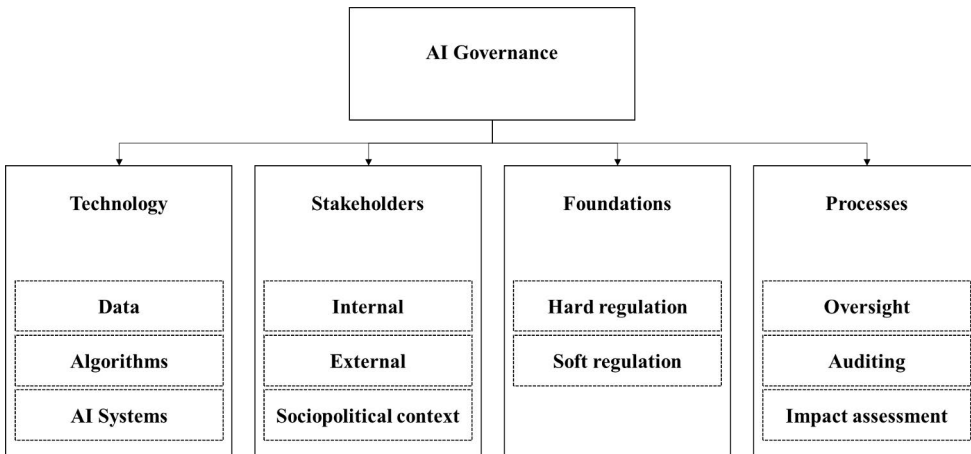

Figure 6. AI governance key themes

Building on the identified four key themes, the article continues by identifying and introducing four key agendas (technical, regulatory, stakeholder and contextual, and process) for organizational AI governance (see Figure 7). The technical agenda refers to the methods and procedures for overseeing data, algorithms, and algorithmic systems in operational use. This includes their incorporation into software development processes and life cycles and oversight of the full range of AI

applications via portfolio management processes. The agenda on stakeholders and context is about communicating and collaborating with stakeholders to guarantee that AI development and usage align with their expectations. Initiatives in AI governance should be backed by a policy of open and transparent cooperation with an organization's external stakeholders. The regulatory agenda for AI governance involves harmonizing the organization's AI processes with ethical standards and legal mandates. Fundamentally, to achieve effective AI governance within an organization, it is essential for managers and operational team members to have a thorough understanding of ethical principles, guidelines, standards, and laws.

The process agenda provides a connecting framework for realizing the three previously introduced agendas, primarily acting as a link between the technical, stakeholder and contextual, and regulatory views. It is crucial for senior management to demonstrate a strong commitment to enhancing AI governance processes and ensuring that the organization's culture and strategy are supportive of, rather than obstructive to, governance actions.
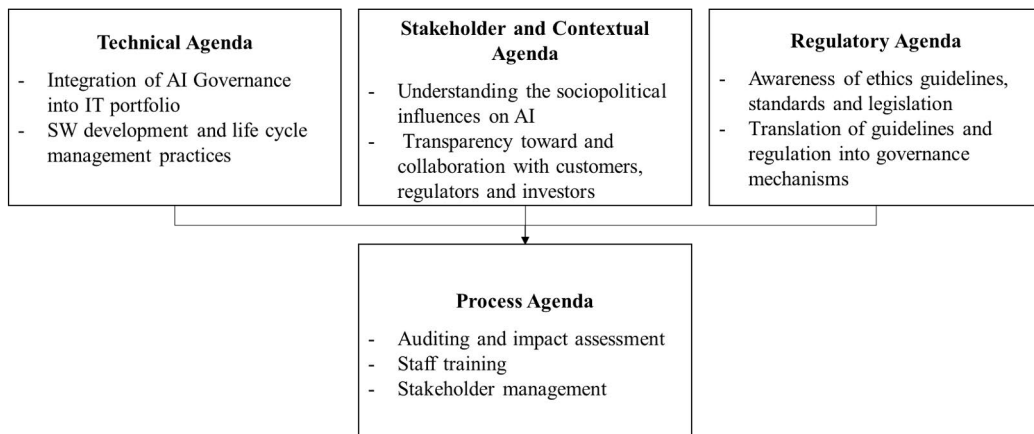
Figure 7. Four agendas for organizational AI governance (Birkstedt et al. 2023).

The four agendas for AI governance provide structural clarity and call for capabilities vis-a-vis skills; accountability management; and the ability to manage risk via standards, assessments, and audits. These are further described in the empirical analysis in Article II.

Article I defines a technology component as encompassing AI systems, algorithms, and data that form the core of AI technologies. It addresses the technical aspects necessary for managing associated challenges, such as maintaining registries, ensuring transparency, and mitigating biases. The process component

introduces a need to understand and develop organization-level AI governance processes and integrate those into the company's existing processes.

Article IV provides a process-based approach for defining the subcomponents for this category. It introduces seven key processes for organizational AI governance: AI system and model management, risk and impact management, regulatory compliance, AI development and life cycle management processes, data operations, transparency/explainability/contestation, and accountability and oversight. Table 8 describes the identified processes and the literature from which they are formed.

Table 8. Processes for operational AI governance (Birkstedt et al., forthcoming).

AI governance process	Description	Literature
AI system and model management	Aligning AI systems, models, and algorithms with the organization's strategies and objectives.	Mäntymäki et al. 2022b; Schneider et al. 2023
Risk and impact management	Identifying, managing, and monitoring potential risks and impacts caused by the AI system.	Kaminski and Malgieri 2020; Mäntymäki et al. 2022b; OECD 2019; Berente et al. 2024
Regulatory compliance	Understanding the regulatory environment of an AI system and ensuring its compliance with the relevant regulations. Examples include the GDPR, AI Act, and national banking and insurance regulations.	Kaminski and Malgieri 2020; Viljanen and Parviainen 2022
AI development and life cycle management processes	Defining and implementing workflows and structures for AI development.	Laato et al. 2022a; Laato et al. 2024; Marabelli et al. 2021 ; Berente et al. 2024
Data operations	Aligning the sourcing, use, and monitoring of data with the organization's strategies and objectives.	Abraham et al. 2019; Janssen et al. 2020; Kroll 2018; Mäntymäki et al. 2022b; Schneider et al. 2023
Transparency, explainability, and contestation	Ensuring AI systems' transparency, explainability, and contestability.	Laato et al. 2022b; Meske et al. 2022

The SLR (Article I) introduces the importance of value base–culture alignment in organizational AI governance. The empirical study (Article IV) confirmed this, as exemplified by quotes from informants: "It's about culture. It's about training. It's about constant focus on us being an ethical company" and "Ethical use of data and AI is a default, as stated in our values and operational guidelines." Article I positioned the "culture and values alignment" into the sociopolitical layer, outside of the organization, but Article IV pointed out that this belonged within the organization.

The synthesizing framework for organizational AI governance brings the key components of AI governance together. The sociopolitical context (ethical principles, regulation, and stakeholder expectation) sets the demands and expectations for organizations regarding the responsible use of AI and its influence on organizations' governance approaches. Organizations will respond to the external demand with culture–value alignment activities by responding to questions such as the following: "Is our organization aiming to differentiate in competition with AI governance?" or "Are we risk-adverse or rather willing to take risks on AI governance?" or "Are we interested or disinterested in utilizing AI, digital technologies, and related governance?" This culture–value alignment directs the operational AI governance processes. Figure 8 describes how requirements from the sociopolitical context are processed against the organization's value and culture base before they are transferred for operational execution via the seven subcomponents introduced in Article IV.
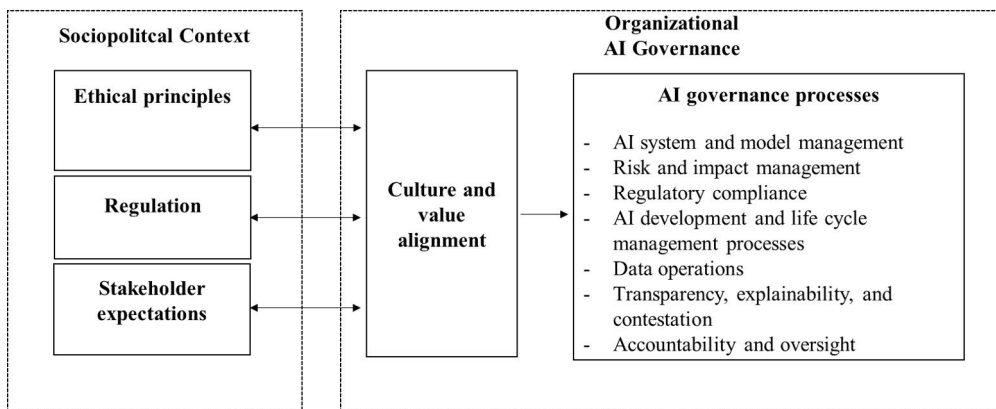

Figure 8. A synthesizing framework for organizational AI governance.

