Designing an AI Governance Framework

Mäntymäki, Matti; Minkkinen, Matti; Zimmer, Markus Philipp; Birkstedt, Teemu; Viljanen, Mika

Published in:
ECIS 2023 Proceedings

Publication date:
2023

Document Version
Peer reviewed version

5-11-2023

Designing an AI governance framework: From research-based premises to meta-requirements

Matti Mäntymäki
University of Turku, matti.mantymaki@utu.fi

Matti Minkkinen
University of Turku, matti.minkkinen@utu.fi

Markus Philipp Zimmer
Leuphana University of Lüneburg, maphzi@icloud.com

Teemu Birkstedt
University of Turku, teemu.birkstedt@utu.fi

Mika Viljanen
University of Turku, mika.viljanen@utu.fi

DESIGNING AN AI GOVERNANCE FRAMEWORK: FROM RESEARCH-BASED PREMISES TO META-REQUIREMENTS

Research Paper

Mäntymäki, Matti, University of Turku, Finland, matti.mantymaki@utu.fi

Minkkinen, Matti, University of Turku, Finland, matti.minkkinen@utu.fi

Zimmer, Markus Philipp, Leuphana University Lüneburg, Germany, markus.zimmer@leuphana.de

Birkstedt, Teemu, University of Turku, Finland, teemu.birkstedt@utu.fi

Viljanen, Mika, University of Turku, Finland, mika.viljanen@utu.fi

Abstract

The development and increasing use of artificial intelligence (AI), particularly in high-risk application areas, calls for attention to the governance of AI systems. Organizations and researchers have proposed AI ethics principles, but translating principles into practice-oriented frameworks has proven difficult. This paper develops meta-requirements for organizational AI governance frameworks to help translate ethical AI principles into practice and align operations with the forthcoming European AI Act. We adopt a design science research approach. We put forward research-based premises, then we report the design method employed in an industry-academia research project. Based on these, we present seven meta-requirements for AI governance frameworks. The paper contributes to the IS research on AI governance by collating knowledge into meta-requirements and advancing a design approach to AI governance. The study underscores that governance frameworks need to incorporate the characteristics of AI, its contexts, and the different sources of requirements.

Keywords: Artificial intelligence, AI governance, Design science research, IT governance.

1 Introduction

The rapid global proliferation of artificial intelligence (AI) has sparked vivid discussions among researchers, professional communities, and the popular media about its systemic risks (Crawford and Calo, 2016; Altman, Wood and Vayena, 2018; Lin, 2019; Mikalef *et al.*, 2022). These risks include algorithmic biases, discrimination against minority groups, and reduced human agency (Bolukbasi *et al.*, 2016; O'Neil, 2016; Veale and Binns, 2017; Bechmann and Bowker, 2019). Aligned with the general surge in research activity around AI in recent years (Russell and Norvig, 2021), academic discourse on AI ethics has gained increasing momentum (Etzioni and Etzioni, 2017; e.g., Dignum, 2018; Müller, 2020; Vakkuri, Kemell and Abrahamsson, 2020). Moreover, there is an emerging body of research on the need to regulate AI and any associated challenges (e.g., Kaminski, 2019; Wallach and Marchant, 2019; Robles Carrillo, 2020). Echoing the societal importance of addressing these issues, international and governmental institutions such as the EU, OECD, and the House of Lords (UK), professional organizations (such as the IEEE), companies, and public sector organizations have created and published their ethical principles and guidelines for AI (e.g., Floridi, 2019; Jobin, Ienca and Vayena, 2019; Mittelstadt, 2019; Hagendorff, 2020).

Even though there is no universally agreed set of core human or societal values (and their priorities), there is a consensus among academic and practitioner communities regarding the importance of ensuring that AI operates in accordance with human and societal values (High-Level Expert Group on Artificial

Intelligence, 2019; Dignum, 2020; Fjeld *et al.*, 2020). The increasingly prominent human-centered AI (HCAI) approach focuses on systems that support human goals, activities, and values (Shneiderman, 2020). Arguments have also been presented in the literature about the importance of moving from articulating principles towards enforcing and applying principles in practice (e.g., Cath, 2018; Mittelstadt, 2019; Hagendorff, 2020), which can be described as the governing of artificial intelligence (Seppälä, Birkstedt and Mäntymäki, 2021; Mäntymäki *et al.*, 2022a). Compared to the volumes of research on principle-based ethics of AI, considerably less research has focused on implementing principles in practice (e.g., Hagendorff, 2020; Morley *et al.*, 2020). Even though principles have been established, there is divergence over what issues they affect and how they should be implemented, and the proliferation of AI governance tools belies the lack of production-ready solutions (Jobin, Ienca and Vayena, 2019; Morley *et al.*, 2020; Kazim, Denny and Koshiyama, 2021). Meanwhile, policymakers and practitioners are increasingly pushing AI governance. As an example of the shift from principle-based AI ethics towards the practical governance of AI, the EU published the Assessment List for Trustworthy Artificial Intelligence in 2020, and the proposal for an EU AI Act in 2021 (High-Level Expert Group on Artificial Intelligence, 2020; European Commission, 2021b).

The continuing advancement of AI in high-risk application areas, such as healthcare, traffic, and finance, and stakeholders' alertness to its potential risks make the effective governance of AI systems necessary in the coming years. The growing awareness of AI risks has thus far yielded numerous guidelines on AI ethics principles (Jobin, Ienca and Vayena, 2019) and increasing regulatory pressure. Aiming to operationalize AI ethics principles, scholars and practitioners have started to discuss organizational and societal AI governance (Dafoe, 2018; Eitel-Porter, 2021; Mäntymäki *et al.*, 2022a; Schneider *et al.*, 2022). Only recently, research has started to converge toward explicit definitions of AI governance (Mäntymäki *et al.*, 2022a). A summary of the current state of the literature reveals that AI governance comprises tools, rules, processes, procedures, and values that aim to ensure the legally compliant and ethically aligned development and use of AI (Winfield and Jirotka, 2018; Butcher and Beridze, 2019; Gahnberg, 2021; Mäntymäki *et al.*, 2022a). While the importance of AI governance has been repeatedly noted (Gasser and Almeida, 2017; Cath, 2018; Butcher and Beridze, 2019; Schmitt, 2021), comprehensive, practice-oriented frameworks for governing AI are few (Benjamins, Barbado and Sierra, 2019; Eitel-Porter, 2021). Collections, reviews, and syntheses of AI ethics principles are in plentiful supply (Jobin, Ienca and Vayena, 2019; Hagendorff, 2020), but the outlines of organizational processes and practices necessary for ensuring responsible AI development are in a nascent state.

Typically, AI governance models touch on particular aspects, such as fairness or transparency (Benjamins, Barbado and Sierra, 2019), and focus on specific stages of system development, such as system design. However, organizations need to govern AI systems over their life cycles and consider the requirements vis-à-vis ethics, legislation, and stakeholders (Laato *et al.*, 2021; Laato, Mäntymäki, *et al.*, 2022). Moreover, as most organizations cannot tackle complex AI governance problems alone, they face two challenges. First, they need to understand the different elements of AI governance including their role in a multi-actor ecosystem for responsible AI (Minkkinen, Zimmer and Mäntymäki, 2023). Second, they must grapple and keep-up with the continuously changing nature of AI governance requirements stemming from ethics and regulation.

To address this difficulty of adequately translating AI ethics principles into organizational AI governance models, *the objective of the paper is to develop meta-requirements of AI governance frameworks for organizations deploying AI to help them translate ethical AI principles into practice and align their operations with the forthcoming European AI Act.* Because translation from AI ethics principles into AI governance is a design problem close to the needs of practitioners, we have adopted a design science research (DSR) approach (Hevner *et al.*, 2004; Peffers *et al.*, 2007; Kuechler and Vaishnavi, 2008). We first put forward the research-based premises for the meta-requirements. We then report the design method that we have employed in an industry-academia research project established to develop an organizational AI governance framework.

