AI governance: themes, knowledge gaps and future agendas

Teemu Birkstedt, Matti Minkkinen, Anushree Tandon and Matti Mäntymäki
Turku School of Economics, University of Turku, Turku, Finland

Received 13 January 2022
Revised 1 September 2022
25 January 2023
15 May 2023
Accepted 19 May 2023

Abstract

Purpose – Following the surge of documents laying out organizations' ethical principles for their use of artificial intelligence (AI), there is a growing demand for translating ethical principles to practice through AI governance (AIG). AIG has emerged as a rapidly growing, yet fragmented, research area. This paper synthesizes the organizational AIG literature by outlining research themes and knowledge gaps as well as putting forward future agendas.

Design/methodology/approach – The authors undertake a systematic literature review on AIG, addressing the current state of its conceptualization and suggesting future directions for AIG scholarship and practice. The review protocol was developed following recommended guidelines for systematic reviews and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

Findings – The results of the authors' review confirmed the assumption that AIG is an emerging research topic with few explicit definitions. Moreover, the authors' review identified four themes in the AIG literature: technology, stakeholders and context, regulation and processes. The central knowledge gaps revealed were the limited understanding of AIG implementation, lack of attention to the AIG context, uncertain effectiveness of ethical principles and regulation, and insufficient operationalization of AIG processes. To address these gaps, the authors present four future AIG agendas: technical, stakeholder and contextual, regulatory, and process. Going forward, the authors propose focused empirical research on organizational AIG processes, the establishment of an AI oversight unit and collaborative governance as a research approach.

Research limitations/implications – To address the identified knowledge gaps, the authors present the following working definition of AIG: AI governance is a system of rules, practices and processes employed to ensure an organization's use of AI technologies aligns with its strategies, objectives, and values, complete with legal requirements, ethical principles and the requirements set by stakeholders. Going forward, the authors propose focused empirical research on organizational AIG processes, the establishment of an AI oversight unit and collaborative governance as a research approach.

Practical implications – For practitioners, the authors highlight training and awareness, stakeholder management and the crucial role of organizational culture, including senior management commitment.

Social implications – For society, the authors review elucidates the multitude of stakeholders involved in AI governance activities and complexities related to balancing the needs of different stakeholders.

Originality/value – By delineating the AIG concept and the associated research themes, knowledge gaps and future agendas, the authors review builds a foundation for organizational AIG research, calling for broad contextual investigations and a deep understanding of AIG mechanisms. For practitioners, the authors highlight training and awareness, stakeholder management and the crucial role of organizational culture, including senior management commitment.

Keywords Artificial intelligence, AI, AI governance, AI ethics, Governance

Paper type Literature review

The research reported here was conducted within the Artificial Intelligence Governance and Auditing (AIGA) project (https://ai-governance.eu), funded by Business Finland's AI Business program.

Internet Research
Vol. 33 No. 7, 2023
pp. 133-167
Emerald Publishing Limited
1066-2243
DOI 10.1108/INTR-01-2022-0042

1. Introduction

Governance of artificial intelligence (AI) has been defined 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; fulfills legal requirements; and meets principles of ethical AI followed by the organization" (Mäntymäki *et al.*, 2022a). AI has become an object of governance because AI applications are increasingly widespread in many application areas across the private and public sectors (Lütge *et al.*, 2021; Reddy *et al.*, 2020). As an umbrella term, AI refers to a research field (Zhang and Lu, 2021), a set of information system capabilities of interpreting data, learning and adaptation (Kaplan and Haenlein, 2019), as well as a more general moving frontier of cutting-edge computing (Berente *et al.*, 2021). AI includes diverse techniques, with machine learning, robotics and pattern recognition being notable research areas (Zhang and Lu, 2021). The current information systems (IS) and management literature recognize AI as an essential driver of industrial development in the next generation of Industry 4.0, integrating technologies and assisting in various industries, including manufacturing, agriculture and education (Lu, 2019; Mazurek and Małagocka, 2019; Sigov *et al.*, 2022; Zhang and Lu, 2021).

Importantly, AI is also becoming prevalent in so-called high-risk application areas (European Commission, 2020), including healthcare (Reddy *et al.*, 2020), automobiles (Lütge *et al.*, 2021) and finance (Lee, 2020). As a result, there are increasing concerns over the potential risks and negative consequences of using AI (Jobin *et al.*, 2019). These concerns include accounts of privacy violations and discriminatory practices related to AI usage in healthcare (Lysaght *et al.*, 2019); divergence in the moral judgments deduced by AI algorithms (Aliman and Kester, 2019); accidents caused by autonomous automobiles (Stilgoe, 2018); and accountability for the norm violations of autonomous systems, such as military drones (Verdiesen *et al.*, 2021).