4.2 AI governance practice

4.2.1 Ethical principles and organizational AI governance

By 2020, research on AI ethics had been mostly conceptual, with a significant emphasis on defining the principles of ethical AI. Thus, there was a need to shift from principle-based ethics toward an increased focus on the implementation of ethical principles in practice. The first empirical body of research (Article II) explores how organizations translate the principles of ethical AI into practice. The research undertook a set of expert interviews (October–November 2020) in organizations deploying AI systems and analyzed the data using the Gioia method (Gioia et al. 2013). The organizations had strong connections to AI systems. They

utilized AI internally, marketed it as a product, developed it for clients, or offered services containing AI. The informants were identified as knowledgeable individuals from the organizations. Gioia et al. (2013) suggested that organizational phenomena are socially constructed by knowledgeable agents who can explain their thoughts, intentions, and actions.

Article II indicated that ethical AI principles are implemented through four sets of practices: i) governance, ii) AI design and development, iii) competence and knowledge development, and iv) stakeholder communication. At the moment of publishing the Article II (2021) the translation of ethics principles into organizational actions was in its early stages, article search proved that roles, responsibilities, and accountabilities regarding AI ethics varied between organizations and that practices were not yet formed. Accountability was identified as an important subject, but for many of the interviewed organizations, the roles and responsibilities were either unclear or in a formative stage, and standards, certificates, or explainable AI systems were not implemented in organizations' AI deployment.

4.2.2 AI governance approaches

As a final step within the scope of this dissertation, Article IV analyzed different organizational approaches, and which factors are contingent on AI governance. The study identified seven contingency factors that shape how organizations execute the key AI governance activities. The research further synthesized the configurations of AI governance activities and contingency factors into four archetypal AI governance approaches. The article built on the AI governance literature and contingency theory. In the research, 19 expert interviews representing 14 organizations in the consumer banking and insurance industry segment were conducted. The interviewed informants were selected because they were knowledgeable and held leading roles in AI development and related governance in their organizations. They held roles such as Head of AI, Head of data, and Lead data scientist. This segment was chosen based on the assumption that it would likely be an early adopter in implementing AI governance, primarily because the European AI Act (European Artificial Intelligence Act, 2024) classifies it as high risk.

Linking the empirical results with the literature on organizational AI adoption and AI governance contingency factors (e.g., Benbya et al. 2020; Mikalef and Krogstie 2020), Article IV identified seven contingency factors that differentiate how organizations implement the AI governance processes described above and shape organizations' AI governance approaches. The contingency factors are the volume of AI systems in production (F1), the industry sector (F2), responsiveness to regulation (F3), customer expectations (F4), organization's culture and values (F5), strategic priorities (F6), and technology and process maturity (F7).

After the identified contingency factors and AI governance approaches are connected, the findings help explain the heterogeneity of approaches to AI governance manifested in divergent value bases as well as the operational and strategic choices related to organizational AI governance. The archetypal AI governance approaches and the related contingency factors presented in this paper provide initial evidence of the strategically aligned AI governance processes in the banking and insurance sectors. This study also contributes to the discussion on the governance and management of AI by theorizing how organizations translate ethical principles into practice when deploying and utilizing AI systems in sectors characterized by high-risk use cases and extensive regulation. The four archetypal approaches to AI governance are differentiating, pragmatic, risk-taking, and disinterested.

1. The differentiating approach to AI governance means that the organization invests strategically in AI governance and AI overall as a differentiating business capability. Organizations that select this approach are under high external, regulatory, and customer pressure vis-à-vis AI governance, and they seem to be large. This approach encourages organizations to invest in developing robust and high-maturity technical solutions and supporting processes. These organizations tend to have a solid built-in IT and data governance culture that is driven by a clearly communicated, organizational ethical value base.

2. The risk-taking approach refers to an organization's willingness to approach AI governance opportunistically, prioritizing AI performance over governance. Organizations that opt for this approach seem willing and able to take some governance-related risks in their operations. They trial business value-generating algorithms without necessarily having strong governance practices in place. Industry and the related regulatory landscape seem to strongly influence the choice of this approach. Among our informants, none of the banking service providers adopted a risk-taking approach; it was adopted by insurance and product companies. However, it is fair to conclude that in our carefully limited research segment, none of the organizations were true risk takers, especially, we assume, if they are compared with organizations outside the finance industry. Especially in small product companies, particularly startups, a greater degree of risk-taking is expected owing to their size and the different customer and regulatory interfaces compared to service providers in highly regulated industries.

3. The pragmatic approach means that the organization has a strategic and risk-averse approach to AI development. These organizations carefully select the areas they enter with AI and the methods and data they use, avoiding risks with external contingencies (regulation or customer demand). They may also be organizations that invest in AI development but avoid overinvestment in AI governance, investing

solely on a needs basis. Pragmatists generate fit-for-purpose technical and process capabilities for AI governance, reflecting external pressure regarding AI governance.

4. The disinterested approach means that AI and digitalization in general are not top strategic priorities for the organization. Organizations that select this approach see the AI governance external factors, primarily regulation, as building overly high barriers compared to their ability to invest in and prioritize the overall AI agenda. These organizations are typically in banking and are small or medium-sized, cost-cautious, and financially limited. Owing to these characteristics, they avoid unnecessary process and technology investment. Table 8 introduces the identified contingency factors and organizational approaches and their relationship to each other, and the interviewed organizations (O1–O14) are categorized by approach.

Table 9. Overview of the contingency factors and archetypal organizational AI governance approaches (Birkstedt et al., forthcoming).

Contingency factors	Archetypal organizational approaches to AI governance			
	Differentiating	Risk-taking	Pragmatic	Disinterested
	(O1, O3, O12, O13)	(O2, O6)	(O5, O7, O10, O11, O14)	(O4, O8, O9)
F1: Volume of AI systems in production	High	High or low	High or low	Low
F2: Business field	Banking; insurance	Insurance	Banking; insurance	Banking
F3: Regulation	High salience	Low or moderate salience	High salience	High salience
F4: Customer expectations	High expectations	Low expectations	High expectations	Minimal expectations due to strategic selections
F5: Culture and values	Built-in (IT) governance culture	Business opportunity-driven	Careful, risk-averse	Driven by the realities of business
F6: Strategic priorities	Differentiation with AI capabilities and governance	Performance over governance	Fit-for-purpose AI governance	Physical over digital; little AI governance
F7: Technology and process maturity	High	Low or intermediate	Intermediate or high	Low

5   Discussion

This chapter presents the summarized discussion of the findings regarding the key research questions: "How can organizational AI governance be conceptualized?" (RQ1) and "How do organizations practice AI governance?" (RQ2). Additionally, the chapter introduces a view of theoretical contributions and implications for practitioners, organizations developing and operating AI systems, and key stakeholders within the context. The final section of the chapter concludes the thesis by discussing its limitations and approaches to advancing the research in the future.