We contribute to the IS literature by responding to the calls for actionable tools for organizations deploying AI systems to translate ethical principles to practice (Schiff *et al.*, 2021; Seppälä, Birkstedt

and Mäntymäki, 2021) and advancing the understanding of the components needed to translate AI ethics principles into AI governance (Laato, Tiainen, *et al.*, 2022; Schneider *et al.*, 2022). Our study draws on a DSR study conducted in a research project that included researchers and public and private organizations acting as design partners and domain experts. Thus, we take AI governance research forward from conceptual and exploratory studies into a design-based direction.

The paper proceeds as follows. Section 2 presents the knowledge base, i.e., the streams of literature we build on to establish the research-based premises for developing meta-requirements. The third section describes our DSR approach. The fourth section outlines the meta-requirements. In the fifth section, we discuss the implications, limitations, and future research areas, and conclude the paper.

2 Knowledge Base: Governance of AI Systems

The knowledge base that informed the design process for meta-requirements for AI governance frameworks comprises four streams of literature, which we distil into four research-based premises: AI as an IT artifact category, high-impact and high-risk use cases, AI governance in an organization's governance system, and layers of AI governance.

2.1 AI as an IT Artifact Category

The AI and algorithm studies literature present the key features of AI and other algorithmic technologies that pose AI-specific governance problems (Mittelstadt *et al.*, 2016; Kitchin, 2017; Dignum, 2020; Berente *et al.*, 2021). These features differentiate AI systems from other IT artifacts and, thereby, imply that AI governance requires new approaches to complement existing IT governance frameworks (e.g., Brown and Grant, 2005; Gregory *et al.*, 2018).

The characteristics that differentiate AI systems from other technologies are subject to ongoing debates, and authors tend to emphasize aspects related to inscrutability, fairness, and responsibility (Mittelstadt *et al.*, 2016; Dignum, 2020; Berente *et al.*, 2021). Dignum (2020) and Berente et al. (2021) have argued that autonomy, adaptation through learning, inscrutability, and interactivity are important sources of ethical governance challenges. Our engagement with organizations developing AI applications pushed us towards Mittelstadt et al.'s (2016) account of AI features that, instead, focused on epistemic, normative, and traceability-related concerns as key to understanding the ethics-related features of AI systems.

According to Mittelstadt et al. (2016), epistemic AI ethics concerns arise out of the propensity of AI technologies to base decisions on possibly inconclusive, inscrutable, or misguided evidence. In particular, machine learning and deep learning technologies work by identifying patterns in large datasets and folding them into decision-making algorithms. However, developers often lack certainty over whether the patterns reflect real causal patterns or simply track phantoms. Moreover, the resulting correlational accounts of relationships in data are often uninterpretable and unexplainable, which leads to difficulties for developers to assess system performance and justify decisions. Third, the technologies may detect and perpetuate existing but unacceptable patterns, such as social biases, in the data. In sum, the normative concerns to which Mittelstadt et al. (2016) refer open AI ethics toward the interaction between AI systems and society. Concerning this interaction, AI system outcomes may be unfair either inadvertently or by design. Further, the technologies affect societal affordances, changing the distribution of action and cognitive capabilities. AI systems also affect how action can be traced back to the humans who control or initiate it, typically muddying the waters and making assigning blame and responsibility increasingly difficult.

2.2 High-Impact, High-Risk Use Cases

Because of their capabilities, AI-enabled or AI-assisted decision-making is used in high-impact, high-risk application areas such as finance, healthcare, and traffic. This underscores the importance of ensuring that AI systems deployed by an organization operate according to societal values and norms

and the organization's values. In recent years, there has been a surge of documents published by organizations laying out their ethical principles for AI (Jobin, Ienca and Vayena, 2019; Morley *et al.*, 2020; Koniakou, 2022). However, since ethical principles do not automatically translate into ethical actions, the literature has acknowledged the need to address the so-called translation problem of AI ethics, i.e., how to translate ethical principles into practice. AI governance has been considered as one approach to address this translation problem (Seppälä, Birkstedt and Mäntymäki, 2021; Koniakou, 2022; Mäntymäki *et al.*, 2022b, 2022a). However, there is no universally agreed set of core human or societal values (and their priorities), but the values, social norms, and standards of desirability are culture-specific and time-bound (Awad *et al.*, 2018). In fact, there are different schools of thought regarding ethics (Hagendorff, 2020).

In addition to non-binding ethical principles, there is also binding regulation to mitigate the potential negative impacts of AI. Hence, compliance with the regulation is a self-evident objective for AI governance. This, in turn, necessitates understanding the totality of regulations affecting the use of AI in a specific use case. All in all, while AI has been labeled as a dynamic frontier of computing (Berente et al., 2021), the regulation influencing the use of AI by organizations is developing (Koniakou, 2022). Hence, AI governance – as a solution to the translation problem – needs to tackle both technological and regulatory developments.

2.3 AI Governance in an Organization's Governance System

While characteristics such as autonomy and learning have been present in previous IS, their distinct combination in AI systems poses governance challenges. Due to their autonomy, learning, inscrutability, and interactivity (Berente *et al.*, 2021), AI systems challenge existing notions of IT governance. IT governance is generally defined as ensuring desirable (human) behavior in the use of IT (Weill and Ross, 2005, p. 2). In addition, the application of AI in high-impact use cases and at scale necessitates the consideration of potential risks and harms to individuals and groups.

Moreover, due to their self-learning and self-adaptive nature, AI systems place high demands on data governance, meaning the exercise of authority and control over data management (DAMA International, 2009, p. 19; Abraham, Schneider and vom Brocke, 2019). Further, while the governance of behavior when using IT remains important (Weill and Ross, 2005; Tiwana, Konsynski and Venkatraman, 2013), governance of the learning algorithms and systems is also crucial (Doneda and Almeida, 2016).

As a result, the questions of who governs what and how are potentially more complicated in AI governance compared to IT governance. Moreover, ethical principles, potential harms, and technology-specific regulation are critical in AI governance, which differentiates it from IT governance, where IT infrastructure and strategy are primarily aligned with business needs (e.g., Brown and Grant, 2005). This also has a bearing on any proposed AI governance framework.

2.4 Layers of AI Governance

Due to its complexity, AI governance has been defined in the literature as a multi-layered phenomenon. The AI governance literature outlines different levels of AI governance (e.g., Gasser and Almeida, 2017; Wirtz, Weyerer and Sturm, 2020). The multi-layered nature of governance suggests many potential answers to the questions "What is governed?" and "How?" (cf. Tiwana, Konsynski and Venkatraman, 2013). Researchers have generally suggested that AI governance entails social/ethical, legal, and technical elements (Doneda and Almeida, 2016; Cath, 2018; Butcher and Beridze, 2019).

The AI governance literature includes several models that propose a layered structure of AI governance issues (see Table 1). While the literature presents different sets of layers intended for different purposes (e.g., ethical management, public administration), they share two key commonalities. First, ethics, law, social norms, and technology are incorporated in all frameworks, either explicitly or implicitly. Second, they feature different levels of abstraction on a micro-macro continuum (e.g., specific AI applications, societal norms).

Source	Description	Layers
Brendel et al. (2021)	Framework for the ethical management of AI	Ethical considerations in AI-related managerial decisions Ethical reference frame for managers Consideration of the dimensions of the organizational environment (e.g., stakeholder groups)
Cath (2018)	Guiding forces in AI governance	Technology Ethics Law
Gasser and Almeida (2017)	A model illustrating the interaction between society and AI systems	Social and legal layer (social norms, regulation, legislation) Ethical layer (ethical criteria, principles) Technical layer (data governance, algorithm accountability, standards)
Shneiderman (2020)	Levels of AI governance	Team (software engineering practices) Organization (safety culture) Industry (oversight and trustworthiness certification)
Wirtz, Weyerer, and Sturm (2020)	An integrated AI governance framework for public administration	AI applications and technology AI challenges AI regulation processes Public AI policy Collaborative AI governance

Table 1. Layered Models of AI Governance

The AI governance literature suggests different criteria for the distinct layers. Governance layers can be interpreted as qualitatively different requirements with different logics (Gasser and Almeida, 2017), levels of action and leverage over algorithmic systems (Shneiderman, 2020), and managerial decision-making horizons (Brendel *et al.*, 2021). We can discern the levels involved in organizational AI governance by distinguishing between concentric layers that cut across themes such as ethics, law, and technology. These levels can be seen as different levels of (socio-technical) complexity (Luhmann, 2012; Schneider, Wickert and Marti, 2017).

Based on this concentric approach, we synthesize three layers from the AI governance literature: environment, organization, and AI system (see Figure 1). Shneiderman's (2020) layers of team, organization, and industry come closest to this structure, but we also highlight the AI system as the governed entity. The concentric approach means that organizations with AI systems operate in particular industries and fields (such as healthcare, education, legislation, or finance), each with particular norms, laws, and governance requirements (Davis and Marquis, 2005; Butcher and Beridze, 2019; Martin, 2019). In other words, each organization may use several AI systems, and each operating environment may host numerous organizations. Further, AI systems may challenge different norms in different industries. For example, automated diagnoses in healthcare can raise questions of transparency and accountability in cases of false positives and false negatives (Ho *et al.*, 2019).

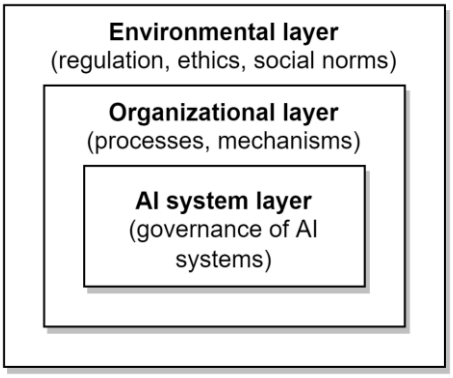

Figure 1. The layered structure of AI governance

This conceptualization of AI governance layers yields a structure whereby the AI system is the concrete governed entity that exists in an organizational context. We conceptualize the AI system as an information technology (IT) artifact that includes AI technologies and is surrounded by a socio-technical system that consists of people, organizations, work systems, and institutions (Dignum, 2019, 2020). The organizational and environmental layers constitute the scaffoldings for AI development, use, and governance. Importantly, all AI governance layers are dynamic and processual, i.e., continuously developing. For example, the European regulatory landscape on AI (environmental layer) is evolving fast, and the final form and enforcement of the EU AI Act, for example, remain to be seen, as well as the connections to other legislative initiatives such as the EU Data Act.

2.5 Research-Based Premises for AI Governance Frameworks

To summarize, we distil the knowledge base into four research-based premises that act as starting points for translating AI ethics principles into governance and provide inputs for developing the meta-requirements (see Table 2). The research-based premises synthesize relevant insights from relevant literature streams that have a bearing on designing AI governance frameworks. They cover the nature of AI as a particular type of IT artifact, the multitude of use cases of AI systems, AI governance as a part of an organization's governance system, and the three central layers of AI governance.

Research-based premise	**Description**
1. AI as an IT artifact category	Unique characteristics of AI systems differentiate AI from other types of information systems (Mittelstadt *et al.*, 2016; Dignum, 2020; Berente *et al.*, 2021).
2. High-impact, high-risk use cases	AI systems deployed in application areas such as healthcare, finance, and traffic call for attention to ethical implications (e.g., Trocin *et al.*, 2021). A multitude of ethical positions and cultural contexts related to the development and deployment of AI systems (Feijóo *et al.*, 2020; Hagendorff, 2020). Uptake of documents outlining ethical principles for AI (Jobin, Ienca and Vayena, 2019; Morley *et al.*, 2020). Developing regulatory landscape (Koniakou, 2022).
3. AI governance in an organization's governance system	Governing AI systems as part of the organization's overall governance system. Intersections between corporate governance, IT governance, data governance, and AI governance. AI governance highlights alignment with social norms and ethical principles (Mäntymäki *et al.*, 2022a; Schneider *et al.*, 2022).
4. Three layers of AI governance	Governance of AI influenced by factors of different levels, including the operating environment, the organization, and the AI system (Gasser and Almeida, 2017; Shneiderman, 2020).

Table 2. Research-based premises for the meta-requirements for AI governance frameworks

3 Research Approach

Because moving from AI ethics principles and research-based premises toward AI governance frameworks is a design problem, we adopted a DSR approach with the overall aim of designing a framework for governing AI systems. This paper presents the design process until the formulation of meta-requirements for AI governance frameworks. In this section, we outline the details of the DSR project. The design process started in August 2020 with the initiation of a two-year research industry-academia project jointly funded by a national research funding agency and the consortium partners.

The project team comprised researchers and a consortium of public and private organizations acting as design partners (see Table 3). Since AI governance presents a multi-layered and complex design challenge, we arranged for the research team and design partners to be transdisciplinary. Consequently, we set up a research team of experts from different research disciplines (e.g., IS, computer science, and law). Similarly, the consortium of design partners involved organizations of different sizes and industries. We drew on this plurality in perspectives throughout the design process to specify meta-requirements for AI governance frameworks.

Consortium partner	Description	Core team members
Alpha (Research team)	Large Public University	2 Professors (IS & Law), 2 Senior Researchers, Research Assistant
Beta (Research team)	Large Public University	Professor (Computer Science), Senior Researcher, Post-doctoral researcher
Gamma (Design partner)	Large consulting company (<1,500 employees) offering strategic consulting, service design, software development, AI, analytics, and cloud and cloud integration services.	Head of Research, Head of Sustainable AI, Business Lead (Data-Driven Business), Insight Lead, Data Scientist, Data Business Designer.
Delta (Design partner)	Small/medium-sized consulting company (<100 employees) offering digital solution design.	Executive Advisor, Sales Director, Principal Consultant.
Epsilon (Design partner)	Large consulting company (<1,000 employees) offering digital strategy, software engineering, and data and intelligence services.	Head of AI and Data Works, Competence Lead, Design Researcher, Service and UX Designer.
Zeta (Design partner)	Small/medium-sized company (<50 employees) offering data and AI strategy, data science, and data architecture services.	Co-founder, Analytics Executive, Chief Data and AI Officer.
Eta (Design partner)	Small/medium-sized company (<50 employees) offering an AI-based cloud service.	Founder, CEO.
Theta (Design partner)	Large (>10,000 employees) financial services provider operating in a high-risk application domain.	Head of AI, Chief Data Scientist, Data Scientist, Legal Counsel
Iota (Design partner)	Large (>5,000 employees) public sector organization operating in a high-risk application domain.	Chief Information Officer, Chief Analyst, Analyst

Table 3. The consortium partners

3.1 The Design Process

The design process can be analytically structured to comprise a sequence of cycles, drawing on Kuechler and Vaishnavi (2008). Accordingly, we can present our design process within five cycles. These are the cycle of (1) problem awareness, (2) design suggestions, (3) development, (4) evaluation, and (5)

conclusion. We refer to this as an analytical structuring since the actual design process was not sequential but iterative. This means we followed a cyclical process of refining the problem awareness and the design but report the process following Kuechler and Vaishnavi (2008) for simplification. To keep the scope of the paper manageable for a conference article and to adhere to the space limitations, we focus on reporting the first cycle of our design process, i.e., problem awareness.

The problem awareness cycle started with the funding application process for the reported research project. In collaboration with the research team and the design partners, the principal investigator prepared a research proposal and funding application to a national Finnish funding agency. The research-based premises were initially outlined in the research proposal included in the funding application.