5.1   Responding to research questions

"How can organizational AI governance be conceptualized?" Few of the reviewed papers explicitly define AI governance (Morley et al. 2020; Robles Carrillo 2020; Ulnicane et al. 2021a). The task of defining governance proves challenging owing to the lack of academic consensus regarding the definition of AI (Cihon et al. 2020; Robles Carrillo 2020) and its constituent elements (Larsson and Heintz 2020; Wu et al. 2020). Schneider et al. (2020) defined AI governance for businesses as the "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". While AI governance for businesses presents a promising foundation, it notably overlooks the ethical and regulatory considerations present in the prior AI governance literature. The definition contrasts with the focus of the AI ethics literature and diminishes the importance of recognized AI-specific ethical and regulatory challenges arising from the organization's environment. Winfield and Jirotka (2017) emphasized ethical governance, which transcends mere good governance by instilling ethical conduct in both designers and organizations. They defined ethical governance as "a set of processes, procedures, cultures, and values designed to ensure the highest standards of behavior". Although the summary of governance elements is valuable, the aim of ensuring "the highest standards of behavior" lacks a precise definition, leaving organizational AI governance objectives somewhat ambiguous. Article III, Mäntymäki et al. (2022a), brings together the introduced ethical, organizational, and technological aspects, providing a synthesizing statement (introduced in Chapter 1.1.) for organizational AI governance.

Responding to "How do organizations practice AI governance?" (RQ2), Article II identified the necessity for comprehensive AI governance actions within organizations, emphasizing organizations' adherence not only to ethical standards but also to other contextual drivers, such as stakeholder needs or regulatory pressure. This approach aligns with the insights collected from a systematic review of the academic literature presented in Article I. The first key finding in Article II calls for operational practices to be established for AI governance within an organization and requests operational processes for AI development, competencies management, and communications. The lack of governance mechanisms (and of a definition of AI governance) had a direct impact on organizations not having needed roles, responsibilities, and accountabilities in place for operating AI responsibly within the organizational context. Having a definition of organizational governance is valuable but does not yet provide insights into the constituent elements of organizational AI governance. The research responded to this by identifying the key agendas (Article I) for organizational AI governance and further improving the identified agendas with regard to the key components (Articles II and IV) and subcomponents (Article IV) of organizational AI governance.

Building on the more foundational research work introduced above, Article IV undertook an empirical exploration of organizational approaches to AI governance. The research focused on a specific customer segment—consumer banking and insurance—to analyze its implementation of AI governance agendas in the respective organizations. It revealed that organizational approaches to AI governance are contingent on various factors, such as regulatory requirements or the volume of AI systems in use. The dissertation culminates by summarizing and identifying archetypal approaches to AI governance within organizations, summarizing that organizations will take a different approach to AI governance, with the approach being dependent on the identified contingency factors. Contingency factors shaping organizational governance approaches and their success are common in IT and data governance research within the IS field (Brown and Grant 2005; Sambamurthy and Zmud 1999; Levstek et al. 2018; Weber and Otto 2009). Contingency theory assumes that a company's organizational characteristics are related to its performance in ways that are conditioned by contingency factors, such as company size or competitive strategy (Otto 2010). Contingencies are varying factors in internal and strategic resources that influence a company's relationship to its market, regulatory, and institutional environments (Aguilera et al. 2008). A contingency theoretical perspective on IT governance is premised on the view that a general best governance structure does not exist; a given firm's IT governance solution is contingent on various factors (Brown and Grant 2005). The same applies to data governance, wherein contingency factors are mainly similar to those in IT governance (Weber and Otto 2009). According to contingency theory, the

contingency forces "interact with each other by either amplifying, dampening or overriding their mutual influences on the IT governance mode" (Sambamurthy and Zmud 1999). Article IV identified and introduced seven contingency factors (F1–F7) that differentiate organizations' AI governance processes as described above and shape organizations' AI governance approaches. Based on different observed configurations of governance processes and contingency factors, four archetypal approaches to organizational AI governance were identified: differentiating, pragmatic, risk-taking, and disinterested.

5.2   Contributions

5.2.1   Contributions to theory

This study and its key findings enhance the field of IS research in three distinct ways: i) by contributing to the ongoing discussion of responsible AI and AI governance; ii) by defining the term "AI governance" and positioning it in relation to the increasingly complex organizational governance landscape; and iii) by integrating the IS literature on strategic IS planning, IT governance, and contingency theory. Table 10 summarizes the main contributions of the research.

The research contributes to the ongoing *discussion of responsible AI and AI governance* by providing empirical insights into how ethical AI principles are put into practice in AI development and deployment (see Mayer et al. 2021), as principle-based ethics do not guarantee that the principles are implemented in practice (Hagendorff 2020; Mittelstadt 2019; Morley et al. 2020). Principles focus on the what, rather than the how, within AI ethics (Morley et al. 2020). Ethical principles should be translated into sufficiently concrete governing actions to guide organizations that deploy AI (Morley et al. 2020; Schiff et al. 2021b; Whittlestone et al. 2019). To ensure their applicability in practice, these principles should be enforceable through governance (Cath 2018; Minkkinen et al. 2021; Morley et al. 2020). Mikalef et al. (2022) and Åkerman et al. (2022) studied responsible AI, negative AI outcomes, and the dark side of AI. This research introduces key dimensions of responsible AI (such as fairness, safety, and transparency), related issues, and expected dark-side assumptions. Following on from the introduced issues, there has been an increasing and continuing emphasis on AI governance in academia (Barn 2020; Koniakou 2023; Mäntymäki et al. 2022a, 2022b; Minkkinen et al. 2023; Papagiannidis et al. 2023; Seppälä et al. 2021; Zimmer et al. 2022) and industry by professional services firms such as Deloitte or PwC (Mökander&Floridi, 2021). The articles included in this dissertation contribute to developing an understanding of how principle-based AI ethics discourse translates into the organizational context and, more specifically, can be put in place within

organizations as development and deployment actions. Article I systematically reviewed the state-of-the-art literature on AI ethics and governance, providing insight into key themes and focusing on identifying existing knowledge gaps, as well as providing an initial conceptual framework for researchers to address to improve organizational AI governance. Article I also identifies the four key agendas around technology, regulation, stakeholder engagement, and operational processes enabling organizational AI governance.

Defining the term AI governance and positioning it in relation to the increasingly complex organizational governance landscape contributes to IS research. Within the academic literature, efforts have been made to define AI governance (e.g., Floridi 2018; Butcher and Berdize 2019; Gahnberg 2021; Gasser and Almeida 2017; Shneiderman 2020), but contributions specifically defining AI governance for organizations were needed. IT governance is an established area of research in IS (e.g., Sambamurthy and Zmud 1999; Tiwana et al. 2013; Weill and Ross 2004), but AI governance is still an emerging topic within IS research. Since 2021, the volume of IS literature on AI governance has increased considerably. One key aspect of IT governance and data governance is that they specify a framework of decision rights and accountability (DAMA International 2009, p. 19; Sambamurthy and Zmud; Weber and Otto 2009; Weill and Ross 2004), outlining how IT and data are governed (Tiwana et al. 2013). IT governance and data governance are implemented through specific mechanisms, which are often grouped under structural, processual, and relational and which contribute to the effectiveness of governance (Abraham et al. 2019; Van Grembergen et al. 2004; Wu, Straub, and Liang 2015). IT and data governance aim to ensure desirable behavior regarding the use of IT (Weill and Ross 2004) and data (Weber and Otto 2009). However, the adequacy of these governance methods for governing AI is currently in question. The evolving nature of AI technologies presents new challenges and governance needs, particularly because of AI systems' learning abilities, adaptiveness, and opacity, which differ significantly from traditional rule-based, data-centric IS (Berente et al. 2020; Kaplan and Haenlein 2019). To address the identified and outlined research gap, the research contributed by providing a definition of organizational AI governance. The research explored the placement of AI governance within the already crowded space of corporate, IT, and data governance. The presented definition of organizational AI governance has been a crucial element in bridging ethics principles with practical applications. The literature-based analysis identified the key agendas that need to be developed to enable organization-level AI governance (stakeholder and context, technical, regulatory, and process agendas), which were further expanded to key components and subcomponents with the support of the conducted empirical research.

The research contributes to IS domain by *integrating with IS research and related literature on strategic IS planning (SISP)*, IT governance and contingency theory.

Archetypes and approaches have been successfully utilized in IS research (Earl 1993; Segars and Grover 1999) to design organizational IT governance (Weill and Ross 2005), and contingency theory has been commonly used for IS theorizing. The empirical research in the dissertation and observations from operational AI governance shed light on the existing governance practices and priorities at the organizational level within the consumer banking and insurance sectors but also provide insights that can generalized not depending on a specific industry. The categorization of organizational archetypes for AI governance introduces a novel way to understand and manage organizations' strategic approaches to governing AI, and the contingency factors list the key characteristics that influence the choice of an AI governance approach. The study elucidates how contingency factors both provide opportunities and set boundaries for organizational decision-making on AI governance.

Table 10. Research contributions per area.

Areas of contribution	Contribution	Contributing to IS literature
Contributing to the ongoing discussion of responsible AI and AI governance	Developing an understanding of how AI ethics and responsible AI discourse translate into the organizational context and can be put in place within organizations as development and deployment actions.	Cath 2018; Mikalef et al. 2022; Morley et al. 2020; Schiff et al. 2021b; Whittlestone et al. 2019; Åkerman et al. 2022
Defining the AI governance term and positioning it in relation to the increasingly complex organizational governance landscape	Examining and analyzing the emerging literature on AI governance and positioning it with IT and data governance academic research. Identifying the key agendas to enable organization-level AI governance: stakeholder and context, technical, regulatory and process agendas.	Berente et al. 2020; Butcher and Berdize 2019; Floridi 2018; Gahnberg 2021; Gasser and Almeida 2017; Sambamurthy and Zmud 1999; Shneiderman 2020; Tiwana et al. 2013; Weill and Ross 2004
Integrating IS research on strategic IS planning, IT governance, and contingency theory	Integrating contingency theory, SISP, and organizational IT governance planning in IS. Conducting research utilizing empirically identified contingencies and organizational archetype approaches.	Earl 1993; Segars and Grover 1999; Weill and Ross 2005; Weill and Olson 2015

5.2.2 Contributions to organizations and society

This dissertation offers practical contributions to organizations and society in four key areas: i) Practical guidance on integrating AI governance with existing organizational IT governance practices, ii) Identification of organizational archetypes for AI governance along with related contingencies, iii) Providing

proposals for organizational arrangements governing AI, and iv) Bridging the regulatory agenda with organizations implementing it.

A practical consideration is how AI governance will be integrated into the broader governance structures of organizations. Although the literature on AI governance has only briefly addressed its relationship with IT and data governance, obviously AI governance cannot function in isolation, neither in organizations. It will likely draw on existing IT governance frameworks, such as COBIT (Article III). However, it remains uncertain to what extent these frameworks can fully accommodate the unique features of AI technologies, including the heightened focuses on ethical considerations and opaque nature of AI systems (Berente et al. 2022; Laato et al. 2022b). Regardless of how AI and IT governance may eventually converge, organizations must critically assess whether their current IT and data governance frameworks are sufficient to address the specific challenges presented by AI, framework development within scope of this dissertation may be utilized. Conducted empirical analysis of AI governance within organizations offers insights into real-world practices and priorities, particularly in the consumer banking and insurance sectors. While these sectors provide a rich context for examining the practical application of AI governance frameworks, the findings are also applicable to other industries facing similar governance questions. Another practical contribution of this research is the identification of organizational archetypes for AI governance, introducing practitioners a guiding framework to understand and manage strategic approaches to AI governance. These archetypes not only serve as models for structuring AI governance but also highlight the contingency factors—such as organizational size, industry-specific regulations, technological capabilities, and risk tolerance—that influence the selection and implementation of governance approaches. These factors can be seen as mechanisms that translate ethical principles into actionable governance practices. In other words, they act as a bridge between high-level ethical frameworks and the practical demands of implementing AI governance. This translation logic is crucial for aligning AI governance strategies with broader ethical considerations, ensuring the responsible and transparent use of AI technologies. By highlighting these dynamics, this research provides a deeper understanding of how organizations are currently approaching AI governance and offers practical insights that can inform the development of more robust and adaptive governance models across sectors. Although managers have some flexibility in choosing a differentiating, risk-taking, pragmatic, or disinterested AI governance approach, consideration of the relevant contingency factors ensures that this strategic choice is sustainable in the years to come. Managers need to bear in mind that no approach is inherently superior, inferior, or future proof. Every organization requires a unique approach that is best suited to the requirements stemming from its evolving environment. By introducing contingency factors and archetypal approaches, the

study helps organizations position their AI governance development as a strategic and actionable activity instead of each organization individually interpreting the multitude of published principles and guidelines and investing in developing its own understanding of key drivers and related activities in governing AI. In other words, the contingency factors and archetypes give organizational managers reference points to facilitate strategic discussions and decisions on AI governance. These reference points foster efficiency in AI governance and build a shared knowledge base for making informed decisions on AI governance and AI strategy more generally. As a third contribution, a practical recommendation, we propose establishing an organizational AI Oversight Unit (AOU) to oversee AI governance processes and facilitate stakeholder engagement. This unit would serve as a centralized point of accountability, ensuring the responsible deployment of AI technologies. While the AOU would provide an initial contact for AI governance, it is essential for organizations to embed AI governance practices comprehensively within their existing structures.

Finally, the emerging regulatory framework for AI, as well as the broader data and digital domains, is inherently a risk-based approach, allowing flexibility in implementation, and blending hard enforcement with softer elements, such as guidelines or codes of conduct. The definitive position of these regulations will formalize once both, regulations and organizational AI governance practices are established. The impact of these regulations on organizations will largely depend on the outcomes of the initial legal decisions, which will set important precedents for how the regulations are interpreted and enforced. In this context, having robust organizational AI governance practices will be a crucial enabler in navigating and shaping this regulatory landscape.

5.3 Limitations

Three key limitations were identified in the research. First, the area of research (AI governance) is novel and unestablished, and it is quickly developing. The second limitation evolves from the first: The concept's development and empirical research were conducted in parallel. Third, owing to the low level of integration with traditional IS research streams, the research contained frameworks development, based on the analyzed literature and the data collected from expert interviews, but did not integrate with the existing theories, beyond contingency theory in one article (Article IV).

The primary constraint, or, rather, the nature of the research area, relates to timeliness. The research domain is nascent, unestablished, and inherently volatile in nature. This domain is undergoing a rapid evolution, whereby technologies and associated processes are continually advancing. Furthermore, most of the existing

AI governance research is predicated upon a narrow conceptualization of AI, thereby not being fully able to analyze the broader transformative impacts of generative AI. Timeliness also applies to the state of the regulatory landscape within the domain. Where the foundational ethics and the principles remain, the reflecting regulation is in development. For example, the European Union AI Act was introduced, commented on, and finally approved during the dissertation research. The area is also moving fast. One could say that there is even a mismatch in academic research processes and their suitability for this quickly evolving area. There is a noticeable disconnect between the pace of academic research and the rapidly changing nature of the field. For instance, it took 18 months to develop the second empirical study to meet the (assumed) high standards of a prestigious journal. However, during these 1.5 years, the field of AI underwent drastic changes, exemplified by the launches of GPT-3.5 (and the related ChatGPT citizen interface), GPT-4, Dalle-2, Midjourney, and Google Bard, placing AI in the hands of citizens across the globe. It is unclear how established ethical principles are developing while the focus is moving from predictive AI toward generative AI.