The research proposal preparation took place in parallel with the EU's white paper on AI (European Commission, 2020). The white paper preceded the EU's AI Act proposal later in April 2021 (European Commission, 2021b). The EU's white paper was a critical event as it indicated that EU-level binding AI regulation was to be introduced. After this point, no one in the project consortium or the funding body questioned the importance of the topic, as it was clear that organizations operating in the EU will need to ensure their compliance with the AI regulation in the making. The project proposal, as well as the research activities, were aligned with EU activities on developing the AI regulation. At the same time, the EU regulatory developments shaped the problem awareness: they underscored the necessity of considering legal compliance throughout the design process compared to the initial funding application.

Besides the need to prepare for compliance with the EU's coming AI regulation, the need to bridge the gap between laying out AI ethics principles and implementing them in practice emerged as a key starting point for the design activities. Through discussions with potential design partners, we realized that there was a consensus among the practitioners involved in the talks that ethical principles are too abstract to provide concrete guidance on implementing responsible AI in practice. Thus, developing an organizational AI governance framework for operationalizing ethics, rules, and principles on AI systems became a key selling point of the funding application and a focal deliverable of the subsequent project.

After securing the research funding, we continued establishing problem awareness through literature reading and discussions with the design partners. A key activity was the specification and articulation of the research-based premises. The purpose of this activity was to increase the research team's knowledge of the literature and provide the design partners with overviews of the research on the project's theme. We first undertook a scoping review of the AI governance literature, including a concept map. The concept map appeared to be a valuable boundary object for discussions with the design partners. These discussions indicated a need to execute a systematic literature review focusing on AI governance at the organizational level.

We established a routine of regular meetings of different formats (e.g., research team meetings, design partner meetings, and workshops). We leveraged these occasions to sound our problem awareness, gained through literature reading, from a practical relevance perspective. Moreover, the discussions with the design partners also directed our reading of the literature.

Drawing on our increasing awareness of the literature and the interactions with the design partners, we formulated the research-based premises into meta-requirements. Meta-requirements present a class of objectives for a design artifact (Jones and Gregor, 2007; Arazy, Kumar and Shapira, 2010), and their formulation precedes the design suggestion and development cycles. In other words, they form prescriptive statements based on justificatory knowledge to guide the artifact's design and evaluation (Lins *et al.*, 2019; Järveläinen, Niemimaa and Zimmer, 2022).

3.2 Data Collection

We followed general recommendations for qualitative data collection. We took notes on the research team meetings and interactions with the design partners. These interactions followed a regular routine, but we also engaged in ad-hoc or planned interactions as necessary. Besides meeting notes, we conducted interviews with the design partners, which we recorded and transcribed. Lastly, we – the research team – kept individual notes. These contained suggestions for the artifact design but also reflections on the design process and the interactions with the design partners. While this data set

focused on the design process, we also systematically kept a record of the designed artifact. This record included illustrations, textual descriptions, linking design features to justificatory knowledge, and a design versioning. Within a spreadsheet, we documented the artifact's different versions and the changes implemented between these versions (vom Brocke, Gau and Mädche, 2021). Figure 2 illustrates the research project's setup and our data collection points within this setup using the DSR framework of Hevner et al. (2004).

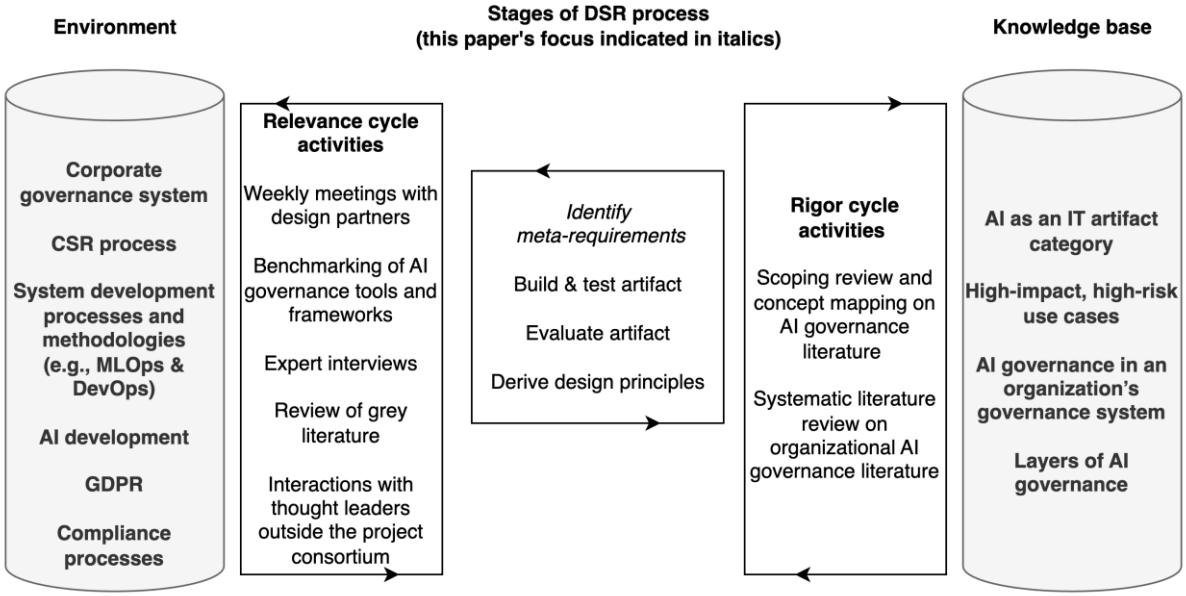

Figure 2. The relevance and rigor cycle activities to enable identifying the meta-requirements for AI governance frameworks based on the environment and knowledge base (adapted from Hevner et al., 2004)

3.3 Data Analysis

We drew from both the knowledge base and the environment to identify and evaluate the meta-requirements. We analyzed the collected data during the design process and after its completion. During the process, we analyzed our notes from the regular meetings, workshops, and interviews within the research team and discussed with the design partners. Since this analysis occurred on the fly, we did not systematically code the collected data but screened and structured them for suggestions or hints on designing AI governance frameworks. In the problem awareness cycle, we specifically focused on identifying requirements that emerged from three analytical activities: (1) discussions with the consortium partners, (2) synthesizing the literature, and (3) relating these two sources of justificatory knowledge. After completing the design process, we used the collected data to reconstruct the design process. We chronologically listed the events and the artifact's design versions using the kept record. We enriched this listing with our notes on critical insights from the interactions within the research team and with the design partner and notes on major changes to the design artifact. Considering the research team's transdisciplinary composition and the diversity within the consortium of design partners, the meta-requirements emerged from the multiplicity of interactions within the presented DSR project. Next, we outline the formulated meta-requirements for AI governance frameworks.

4 Formulating Meta-Requirements for AI Governance Frameworks

As a result of the problem awareness cycle in our DSR process, we formulated seven meta-requirements that AI governance frameworks need to fulfill.

MR1: Accommodate the characteristics of AI. Any suitable AI governance framework should deal with the epistemic, normative, and traceability concerns (Mittelstadt *et al.*, 2016; Dignum, 2020; Berente *et al.*, 2021) that we discussed in the knowledge base section. A governance framework needs to specifically address the risks and challenges brought by inconclusive, inscrutable, and misguided evidence created by machine learning technologies that distinguish AI governance from governing and managing any kind of IT system. Addressing the problems requires, for example, analyzing system data categories, inferences, and proxies biases stemming from historical learning data (Martin, 2019) and devising explainability strategies that justify decisions to affected parties (Laato, Tiainen, *et al.*, 2022).

MR2: Position AI governance in the organization's overall governance system. AI governance enters a crowded governance landscape with corporate governance, IT governance, and data governance already dealing with issues such as accountability, decision rights, and compliance with data regulations (Mäntymäki *et al.*, 2022a). An AI governance framework should be designed to avoid redundancy and bring added value. Therefore, such frameworks should position AI governance in existing governance domains within an organization's governance system.