The area is becoming established and developing quickly, reflecting the second research limitation: concept development and the ability to conduct empirical research simultaneously. The SLR (Birkstedt et al. 2023) revealed a need for terminology and structure-related definition and analysis work vis-à-vis organizational AI governance. As explained in Chapter 4, a qualitative research approach is selected when the concepts and theories are "immature" owing to a lack of theory and previous research (Morse 1991, Creswell 2019). The expert interviews that were conducted were semi-structured and were helpful for exploring new insights or ideas that emerged during the interview. The research was largely conducted utilizing the externally funded AI Governance and Auditing (AIGA) program's North European university and organization network, largely limiting the empirical viewpoint to the European perspective.

The third limitation, the low level of integration with traditional IS research streams, is equally interconnected with the first (i.e., the novelty and development speed of the research area). The research conducted in this dissertation has largely focused on defining organizational AI governance; positioning it within the context of corporate, IT, and data governance; and identifying the key agendas and components of organizational AI governance. There is room for development to gain an in-depth understanding of AI governance's relationship to, and possibly even integration with, IT governance. The final empirical paper (Article IV) was developed using contingency theory, but there is room for further integration with existing theory development in the IS research area.

5.4   Future research directions

There are three identified directions for future research that would continue the developments discussed in the dissertation: conducting in-depth organizational AI governance research to collect more comprehensive data at the organizational level and focus more specifically on generative AI (versus traditional machine learning); continuing the examination of the relationship between AI governance and IT governance; and introducing theoretical perspectives, such as actor network theory (ANT), to improve future research.

Foundational work (definition, positioning, and structuring) has progressed for the past four years and has been supported by this research. This makes room for further research, enabling, for example, longitudinal case studies within organizations or quantitative data collection using more structured data collection and analysis mechanisms. The area of research is most likely to develop, as the focus has been shifting from narrow AI to generative AI. There will be a continuous need to revalidate the set research questions and developed research outputs (e.g., definition of organizational AI governance and key components of AI governance) against the developing world dominated by generative AI developments.

Previous research on IS indicates that IT governance acts as an intermediary between IT investment and organizational performance, thereby influencing the business value generated by IT (Turedi and Zhu 2019). In particular, the ability of an organization to identify, design, implement, and utilize IT governance processes has been found to enhance IT performance. This improvement in IT performance subsequently leads to better business performance (Joshi, Benitez, Huygh, Ruiz, and De Haes 2022). Based on this research, managing the ethics-related risk and governing AI will require a more system-based view than traditional IT governance. This antagonism between IT governance and AI governance may provide opportunities for more research, possibly leading to closer integration or even to the migration of IT governance's role, emphasizing the view that includes the AI governance system.

This research has identified and communicated the importance of translating AI ethical principles into practice. ANT has been recognized as having the potential to advance understanding of the complex social interactions associated with IT (Sarker et al. 2006; Hanseth et al. 2004; Walsham 1997), especially in the areas that are developing and in constant flux (Walsham 1997; Islam et al. 2019). ANT helps explain how agents interact, and it allows for the analysis of both artificial and non-artificial agents within the same context, avoiding the need to think in terms of human/nonhuman barriers and equally ignoring the hierarchical distribution of actors (Latour 2005; Tatnall 2005). ANT offers a methodology for analyzing the development of phenomena over time, providing tools to continue analysis beyond the finalization of the dissertation. The area under study is quickly evolving, and the

empirical information in the scope of the research is primarily collected on the responsibility of narrow AI solutions. ANT might provide a useful methodology for further analyzing the implications of generative AI for organizational practices, which has been a limited new capability in the research context to date.

In summary, I believe that future research on organizational AI governance will be important in developing both an empirical understanding and a robust theoretical foundation. Additionally, a remarkable area of research lies in exploring the role of AI governance within the broader context of corporate governance, including AI governance's relationship to IT and data governance.

List of Abbreviations

AEDT	Automatic Employment Decision Tools Law (New York City)
AI	Artificial Intelligence
AIGA	AI Governance and Auditing program
ANT	Actor Network Theory
COBIT	Control Objectives for Information Technologies
GDPR	General Data Protection Regulation
IS	Information Systems
IT	Information Technology
LLM	Large Language Model
PAI	Partnership on AI
RQ	Research Question
SISP	Strategic IS Planning
SLR	Systematic Literature Review

List of References

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. ISSN 0268-4012. https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Aguilera, R.V., Filatotchev, I., Gospel, H., and Jackson, G. An Organizational Approach to Comparative Corporate Governance: Costs, Contingencies, and Complementarities. Organization Science, 19, 3 (2008), 475–492.

Agrawal, A., Gans, J.S., Goldfarb, A. (2019). Exploring the impact of artificial Intelligence: Prediction versus judgment. Information Economics and Policy, 47, 1-6. https://doi.org/10.1016/j.infoecopol.2019.05.001

Aliman, N., Kester, L. (2019). Transformative AI Governance and AI-Empowered Ethical Enhancement Through Preemptive Simulations. ResearchGate. DOI:10.21552/delphi/2019/1/6

Algorithmwatch (2023) (https://inventory.algorithmwatch.org/), visited 12.1.2024

Allam, H., Dempere, J., Akre, V., Parakash, D., Mazher, N., & Ahamed, J. (2023, May). Artificial intelligence in education: an argument of Chat-GPT use in education. In 2023 9th International Conference on Information Technology Trends (ITT) (pp. 151-156). IEEE.

Armat, M., Assarroudi, A., Rad, M., Sharifi, H., Heydari, A. (2018). Inductive and Deductive: Ambiguous Labels in Qualitative Content Analysis. The Qualitative Report, 23(1), 219-221. https://doi.org/10.46743/2160-3715/2018.2872

Artificial Intelligence Incident Database. Responsible AI Collaborative. Retrieved on May 31, 2024 from incidentdatabase.ai/cite/92.

Ashok, M., Madan, R., Joha, A., Sivarajah, U. (2022). Ethical framework for Artificial Intelli-gence and Digital technologies. International Journal of Information Management 62(102433). https://doi.org/10.1016/j.ijinfomgt.2021.102433

Assarroudi A., Heshmati Nabavi F., Armat M.R., Ebadi A., Vaismoradi M. (2018). Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. Journal of Research in Nursing. 23(1), 42-55. doi:10.1177/1744987117741667

Baily, M., Brynjolfsson, E., & Korinek, A. (2023). Machines of mind: The case for an AI-powered productivity boom.

Barn, B. S. (2020). Mapping the public debate on ethical concerns: algorithms in mainstream media. Journal of Information, Communication and Ethics in Society, 18(1), 124-139.

Benbya, H., Davenport, T.H., and Pachidi, S. Artificial intelligence in organizations: current state and future opportunities. MIS Quarterly Executive, 19, 4 (2020).

Berente, N., Gu, B., Recker, J., Santhanam, R. (2021). "Managing Artificial Intelligence," MIS Quarterly Special Issue: Managing AI

Berente, N., Kormylo, C., & Rosenkranz, C. (2024). Test-Driven Ethics for Machine Learning. Communications of the ACM, 67(5), 45-47.Beulen, E., Plugge, A., van Hillegersberg, J. (2022). Formal and relational governance of artifi-cial intelligence outsourcing. Information Systems and E-Business Management, 20, 719–748. https://doi.org/10.1007/s10257-022-00562-7

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. Internet Research, 33(7), 133-167.

Breidbach, C., Maglio, P. (2020). Accountable algorithms? The ethical implications of data-driven business models. Journal of Service Management, Vol. 31 No. 2, 163-185. https://doi.org/10.1108/JOSM-03-2019-0073

Brendel, A. B., Mirbabaie, M., Lembcke, T-B., Hofeditz, L. (2021). Ethical Management of Artificial Intelligence. Sustainability 2021, 13(4), 1974; https://doi.org/10.3390/su13041974

Brown, A., & Grant, G. (2005). Framing the Frameworks: A Review of IT Governance Re-search. Communications of the Association for Information Systems, 15. https://doi.org/10.17705/1CAIS.01538

Butcher, J., Beridze, I. (2019). What is the state of artificial intelligence governance globally? RUSI J. 164, 88–96. https://doi.org/10.1080/03071847.2019.16942 60

California Department of Technology. GenAI Executive Order. 2024. https://cdt.ca.gov/technology-innovation/artificial-intelligence-community/genai-executive-order/.

California Legislative Information. SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. 2024. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047.

Castelvecchi, D. (2016). Can we open the black box of AI? Nature, 538(7623).

Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. The Royal Society Publishing, Philosophical Transactions A 376(2133):20180080. DOI:10.1098/rsta.2018.0080

Chan, A. (2023). GPT-3 and InstructGPT: technological dystopianism, utopianism, and "Contextual" perspectives in AI ethics and industry. AI and Ethics, 3(1), 53-64.

Chiao, V. (2019) Fairness, accountability and transparency: notes on algorithmic decision making in criminal justice. International Journal of Law in Context 15(2), 126-139. DOI:10.1017/S1744552319000077

Cihon, P., Maas, M.M., Kemp, L. (2020). Fragmentation and the Future: Investigating Architec-tures for International AI Governance. Global Policy, 11(5), 545-556. https://doi.org/10.1111/1758-5899.12890

Cho, K., van Merrienboer, B., Bahdanau, D., Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. arXiv:1409.1259.

Connelly, L. M. (2016). Trustworthiness in qualitative research. Medsurg nursing, 25(6), 435.

Cooper, H. (1998). Synthesizing research: A guide for literature reviews (3rd ed.). Sage Publi-cations, Inc.

Cordella, Antonio & Shaikh, Maha. (2006). From Epistemology to Ontology: Challenging the Constructed Truth of ANT.

Creswell, J. W. (2009). Research designs. Qualitative, quantitative, and mixed methods approaches. Sage Publications, Inc.Crotty, M. J. (1998). The foundations of social research: Meaning and perspective in the research process. The foundations of social research, 1-256.

Crotty, M. (1998). The Foundations of Social Research. Meaning and Perspective in the Research Process. SAGE Publications Ltd.

DAMA International (2009). The DAMA Guide to The Data Management Body of Knowledge. 19. https://www.academia.edu/19992490/The_DAMA_Guide_to_the_Data_Management_Body_of_Knowledge_First_Edition, accessed 3.2.2024

Davenport, T.H. (2018). The AI Advantage: How to put the artificial intelligence revolution to work, 7-8.

De Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation?. Philosophy & technology, 34(4), 1135-1193.

Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R. (2023). Navigating the Jagged Technological Fron-tier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013. SSRN: https://ssrn.com/abstract=4573321.

Domanski, R.J. (2019), "The A.I. pandorica: linking ethically-challenged technical outputs to prospective policy approaches", Proceedings of the 20th Annual International Conference on Digital Government Research, pp. 409-416, doi: 10.1145/3325112.3325267.

Dwivedi, Y.K., Kshetri, N., Hughes, L. et al. (2023). "So what if ChatGPT wrote it?" Multidis-ciplinary perspectives on opportunities, challenges and implications of generative conversa-tional AI for research, practice and policy. International Journal of Information Manage-ment, 71(102642). ISSN 0268-4012.

Earl, M.J. (1993). Experiences in Strategic Information Systems Planning. MIS Quarterly 17(1), 1-24. https://doi.org/10.2307/249507

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. The Academy of Management Review, 14(4), 532–550. https://doi.org/10.2307/258557

Eriksson, P., & Kovalainen, A. (2015). Qualitative methods in business research: A practical guide to social research. Sage. p 294-296

European Artificial Intelligence Act (2024). Artificial intelligence act https://artificialintelligenceact.eu/the-act/

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M. (2020), Principled Artificial Intelli-gence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication No. 2020. https://dx.doi.org/10.2139/ssrn.3518482

Floridi, L. (2018). Soft ethics, the governance of the digital and the general data protection regulation. Phil. Trans. R. Soc. A 376, 20180081. https://doi.org/10.1098/ rsta.2018.0081

Floridi, L., Cowls, J., Beltrametti, M., et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28, 4 (2018), 689–707.

Gahnberg, C. (2021). What rules? Framing the governance of artificial agency. Policy Soc. 40, 194–210. https:// doi.org/10.1080/14494035.2021.1929729

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. Machine learning and the city: Applications in architecture and urban design, 535-545.

Gasser, U., Almeida, V. (2017). A Layered Model for AI Governance. Institute of Electrical and Electronics Engineers (IEEE), Internet Computing 21(6), 58-62. https://doi.org/10.1109/MIC.2017.4180835

Giffen, B. V., & Ludwig, H. (2023). How Boards of Directors Govern Artificial Intelligence. MIS Quarterly Executive, 22(4), 7.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational research methods, 16(1), 15-31.

Gilson, L. L., Goldberg, C. B. (2015). Editors' Comment: So, What Is a Conceptual Paper? Group & Organization Management, 40(2), 127-130. https://doi.org/10.1177/1059601115576425

Gregory, R.W., Kaganer, E.A., Henfridsson, O., Ruch, T.J. (2018). IT consumerization and the transformation of IT governance. MIS Quarterly 42(4), 1225-1253. DOI:10.25300/MISQ/2018/13703

Hagendorff, T. (2020), The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Ma-chines, Volume 30, 99–120. https://link.springer.com/article/10.1007/S11023-020-09517-8

Hanseth, O., Aanestad, M., Berg, M. (2004). Guest editors' introduction: Actor-network theory and information systems. What's so special? Information Technology & People, 17(2), 116-123. https://doi.org/10.1108/09593840410542466

Harlow, H. (2018). Ethical Concerns of Artificial Intelligence, Big Data and Data Analytics. European Conference on Knowledge Management; Kidmore End (Sep 2018). https://www.proquest.com/openview/56146799514cd6f3c4e92e82079fb128/1?pq-origsite=gscholar&cbl=1796412

Hathaway, R.S. (1995). Assumptions underlying quantitative and qualitative research: Implica-tions for institutional research. Higher Education 36(5), 535-562. <http://hdl.handle.net/2027.42/43615>

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

Hsieh H-F., Shannon S.E. (2005). Three Approaches to Qualitative Content Analysis. Qualita-tive Health Research. 15(9), 1277-1288. doi:10.1177/1049732305276687

Holden, M.T., Lynch, P. (2004). Choosing the Appropriate Methodology: Understanding Re-search Philosophy. The Marketing Review 4(4). DOI:10.1362/1469347042772428

Islam, N. A K. M., Mäntymäki, M., Turunen, M. (2019). Why do blockchains split? An actor-network perspective on Bitcoin splits. Technological Forecasting and Social Change
Volume 148, 119743. https://doi.org/10.1016/j.techfore.2019.119743

Jia, K., Zhang, N. (2022). Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines. Electron Markets 32, 59–71. https://doi.org/10.1007/s12525-021-00480-5

Joshi, A., Benitez, J., Huygh, T., Ruiz, L., De Haes, S. (2022). Impact of IT governance process capability on business performance: Theory and empirical evidence. Decision Support Sys-tems, 153. Article I13668. https://doi.org/10.1016/j.dss.2021.113668

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence volume 1, 389–399. https://www.nature.com/articles/s42256-019-0088-2

Kaminski, M.E., Malgieri, G. (2021). Algorithmic Impact Assessments Under the GDPR: Pro-ducing Multi-Layered Explanations. International Data Privacy Law, 11(2), 125-144. DOI https://doi.org/10.1093/idpl/ipaa020

Kaplan, A., Haenlein, M (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. Sage Journals, 61(4). https://doi.org/10.1177/0008125619864925

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103(102274).

Khowaja, S. A., Khuwaja, P., Dev, K., Wang, W., & Nkenyereye, L. (2024). Chatgpt needs spade (sustainability, privacy, digital divide, and ethics) evaluation: A review. Cognitive Computation, 1-23.