MR3: Map the relevant regulatory landscape and update when regulation changes. In addition to organizations' governance systems, AI governance is positioned within a regulatory landscape that includes general regulation, such as the GDPR, and sectoral legislation, such as healthcare-specific regulations (Viljanen and Parviainen, 2022). An actionable governance framework should take into account the boundaries set by the regulatory landscape, such as provisions for so-called high-risk AI systems (European Commission, 2021b). Moreover, a governance framework has to deal with the fact that legislation and its enforcement are moving targets, and thus AI governance should be adaptable to changing requirements and not too tightly coupled to regulation at a particular point in time.

MR4: Address the translation problem of AI ethics. The AI ethics literature recognizes the inadequacy of ethical principles alone in governing the risks brought by AI systems. Hence, AI ethics principles need to be translated into practicable governance mechanisms (Mittelstadt, 2019; Morley *et al.*, 2020; Schiff *et al.*, 2021). An effective AI governance framework should incorporate ethical principles and their translation into more practical mechanisms and processes.

MR5: Incorporate the multi-stakeholder nature of governing AI systems in an organization and the stakeholders' requirements. According to the AI governance literature, no single organizational stakeholder governs AI systems, but rather, responsibilities are often shared in different arrangements and even complex networks of accountability (Orr and Davis, 2020; Shneiderman, 2020; Seppälä, Birkstedt and Mäntymäki, 2021). Moreover, relevant stakeholders reach beyond the focal organization and include, for example, customers, affected individuals, and investors (Stahl *et al.*, 2021; Minkkinen, Niukkanen and Mäntymäki, 2022). Therefore, an AI governance framework should take into account this intra- and inter-organizational set of actors that articulate AI governance requirements and play their parts in responding to them.

MR6: Acknowledge the multitude of ethical viewpoints and cultural contexts with different value systems. The AI ethics literature includes multiple ethical approaches, such as consequentialism and deontology, and a long list of principles, such as fairness, transparency, accountability, and privacy (Jobin, Ienca and Vayena, 2019; Hagendorff, 2020). In addition, AI development and use are global phenomena, meaning that AI systems are used in different cultural contexts, such as the United States, China, and Europe (Feijóo *et al.*, 2020). Therefore, an AI governance framework has to consider the diverse ethical perspectives and cultural norms and adopt a sufficiently value-agnostic stance, while respecting certain fundamental rights. In practice, this could mean a stable core of shared principles, such as non-discrimination according to gender or ethnicity, while leaving space for different priorities in different cultural contexts. The phenomenon of Islamic banking could provide one comparison point.

MR7: Integrate with the organization's AI system development and operations processes. AI systems need to be governed throughout their lifecycles (Laato, Birkstedt, *et al.*, 2022). For AI governance to reach the operational level effectively, it needs to be integrated with organizations' AI system development and operations processes, which may involve methods such as agile development and operations (DevOps) (Gall and Pigni, 2021).

Table 4 summarizes the meta-requirements, and the final column connects them to the knowledge base and environment, including the research-based premises (RPs, see Table 2 in section 2.5).

Meta-requirement	**Description**	**Rationale (connection to the knowledge base and environment)**
1. Accommodate the characteristics of AI	A governance framework should deal with the unique epistemic, normative, and traceability concerns in contemporary AI systems.	A mismatch between existing IT governance frameworks in the knowledge base (RP3) and the AI governance literature, which details the novel characteristics of AI systems (RP1).
2. Position AI governance in the organization's overall governance system	A governance framework should consider AI governance in a complex organizational setting with numerous interlinked governance areas.	Literature streams on corporate governance, IT governance, and data governance highlight the importance of these governance fields (RP3). Design partners particularly emphasized the link to data governance and data management.
3. Map the relevant regulatory landscape and update based on changes in the regulation	A governance framework should be up to date with relevant regulatory developments concerning AI and sectoral AI applications.	GDPR is a prominent part of the environment, and sectoral legislation covers areas such as healthcare and finance (RP2, RP4). The EU AI policy process was strongly developing during the time of the framework design.
4. Address the translation problem of AI ethics	A governance framework should indicate how high-level ethical principles and requirements can be translated into the operational governance of AI systems.	The AI ethics literature extensively discusses the translation problem of AI ethics (RP4). Design partners and interviewed practitioners corroborated the need for practical tools.
5. Incorporate the multi-stakeholder nature of governing AI systems in an organization and stakeholders' requirements	A governance framework should deal with AI governance as an inherently multi-stakeholder set of activities, including intra-organizational, inter-organizational, and stakeholder engagement components.	The literature highlights the multi-stakeholder nature of AI ethics and accountability (RP3, RP4). Design partners and interviewed practitioners indicated many potential organizational arrangements for AI governance.
6. Acknowledge the multitude of ethical viewpoints and cultural contexts with different value systems	A governance framework should consider and arbitrate in a responsible way between different ethical viewpoints and value systems	The AI ethics literature articulates the complexity of ethical perspectives (RP2, RP4). Design partners conducted end-user workshops, which corroborated the complexity of different sets of values.
7. Integrate with the organization's AI system development and operations processes	A governance framework should incorporate AI governance on the level of technical activities for designing, developing, and operating AI systems.	The knowledge base and technical design partners corroborated the necessity of integration with development and operations processes (RP3, RP4).

Table 4. Meta-requirements for an AI governance framework

5 Discussion and Conclusion

5.1 Implications for IS Research and Practice

The current paper contributes design knowledge in the form of meta-requirements for establishing AI governance frameworks in organizations. In doing so, we also contribute to the emerging IS research on AI governance (Seppälä, Birkstedt and Mäntymäki, 2021; Minkkinen, Zimmer and Mäntymäki, 2023) and responsible AI (Trocin *et al.*, 2021; Zimmer, Minkkinen and Mäntymäki, 2022). As its chief contribution, this paper collates the literature and expert knowledge on AI governance and puts forward meta-requirements for designing AI governance frameworks. The meta-requirements are an intermediate step in translating AI ethics principles into fully developed AI governance frameworks. Thus, the paper takes the IS research forward from conceptual papers (Schneider *et al.*, 2022), principle-based research frameworks (Thiebes, Lins and Sunyaev, 2021), and explorative studies (Papagiannidis *et al.*, 2022) into a design-based direction by specifying the meta-requirements that act as foundations for practicable AI governance. In addition to AI governance, this research stream can contribute to design theory on developing governance frameworks more broadly.

The meta-requirements also offer practitioners a first step toward developing AI governance fit for the current generation of AI systems. Moreover, the research-based starting points we have derived from different bodies of literature can help practitioners involved in developing AI governance tools and processes in understanding the potential interfaces and linkages between AI governance and other organizational processes in the overall governance systems of an organization.

As an implication for both research and practice, we highlight the role of organizations as mediators between regulatory and ethical AI requirements, on the one hand, and the design, development, and use of algorithmic systems, on the other hand. Organizational actors, such as managers, heads of AI, and internal responsible AI boards, act as translators of ethical AI requirements into practice. Due to this translating role, they play a key part in ensuring that AI systems work in a socially responsible manner. The need for operationalizing ethical and human-centric AI has been repeatedly articulated (Morley et al., 2020; Seppälä, Birkstedt and Mäntymäki, 2021). Numerous responsible AI initiatives have also been conducted, most notably in the EU (e.g., High-Level Expert Group on Artificial Intelligence, 2019; European Commission, 2021a; European Parliament, 2022). Nevertheless, significant gaps remain in the practical implementation of AI governance, and the meta-requirements outline key focus areas for organizations and organizational researchers, such as the integration of AI governance into governance systems and development and operations processes. AI governance should not focus on restrictions and creating unresolvable unease for organizations. Instead, it should enable organizations to use AI systems in alignment with organizational objectives, values, and ethical AI principles.