Kitchenham, B. and Charters, S. (2007). Guidelines for Performing Systematic Literature Re-views in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report. https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering

Kodiyan, A. A. (2019). An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. Researchgate Preprint, 1-19.

Koniakou, V. From the "rush to ethics" to the "race for governance" in Artificial Intelligence. Information Systems Frontiers, 25, 1 (2023), 71–102.

Kroll, J.A. (2018). The fallacy of inscrutability. The Royal Society Publishing, 376(2133). DOI: https://doi.org/10.1098/rsta.2018.0084

Kumar A., Braud, T., Tarkoma, S., Hui, P. (2020). Trustworthy AI in the Age of Pervasive Computing and Big Data. Institute of Electrical and Electronics Engineers, 2020 IEEE In-ternational Conference on Pervasive Computing and Communications Workshops. https://doi.org/10.1109/PerComWorkshops48775.2020.9156127

Laato, S., Birkstedt, T., Mäntymäki, M., Minkkinen, M., Mikkonen, T. (2021). AI governance in the system development life cycle: insights on responsible machine learning engineering. CAIN '22: Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI. 113–123. https://doi.org/10.1145/3522664.3528598

Larsson, S., Heintz, F. (2020). Transparency in artificial intelligence. Internet Policy Review, 9(2). https://doi.org/10.14763/2020.2.1469

Latour, B. (2005). From realpolitik to dingpolitik. Making things public: Atmospheres of democracy, 1444.

Levstek, A., Hovelja, T., Pucihar, A. (2018). IT Governance Mechanisms and Contingency Fac-tors: Towards an Adaptive IT Governance Model. Organizacija 51(4), 286-310. DOI:10.2478/orga-2018-0024

Lincoln, Yvonna & Guba, Egon. (1986). But Is It Rigorous? Trustworthiness and Authenticity in Naturalistic Evaluation. New Directions for Program Evaluation. 1986. 73 - 84.

Lysaght, T., Lim, H.Y., Xafis, V. and Ngiam, K.Y. (2019), "AI-assisted decision-making in healthcare", Asian Bioethics Review, Vol. 11 No. 3, pp. 299-314, doi: 10.1007/s41649-019-00096-0.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures, 90, 46–60. https://doi.org/10.1016/j.futures.2017.03.006

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. https://doi.org/10.1080/0960085X.2022.2026621

Mikalef, P., Krogstie, J. (2020). Examining the interplay between big data analytics and contex-tual factors in driving process innovation capabilities. European Journal of Information Systems, 29:3, 260-287. DOI: 10.1080/0960085X.2020.1740618

Mayring, P. (2000). Qualitative Content Analysis. Forum Qualitative Sozialforschung 1(2). DOI: https://doi.org/10.17169/fqs-1.2.1089.

Minkkinen, M., Zimmer, M.P., Mäntymäki, M. (2021). Towards Ecosystems for Responsi-ble AI. Lecture Notes in Computer Science ((LNISA,volume 12896)). https://doi.org/10.1007/978-3-030-85447-8_20

Minkkinen, M., Niukkanen, A., Mäntymäki, M. (2022). What about investors? ESG analyses as tools for ethics-based AI auditing. AI and Society, 39, 329–343. https://doi.org/10.1007/s00146-022-01415-0

Minkkinen, M., Mäntymäki, M. (2023). Discerning Between the "Easy" and "Hard" Problems of AI Governance. IEEE Transactions on Technology and Society, 4(2), 188 – 194. DOI: 10.1109/TTS.2023.3267382

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelli-gence 1(11). DOI:10.1038/s42256-019-0114-4

Mora, M., Gelman, O., Paradice, D., & Cervantes, F. (2008, May). The case for conceptual re-search in information systems. In CONF-IRM 2008 Proceedings (p. 52).

Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Prac-tices. Science and Engineering Ethics, Volume 26, 2141–2168. https://link.springer.com/article/10.1007/s11948-019-00165-5%23Tab1

Morse, J. M. (1991). Approaches to Qualitative & Quantitative Methodological Triangulation. Nursing Research, 40(2), 120-123. https://doi.org/10.1097/00006199-199103000-00014

Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M. (2022a). Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance. arXiv, 2206.00335. https://doi.org/10.48550/arXiv.2206.00335

Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M. (2022b). Defining organizational AI governance. AI and Ethics 2(4) 1-7. DOI:10.1007/s43681-022-00143-x

Möhlmann, M., Bentley, U., Zalmanson, L., et al. Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control. MIS Quarterly, 45, 4 (2021), 1999–2022.

Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. Minds and Machines, 31(2), 323-327.

Mökander, J., Juneja, P., Watson, D., Floridi, L. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other? Minds and Machines 32(4), 1-8. DOI:10.1007/s11023-022-09612-y

Murray, D., Stankovic, L., and Stankovic, V. Transparent AI. explainability of deep learning based load disaggregation. In ACM, 2021, pp. 5.

Nasim, S. F., Ali, M. R., & Kulsoom, U. (2022). Artificial intelligence incidents & ethics a narrative review. *International Journal of Technology, Innovation and Management (IJTIM)*, *2*(2), 52-64.

Newlands, G. (2021). Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. Sage Journals. https://doi.org/10.1177/20539517211016026

New York Times (2023) https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html

New York Times (2024) https://www.nytimes.com/2024/04/30/business/media/newspapers-sued-microsoft-openai.html, accessed 15.5.2024

NYC Mayor's Office of the Chief Technology Officer (2021). AI Primer, The New York City Artificial Intelligence Primer, 16-28. The Official Website of the City of New York. https://www.nyc.gov/assets/cto/downloads/ai-strategy/nyc_ai_primer.pdf, accessed 12.3.2024

NYC Consumer and Worker Protection (2023). Automated Employment Decision Tools (AEDT). The Official Website of the City of New York. https://www.nyc.gov/assets/dca/downloads/pdf/about/DCWP-AEDT-FAQ.pdf, accessed 12.3.2024

OECD (2019) https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, accessed 19.1.2024

Okoli, C., & Schabram, K. (2015). A guide to conducting a systematic literature review of information systems research.

Otto, Boris (2010). IT Governance and Organizational Transformation: Findings From an Ac-tion Research Study. AMCIS 2010, Proceedings. 421. http://aisel.aisnet.org/amcis2010/421

Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P., Korgstie, J. (2023). Toward AI Gov-ernance: Identifying Best Practices and Potential Barriers and Outcomes. Information Systems Frontiers, 25, 123–141. https://doi.org/10.1007/s10796-022-10251-y

Partnership of AI (2016). Our tenets. https://partnershiponai.org/about/#:~:text=Our%20Tenets, accessed 3.2.2024

Patton, M.Q. (1990). Purposeful Sampling I: Qualitative Evaluation and Research Methods. Sage, London

Perry, B., Uuk, R. (2019). AI Governance and the Policymaking Process: Key Considerations for Reducing AI Risk. Big Data Cogn. Comput. 2019, 3(2), 26. https://doi.org/10.3390/bdcc3020026

Personal Data Protection Commission (2019) https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf, accessed 6.1.2024

Polyviou, A., Zamani, E.D. (2023) Are we Nearly There Yet? A Desires & Realities Framework for Europe's AI Strategy. Information Systems Frontiers 25, 143–159. https://doi.org/10.1007/s10796-022-10285-2

Rana, M.S. Nobi, M. N. Murali B. and Sung, A. H. (2022) "Deepfake Detection: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 25494-25513, doi: 10.1109/ACCESS.2022.3154404.

Razzaque, A. (2021). Artificial intelligence and IT governance: a literature review. The Big Data-Driven Digital Economy: Artificial and Computational Intelligence, 85-97.

Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. Telecommunications Pol-icy, 44(6). https://doi.org/10.1016/j.telpol.2020.101937

Sambamurthy, V., Zmud, R.W. (1999). Arrangements for Information Technology Governance: A Theory of Multiple Contingencies. MIS Quarterly, 23(2), 261-290. https://doi.org/10.2307/249754

Sarker, S., Sarker, S., Sidorova, A. (2006). Understanding Business Process Change Failure: An Actor-Network Perspective. Journal of Management Information Systems, 23:1, 51-86, DOI: 10.2753/MIS0742-1222230102

Saunders, M., Lewis, P. and Thornhill, A. (2007) Research Methods for Business Students. 4th Edition, Financial Times Prentice Hall, Edinburgh Gate, HarlowSharma, A., Muhammad, A., Turesson, C. (2020). Incidence and prevalence of giant cell arteritis and polymyalgia rheumatica: A systematic

literature review. Semin Arthritis Rheum. 2020 Oct;50(5).1040-1048. doi: 10.1016/j.semarthrit.2020.07.005. Epub 2020 Jul 14. PMID: 32911281.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., Lennon, M. (2021b). Explaining the principles to practices gap in AI. IEEE Technology and Society Magazine, 40(2). DOI: 10.1109/MTS.2021.3056286

Schneider, J., Abraham, R., Meske, C., vom Brocke, J. (2020). AI governance for businesses. arXiv:2011.10672 [cs]

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. AIS Trans-actions on Human-Computer Interaction, 12(3), 109-124. https://doi.org/10.17705/1thci.00131

Schwarz, A. and Hirschheim, R. An extended platform logic perspective of IT governance: managing perceptions and activities of IT. The Journal of Strategic Information Systems, 12, 2 (2003), 129–166.

Segars, A.H., Grover, V. (1999). Profiles of Strategic Information Systems Planning. Infor-mation Systems Research 10(3),199-232. DOI:10.1287/isre.10.3.199

Select Committee of Artificial Intelligence (2018) https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf, accessed 6.1.2024

Seppälä, A., Birkstedt, T., Mäntymäki, M. (2021). From Ethical AI Principles to Governed AI. Forty-Second International Conference on Information Systems, Austin 2021. https://www.researchgate.net/publication/358234837_From_Ethical_AI_Principles_to_Governed_AI

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning?. British Journal of Educational Technology, 51(5), 1450-1484.

Solomonoff, R. J. (1964). A formal theory of inductive inference. Part I. Information and control, 7(1), 1-22.

Spulber, D. F. (2011). Should business method inventions be patentable? Journal of Legal Analysis, 3(1)

Strauss, A., & Corbin, J. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Sage Publications, Inc.

Tallon, P., Short, J., Harkins, M. (2013). The Evolution of Information Governance at Intel. MIS Quarterly Executive, 12(4, Art.5). https://aisel.aisnet.org/misqe/vol12/iss4/5 (2013)

Tandon, A., Dhir, A., Islam, N. A K. M., Mäntymäki, M. (2020). Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda. Elsevier, Computers in Industry, 122, 103290. https://doi.org/10.1016/j.compind.2020.103290

Tatnall, A. (2005). Actor-network theory in information systems research. In Encyclopedia of Information Science and Technology, First Edition (pp. 42-46). IGI Global.

Teodorescu, M., 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). DOI: 10.25300/MISQ/2021/16535.Tiwana, A., Konsynski, B., Venkatraman, B. (2013). Special Issue: Information Technology and Organizational Governance: The IT Governance Cube. Journal of Management Information Systems. 30, 2013 (3). https://doi.org/10.2753/MIS0742-1222300301

Tiwana, A. (2015) Evolutionary Competition in Platform Ecosystems. Information Systems Research, Vol. 26, No. 2, 266-281 https://doi.org/10.1287/isre.2015.0573

Tranfield, D., Denyer, D., Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. British Journal of Management, 14(3), 207-222. https://doi.org/10.1111/1467-8551.00375

Tubaro, P., Casilli, A. (2019). Micro-work, artificial intelligence and the automotive industry. Economia e Politica Industriale: Journal of Industrial and Business Economics, 46(3), 333-345.

Turedi, S., Zhu, H. (2019). How to Generate More Value from IT: The Interplay of IT Invest-ment, Decision Making Structure, and Senior Management Involvement in IT Governance. Communications of the Association for Information Systems, 44. https://doi.org/10.17705/1CAIS.04426

Turing, A. M. (1950). Mind. Mind, 59(236), 433-460.

Ulnicane, I., Knight, W., Leach, T., Carsten Stahl, B., Wanjiku, W-G. (2021). Framing govern-ance for a contested emerging technology: insights from AI policy. Policy and Society, 40(2), 158–177, https://doi.org/10.1080/14494035.2020.1855800

Van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. The information society, 19(4), 315-326.

Van Grembergen, W., De Haes, S., Guldentops, E. (2004). Structures, Processes and Relational Mechanisms for IT Governance. Strategies for Information Technology Governance. DOI: 10.4018/978-1-59140-140-7.ch001

Viljanen, M., Parviainen, H. (2022). AI applications and regulation: Mapping the regulatory strata. Frontiers in Computer Science, 3-2021. https://doi.org/10.3389/fcomp.2021.779957

Walsham, G. (1997). Actor-Network Theory and IS Research: Current Status and Future Pro-spects. Information Systems and Qualitative Research. IFIP — The International Federa-tion for Information Processing. https://doi.org/10.1007/978-0-387-35309-8_23

Weber, K., Otto, B., Oesterle, H. (2009). One Size Does Not Fit All - A Contingency Approach to Data Governance. ACM Journal of Data and Information Quality, 1(1). DOI:10.1145/1515693.1515696

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2), xiii–xxiii. http://www.jstor.org/stable/4132319

Weill, L. (2008). Leverage and Corporate Performance: Does Institutional Environment Matter? Small Business Economics, 30(3), 251-265. DOI:10.1007/s11187-006-9045-7

Weill, P.D., Ross, J.W. (2004). IT Governance: How Top Performers Manage IT Decision Rights for Superior Results. Harvard Business School Press, ISBN: 978-1591392538.

Weill, P.D., Ross, J.W. (2005). A Matrixed Approach to Designing IT Governance. MITSloan Management Review, Winter 2005. https://sloanreview.mit.edu/article/a-matrixed-approach-to-designing-it-governance/, accessed 1.11.2023

White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Whittlestone, J., Nyrup, R., Alexandrova, A., Cave, S. (2019). The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. The 2019 AAAI/ACM Conference. DOI:10.1145/3306618.3314289. https://www.researchgate.net/publication/334378492_The_Role_and_Limits_of_Principles_in_AI_Ethics_Towards_a_Focus_on_Tensions

Winfield, A.F.T., Jirotka, M. (2017). The Case for an Ethical Black Box. Lecture Notes in Com-puter Science. Conference: Towards Autonomous Robotic Systems. DOI:10.1007/978-3-319-64107-2_21

Wirtz, B.W., Weyerer, J.C., Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. Government Information Quarterly, 39(4). https://doi.org/10.1016/j.giq.2022.101685

Woodruff, A., Shelby, R., Kelley, P. G., Rousso-Schindler, S., Smith-Loud, J., & Wilcox, L. (2024, May). How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-26).

Wu, F., Lu, C., Zhu, M., Chen, H. (2020). Towards a new generation of artificial intelligence in China. Nature Machine Intelligence 2(6), 312-316. DOI:10.1038/s42256-020-0183-4Wu, S. P. J., Straub, D. W., Liang, T. P. (2015). How information technology governance mecha-nisms and strategic alignment influence organizational performance: Insights from a matched survey of business and IT managers. Mis Quarterly, 39(2), 497-518. DOI:10.25300/MISQ/2015/39.2.10

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.

Zimmer, M.P., Minkkinen, M., Mäntymäki, M. (2022). Responsible Artificial Intelligence Systems Critical considerations for business model design. Scandinavian Journal of Information Systems, 34(2, art.4).

Ågerfalk, P. J., Conboy, K., Crowston, K., Eriksson Lundström, J., Jarvenpaa, S. L., Ram, S., & Mikalef, P. (2022). Artificial intelligence in information systems: State of the art and research roadmap. Association for Information Systems.