A general implication for the design of governance frameworks is that frameworks need to incorporate the distinct characteristics of the governed entity (such as AI systems), the contexts in which the entity is embedded (e.g., an organization's governance systems, multi-stakeholder networks), and the different sources of requirements (e.g., regulation, AI ethics principles). The characteristics of the governed entity influence both the possibilities and challenges of governance. This can be seen in the case of learning AI systems, where machine learning can pose challenges but can also be used as a leverage point in governance. The governance systems and networks in which governed entities are embedded also define the boundary conditions of effective governance, and the different kinds of requirements necessitate suitably elaborate governance frameworks that tackle different types of inputs.

5.2 Limitations

As an initial design approach to the multi-faceted topic of AI governance, the current study has two evident limitations that need to be acknowledged. First, with this paper, we deliberately focused on meta-requirements, that is, generic objectives that a design artifact needs to fulfill. We provide prescriptions for designing an artifact (design knowledge), but questions about the respective artifacts and the design process are left unanswered until later stages in the design process. Second, the

organizational framing of AI governance means that questions of regulatory development and broad societal debates on AI governance are taken as requirements from the external environment rather than aspects to be influenced through design intervention. However, future developments might elaborate on societal AI governance mechanisms. On the organizational level, we provide meta-requirements stemming from a synthesis of the literature, dialogues with design partners, and expert interviews rather than an implementable organizational AI governance framework.

5.3 Future Research Directions

We discuss five promising areas of inquiry concerning future research directions. The first and most direct future research direction is developing a governance framework that fulfills the meta-requirements. Ultimately, this will also facilitate the creation of measurement instruments, such as key performance indicators (KPIs), to assess organizational performance. In subsequent work, maturity levels of different dimensions of AI governance could be specified (Shneiderman, 2020), which would provide an overview of an organization's AI governance readiness for managers, investors, and other stakeholders (cf. Jöhnk, Weißert and Wyrtki, 2020; Minkkinen, Niukkanen and Mäntymäki, 2022).

Second, designing AI governance frameworks also provides a starting point for AI auditing frameworks. Literature on AI and algorithmic auditing is emerging, and no framework has yet been firmly established (Raji *et al.*, 2020; Koshiyama *et al.*, 2021; Minkkinen, Laine and Mäntymäki, 2022). Although delineating the relevant issues of AI auditing is beyond the scope of this paper, one starting point could be to design an AI auditing framework at least partly based on the same meta-requirements and as a derivative from the design of an AI governance framework.

Third, articulating the research-based premises and meta-requirements can facilitate qualitative research to understand the human and non-human elements, processes, and mechanisms involved in organizational AI governance. For example, studies could examine how organizations govern AI systems, how (and to what extent) strategic and value alignment occurs, and what incipient or more fully developed governance mechanisms are employed (Seppälä, Birkstedt and Mäntymäki, 2021; Stahl *et al.*, 2021). Comparative studies could also provide insights into the industry, sector, and regional differences. For example, differences between highly regulated areas (such as medicine) and less ethically sensitive areas (such as manufacturing) could be explored.

Fourth, concerning contextual differences, different risks and concomitant levels of AI governance are essential topics for subsequent research, especially with the coming risk-based EU AI regulation (European Commission, 2021b). Different AI governance requirements and mechanisms could be triggered by analyzing the risk environment around using a particular AI system. For example, life-critical areas such as healthcare will probably produce more stringent governance requirements than systems that provide consumers with product recommendations.

Fifth, we raise the issue of multi-actor networks. This is because companies increasingly offer AI as a service instead of discrete products (Kozuka, 2019; Javadi *et al.*, 2020). How do AI governance and accountability work in multi-actor settings, for example, where one company develops an AI system, another company uses it, and a third company audits the system (cf. Minkkinen, Zimmer and Mäntymäki, 2023)? We can assume that inter-organizational AI governance will become important as organizational boundaries become blurred. The chains of governance and accountability in employing AI systems require further study to complement the work on organizational AI governance.

References

Abraham, R., Schneider, J. and vom Brocke, J. (2019) 'Data governance: A conceptual framework, structured review, and research agenda', *International Journal of Information Management*, 49, pp. 424–438. Available at: https://doi.org/10.1016/j.ijinfomgt.2019.07.008.

Altman, M., Wood, A. and Vayena, E. (2018) 'A Harm-Reduction Framework for Algorithmic Fairness', *IEEE Security Privacy*, 16(3), pp. 34–45. Available at: https://doi.org/10.1109/MSP.2018.2701149.

Arazy, O., Kumar, N. and Shapira, B. (2010) 'A Theory-Driven Design Framework for Social Recommender Systems', *Journal of the Association for Information Systems*, 11(9). Available at: https://doi.org/10.17705/1jais.00237.

Awad, E. *et al.* (2018) 'The Moral Machine experiment', *Nature*, 563(7729), pp. 59–64. Available at: https://doi.org/10.1038/s41586-018-0637-6.

Bechmann, A. and Bowker, G.C. (2019) 'Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media', Big Data & Society, 6(1), p. 2053951718819569. Available at: https://doi.org/10.1177/2053951718819569.

Benjamins, R., Barbado, A. and Sierra, D. (2019) 'Responsible AI by design in practice', arXiv: Computers and Society [Preprint]. Available at: https://arxiv.org/abs/1909.12838v2 (Accessed: 14 December 2020).

Berente, N. et al. (2021) 'Managing artificial intelligence', MIS Quarterly, 45(3), pp. 1433–1450.

Bolukbasi, T. et al. (2016) 'Man is to computer programmer as woman is to homemaker? debiasing word embeddings', in *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc. (NIPS'16), pp. 4356–4364.

Brendel, A.B. *et al.* (2021) 'Ethical management of artificial intelligence', *Sustainability*, 13(4), p. 1974. Available at: https://doi.org/10.3390/su13041974.

vom Brocke, J., Gau, M. and Mädche, A. (2021) 'Journaling the Design Science Research Process. Transparency About the Making of Design Knowledge', in L. Chandra Kruse, S. Seidel, and G.I. Hausvik (eds) *The Next Wave of* Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 131–136. Available at: https://doi.org/10.1007/978-3-030-82405-1_15.

Brown, A.E. and Grant, G.G. (2005) 'Framing the frameworks: A review of IT governance research', *Communications of the Association for Information Systems*, 15(1). Available at: https://doi.org/10.17705/1CAIS.01538.

Butcher, J. and Beridze, I. (2019) 'What is the state of artificial intelligence governance globally?', *The RUSI Journal*, 164(5–6), pp. 88–96. Available at: https://doi.org/10.1080/03071847.2019.1694260.

Cath, C. (2018) 'Governing artificial intelligence: ethical, legal and technical opportunities and challenges', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133). Available at: https://doi.org/10.1098/rsta.2018.0080.

Crawford, K. and Calo, R. (2016) 'There is a blind spot in AI research', *Nature News*, 538(7625), p. 311. Available at: https://doi.org/10.1038/538311a.

Dafoe, A. (2018) *AI Governance: A Research Agenda*. Future of Humanity Institute, University of Oxford: Future of Humanity Institute.

DAMA International (2009) *The DAMA Guide to the Data Management Body of Knowledge - DAMA-DMBOK*. Denville, NJ, USA: Technics Publications, LLC.

Davis, G.F. and Marquis, C. (2005) 'Prospects for Organization Theory in the Early Twenty-First Century: Institutional Fields and Mechanisms', *Organization Science*, 16(4), pp. 332–343. Available at: https://doi.org/10.1287/orsc.1050.0137.

Dignum, V. (2018) 'Ethics in artificial intelligence: introduction to the special issue', *Ethics and Information Technology*, 20(1), pp. 1–3. Available at: https://doi.org/10.1007/s10676-018-9450-z.

Dignum, V. (2019) *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer International Publishing (Artificial Intelligence: Foundations, Theory, and Algorithms). Available at: https://doi.org/10.1007/978-3-030-30371-6.

Dignum, V. (2020) 'Responsibility and artificial intelligence', in Dignum, V., *The Oxford handbook of ethics of AI*. Edited by M. D. Dubber, F. Pasquale, and S. Das. Oxford University Press, pp. 213–231. Available at: https://doi.org/10.1093/oxfordhb/9780190067397.013.12.

Doneda, D. and Almeida, V.A.F. (2016) 'What is algorithm governance?', *IEEE Internet Computing*, 20(4), pp. 60–63. Available at: https://doi.org/10.1109/MIC.2016.79.

Eitel-Porter, R. (2021) 'Beyond the promise: Implementing ethical AI', *AI and Ethics*, 1(1), pp. 73–80. Available at: https://doi.org/10.1007/s43681-020-00011-6.

Etzioni, A. and Etzioni, O. (2017) 'Incorporating Ethics Into Artificial Intelligence', *The Journal of Ethics*, 21(4), pp. 403–418. Available at: https://doi.org/10.1007/s10892-017-9252-2.

European Commission (2020) *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust*.

European Commission (2021a) *A European approach to artificial intelligence, Shaping Europe's digital future*. Available at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (Accessed: 19 January 2022).

European Commission (2021b) *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS COM/2021/206 final*. Available at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence (Accessed: 4 May 2021).

European Parliament (2022) *Artificial intelligence in a digital age*. Strasbourg. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_EN.html (Accessed: 27 October 2022).

Feijóo, C. *et al.* (2020) 'Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy', *Telecommunications Policy*, 44(6), p. 101988. Available at: https://doi.org/10.1016/j.telpol.2020.101988.

Fjeld, J. *et al.* (2020) 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI', *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.3518482.

Floridi, L. (2019) 'Establishing the rules for building trustworthy AI', *Nature Machine Intelligence*, 1(6), pp. 261–262. Available at: https://doi.org/10.1038/s42256-019-0055-y.

Gahnberg, C. (2021) 'What rules? Framing the governance of artificial agency', *Policy and Society*, 40(2), pp. 194–210. Available at: https://doi.org/10.1080/14494035.2021.1929729.

Gall, M. and Pigni, F. (2021) 'Taking DevOps mainstream: A critical review and conceptual framework', *European Journal of Information Systems*, 0(0), pp. 1–20. Available at: https://doi.org/10.1080/0960085X.2021.1997100.

Gasser, U. and Almeida, V.A.F. (2017) 'A layered model for AI governance', *IEEE Internet Computing*, 21(6), pp. 58–62. Available at: https://doi.org/10.1109/MIC.2017.4180835.

Gregory, R.W. *et al.* (2018) 'IT consumerization and the transformation of IT governance', *MIS Quarterly*, 42(4), pp. 1225–1253.

Hagendorff, T. (2020) 'The ethics of AI ethics: An evaluation of guidelines', *Minds and Machines*, 30(1), pp. 99–120. Available at: https://doi.org/10.1007/s11023-020-09517-8.

Hevner, A.R. *et al.* (2004) 'Design science in information systems research', *MIS Quarterly*, 28(1), pp. 75–105. Available at: https://doi.org/10.2307/25148625.

High-Level Expert Group on Artificial Intelligence (2019) *Ethics Guidelines for Trustworthy AI*. European Commission. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

High-Level Expert Group on Artificial Intelligence (2020) *The assessment list for trustworthy artificial intelligence (ALTAI)*.

Ho, C.W.L. *et al.* (2019) 'Governance of automated image analysis and artificial intelligence analytics in healthcare', *Clinical Radiology*, 74(5), pp. 329–337. Available at: https://doi.org/10.1016/j.crad.2019.02.005.

Järveläinen, J., Niemimaa, M. and Zimmer, M. (2022) 'Designing a thrifty approach for SME business continuity: practices for transparency of the design process', *Journal of the Association for Information Systems*, 23(6). Available at: https://doi.org/10.17705/1jais.00771.

Javadi, S.A. *et al.* (2020) 'Monitoring Misuse for Accountable "Artificial Intelligence as a Service"', in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. *AIES '20: AAAI/ACM Conference on AI, Ethics, and Society*, New York NY USA: ACM, pp. 300–306. Available at: https://doi.org/10.1145/3375627.3375873.

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389–399. Available at: https://doi.org/10.1038/s42256-019-0088-2.

Jöhnk, J., Weißert, M. and Wyrtki, K. (2020) 'Ready or not, AI comes—An interview study of organizational AI readiness factors', *Business & Information Systems Engineering*, 63(1), pp. 5–20. Available at: https://doi.org/10.1007/s12599-020-00676-7.

Jones, D. and Gregor, S. (2007) 'The anatomy of a design theory', *Journal of the Association for Information Systems*, 8(5). Available at: https://doi.org/10.17705/1jais.00129.

Kaminski, M.E. (2019) 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability', *Southern California Law Review*, 92(6), pp. 1529–1616.

Kazim, E., Denny, D.M.T. and Koshiyama, A. (2021) 'AI auditing and impact assessment: according to the UK information commissioner's office', *AI and Ethics* [Preprint]. Available at: https://doi.org/10.1007/s43681-021-00039-2.

Kitchin, R. (2017) 'Thinking critically about and researching algorithms', *Information, Communication & Society*, 20(1), pp. 14–29. Available at: https://doi.org/10.1080/1369118X.2016.1154087.

Koniakou, V. (2022) 'From the "rush to ethics" to the "race for governance" in Artificial Intelligence', *Information Systems Frontiers* [Preprint]. Available at: https://doi.org/10.1007/s10796-022-10300-6.

Koshiyama, A. *et al.* (2021) *Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*. SSRN Scholarly Paper ID 3778998. Rochester, NY: Social Science Research Network. Available at: https://doi.org/10.2139/ssrn.3778998.

Kozuka, S. (2019) 'A governance framework for the development and use of artificial intelligence: lessons from the comparison of Japanese and European initiatives', *Uniform Law Review*, 24(2), pp. 315–329. Available at: https://doi.org/10.1093/ulr/unz014.

Kuechler, B. and Vaishnavi, V. (2008) 'On theory development in design science research: anatomy of a research project', *European Journal of Information Systems*, 17(5), pp. 489–504. Available at: https://doi.org/10.1057/ejis.2008.40.

Laato, S. *et al.* (2021) 'Digital Transformation of Software Development: Implications for the Future of Work', in D. Dennehy et al. (eds) *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 609–621. Available at: https://doi.org/10.1007/978-3-030-85447-8_50.

Laato, S., Birkstedt, T., *et al.* (2022) 'AI governance in the system development life cycle: Insights on responsible machine learning engineering', in *Proceedings of the 1st Conference on AI Engineering - Software Engineering for AI*. *CAIN*, New York, N.Y.: ACM. Available at: https://doi.org/10.1145/3522664.3528598.

Laato, S., Tiainen, M., *et al.* (2022) 'How to explain AI systems to end users: A systematic literature review and research agenda', *Internet Research*, 32(7). Available at: https://doi.org/10.1108/INTR-08-2021-0600.

Laato, S., Mäntymäki, M., *et al.* (2022) 'Trends and trajectories in the software industry: Implications for the future of work', *Information Systems Frontiers* [Preprint]. Available at: https://doi.org/10.1007/s10796-022-10267-4.

Lin, T.C.W. (2019) 'Artificial Intelligence, Finance, and the Law', *Fordham L. Rev.*, 88, p. 531.

Lins, S. *et al.* (2019) 'Designing Monitoring Systems for Continuous Certification of Cloud Services: Deriving Meta-requirements and Design Guidelines', *Communications of the Association for Information Systems*, 44(1). Available at: https://doi.org/10.17705/1CAIS.04425.

Luhmann, N. (2012) *Introduction to Systems Theory*. Edited by D. Baecker. Translated by P. Gilgen. Cambridge: Polity.

Mäntymäki, M. *et al.* (2022a) 'Defining organizational AI governance', *AI and Ethics*, 2, pp. 603–609. Available at: https://doi.org/10.1007/s43681-022-00143-x.

Mäntymäki, M. *et al.* (2022b) *Putting AI ethics into practice: the hourglass model of organizational AI governance*. arXiv:2206.00335. arXiv. Available at: https://doi.org/10.48550/arXiv.2206.00335.

Martin, K. (2019) 'Ethical implications and accountability of algorithms', *Journal of Business Ethics*, 160(4), pp. 835–850. Available at: https://doi.org/10.1007/s10551-018-3921-3.

Mikalef, P. *et al.* (2022) 'Thinking responsibly about responsible AI and "the dark side" of AI', *European Journal of Information Systems*, 0(0), pp. 1–12. Available at: https://doi.org/10.1080/0960085X.2022.2026621.

Minkkinen, M., Niukkanen, A. and Mäntymäki, M. (2022) 'What about investors? ESG analyses as tools for ethics-based AI auditing', *AI & SOCIETY* [Preprint]. Available at: https://doi.org/10.1007/s00146-022-01415-0.

Minkkinen, M., Zimmer, M.P. and Mäntymäki, M. (2023) 'Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame', *Information Systems Frontiers*, 25, pp. 103–121. Available at: https://doi.org/10.1007/s10796-022-10269-2.

Mittelstadt, B. (2019) 'Principles alone cannot guarantee ethical AI', *Nature Machine Intelligence*, 1(11), pp. 501–507. Available at: https://doi.org/10.1038/s42256-019-0114-4.

Mittelstadt, B.D. *et al.* (2016) 'The ethics of algorithms: Mapping the debate', *Big Data and Society*, 3(2). Available at: https://doi.org/10.1177/2053951716679679.

Morley, J. *et al.* (2020) 'From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices', *Science and Engineering Ethics*, 26(4), pp. 2141–2168. Available at: https://doi.org/10.1007/s11948-019-00165-5.

Müller, V.C. (2020) 'Ethics of Artificial Intelligence and Robotics', in E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*. Palo Alto, Cal.: CSLI, Stanford University, pp. 1–70.

O'Neil, C. (2016) *Weapons of math destruction: how big data increases inequality and threatens democracy*. First edition. New York: Crown.

Orr, W. and Davis, J.L. (2020) 'Attributions of ethical responsibility by Artificial Intelligence practitioners', *Information, Communication & Society*, 23(5), pp. 719–735. Available at: https://doi.org/10.1080/1369118X.2020.1713842.

Papagiannidis, E. *et al.* (2022) 'Toward AI governance: Identifying best practices and potential barriers and outcomes', *Information Systems Frontiers* [Preprint]. Available at: https://doi.org/10.1007/s10796-022-10251-y.

Peffers, K. *et al.* (2007) 'A Design Science Research Methodology for Information Systems Research', *Journal of Management Information Systems*, 24(3), pp. 45–77. Available at: https://doi.org/10.2753/MIS0742-1222240302.

Raji, I.D. *et al.* (2020) 'Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing', in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery (FAT* '20), pp. 33–44. Available at: https://doi.org/10.1145/3351095.3372873.

Robles Carrillo, M. (2020) 'Artificial intelligence: From ethics to law', *Telecommunications Policy*, 44(6), p. 101937. Available at: https://doi.org/10.1016/j.telpol.2020.101937.

Russell, S.J. and Norvig, P. (2021) *Artificial intelligence: a modern approach*. Fourth edition. Hoboken: Pearson (Pearson series in artificial intelligence).

Schiff, D. *et al.* (2021) 'Explaining the principles to practices gap in AI', *IEEE Technology and Society Magazine*, 40(2), pp. 81–94. Available at: https://doi.org/10.1109/MTS.2021.3056286.

Schmitt, L. (2021) 'Mapping global AI governance: A nascent regime in a fragmented landscape', *AI and Ethics* [Preprint]. Available at: https://doi.org/10.1007/s43681-021-00083-y.

Schneider, A., Wickert, C. and Marti, E. (2017) 'Reducing Complexity by Creating Complexity: A Systems Theory Perspective on How Organizations Respond to Their Environments', *Journal of Management Studies*, 54(2), pp. 182–208. Available at: https://doi.org/10.1111/joms.12206.

Schneider, J. *et al.* (2022) 'Artificial intelligence governance for businesses', *Information Systems Management*, pp. 1–21. Available at: https://doi.org/10.1080/10580530.2022.2085825.

Seppälä, A., Birkstedt, T. and Mäntymäki, M. (2021) 'From ethical AI principles to governed AI', in *Proceedings of the 42nd International Conference on Information Systems (ICIS2021)*. *International Conference on Information Systems (ICIS)*, Austin, Texas. Available at: https://aisel.aisnet.org/icis2021/ai_business/ai_business/10/.

Shneiderman, B. (2020) 'Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems', *ACM Transactions on Interactive Intelligent Systems*, 10(4). Available at: https://doi.org/10.1145/3419764.

Stahl, B.C. *et al.* (2021) 'Organisational responses to the ethical issues of artificial intelligence', *AI & SOCIETY* [Preprint]. Available at: https://doi.org/10.1007/s00146-021-01148-6.

Thiebes, S., Lins, S. and Sunyaev, A. (2021) 'Trustworthy artificial intelligence', *Electronic Markets*, 31(2), pp. 447–464. Available at: https://doi.org/10.1007/s12525-020-00441-4.

Tiwana, A., Konsynski, B. and Venkatraman, N. (2013) 'Special Issue: Information Technology and Organizational Governance: The IT Governance Cube', *Journal of Management Information Systems*, 30(3), pp. 7–12. Available at: https://doi.org/10.2753/MIS0742-1222300301.

Trocin, C. *et al.* (2021) 'Responsible AI for digital health: A synthesis and a research agenda', *Information Systems Frontiers* [Preprint]. Available at: https://doi.org/10.1007/s10796-021-10146-4.

Vakkuri, V., Kemell, K.-K. and Abrahamsson, P. (2020) 'ECCOLA - a Method for Implementing Ethically Aligned AI Systems', in. *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 195–204. Available at: https://doi.org/10.1109/SEAA51224.2020.00043.

Veale, M. and Binns, R. (2017) 'Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data', *Big Data & Society*, 4(2), p. 2053951717743530. Available at: https://doi.org/10.1177/2053951717743530.

Viljanen, M. and Parviainen, H. (2022) 'AI applications and regulation: Mapping the regulatory strata', *Frontiers in Computer Science*, 3. Available at: https://www.frontiersin.org/article/10.3389/fcomp.2021.779957 (Accessed: 18 January 2022).

Wallach, W. and Marchant, G. (2019) 'Toward the Agile and Comprehensive International Governance of AI and Robotics [point of view]', *Proceedings of the IEEE*, 107(3), pp. 505–508. Available at: https://doi.org/10.1109/JPROC.2019.2899422.

Weill, P. and Ross, J. (2005) 'A Matrixed Approach to Designing IT Governance', *MIT Sloan Management Review*, 46(2), pp. 26–34.

Winfield, A.F.T. and Jirotka, M. (2018) 'Ethical governance is essential to building trust in robotics and artificial intelligence systems', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), p. 20180085. Available at: https://doi.org/10.1098/rsta.2018.0085.

Wirtz, B.W., Weyerer, J.C. and Sturm, B.J. (2020) 'The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration', *International Journal of Public Administration*, 43(9), pp. 818–829. Available at: https://doi.org/10.1080/01900692.2020.1749851.

Zimmer, M.P., Minkkinen, M. and Mäntymäki, M. (2022) 'Responsible artificial intelligence systems: Critical considerations for business model design', *Scandinavian Journal of Information Systems*, 34(2).