To address the concerns over the impacts of AI, there is growing public demand for the ethical use of AI. Accordingly, governmental and international organizations (such as the European Union [EU] and the Organization for Economic Co-operation and Development [OECD]), professional bodies (such as the Institute of Electrical and Electronics Engineers [IEEE]) and various companies have published their ethical AI principles and guidelines (Fjeld *et al.*, 2020; Hagendorff, 2020; Jobin *et al.*, 2019). However, these principle-based ethics provide limited assurance that the principles are met in practice (Hagendorff, 2020; Mittelstadt, 2019) because principles focus on the *what*, rather than the *how*, of AI ethics (Morley *et al.*, 2020). Ethical principles must be sufficiently concrete 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 must also be enforceable through governance (Cath, 2018; Minkkinen *et al.*, 2021; Morley *et al.*, 2020). Echoing these issues, there has been increasing emphasis on AI governance (AIG) in academia (Barn, 2020; Koniakou, 2023; Laato *et al.*, 2022a; Mäntymäki *et al.*, 2022a, b; Minkkinen *et al.*, 2021, 2022a, 2023; Papagiannidis *et al.*, 2023; Seppälä *et al.*, 2021; Zimmer *et al.*, 2022) and industry (Deloitte, 2021; KPMG, 2021; Statista, 2021a). On the industry side, a recent report suggests that a significant majority (87%) of IT decision-makers believe in the need to regulate AI-driven technologies, with a clear focus on ethics and corporate social responsibility (KPMG, 2021).

The emerging academic research on AIG focuses on three primary lines of inquiry: concepts and technical tools (Kroll, 2021; Larsson and Heintz, 2020), the translation of AI ethics principles into practice by connecting tools and principles (Mäntymäki *et al.*, 2022b; Mittelstadt, 2019; Morley *et al.*, 2020) and the analysis of societal AI policy and regulation (Koulu, 2020; Minkkinen and Mäntymäki, 2023; Smuha, 2021). However, prior research has included few attempts to synthesize the extant body of knowledge (for an overview focusing on public administration, see Sigfrids *et al.*, 2022). To address this void, we have undertaken a systematic literature review (SLR) of AIG, answering the following research questions (RQs):

RQ1. How does the literature conceptualize organizational-level AI governance?

RQ2. What are the key themes and knowledge gaps in AI governance research?

RQ3. What future agendas for AI governance can be identified based on the literature?

An SLR is a well-acknowledged tool that can help scholars consolidate current knowledge in a field, identify research gaps and outline the scope for future studies (Kitchenham *et al.*, 2009; Tranfield *et al.*, 2003). Thus, SLRs enable the development of an organized summary of knowledge in a research area through rigorous protocols, allowing scholars to build a foundational platform for understanding the state of the art and determining the scope for further knowledge development (Behera *et al.*, 2019; Kitchenham *et al.*, 2009). To the best of our knowledge, previous SLRs (Sharma *et al.*, 2020; Sigfrids *et al.*, 2022) have touched on AI governance but have not focused on the organizational governance of AI. Sigfrids *et al.* (2022) examined AI governance by public administration, thus excluding private companies. In turn, the SLR by Sharma *et al.* (2020) focused on the use of AI in promoting effective governance (i.e. *governance through AI*).

In the broad AI research landscape (e.g. Lu, 2019; Mazurek and Małagocka, 2019; Sigov *et al.*, 2022; Zhang and Lu, 2021), our study is positioned to address a specific problem: overcoming the principles-to-practices gap in AI ethics through organization-level AIG (Mäntymäki *et al.*, 2022a; Morley *et al.*, 2020; Schiff *et al.*, 2021b; Seppälä *et al.*, 2021; Whittlestone *et al.*, 2019). By organization-level AIG, we mean rules, practices, processes and tools that organizations can implement to govern their AI systems (Mäntymäki *et al.*, 2022a), as opposed to broader societal issues that state-level and supranational actors seek to address (Butcher and Beridze, 2019; Minkkinen and Mäntymäki, 2023). The research is thus placed within the human-centric (Shneiderman, 2020) and socio-technical (Dignum, 2020) AI research traditions, in contrast to the technical AI literature realm (e.g. Huang *et al.*, 2015). Our study contributes to the emerging AIG knowledge by drawing together disparate lines of inquiry into an overall understanding of AIG, identifying critical knowledge gaps in the current scholarly discourse, and presenting future research agendas to advance sufficiently broad and practicable AIG approaches.

The remainder of the paper proceeds as follows. The next section presents the methodological approach employed for the SLR, while the subsequent sections focus on the themes (Section 3) and knowledge gaps (Section 4) in the AIG literature. Section 5 synthesizes a model of four agendas (technical, stakeholder and contextual, regulatory, and process) from the identified themes and knowledge gaps. The model is presented to guide future AIG research and practice. The paper concludes with limitations and directions for future research.

2. Methodology

We developed the review protocol following well-established guidelines for systematic reviews (Kitchenham *et al.*, 2009; Webster and Watson, 2002) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements (Moher *et al.*, 2015; Page *et al.*, 2021). We ensured that our protocol followed both the original and the most recently published PRISMA guidelines to achieve rigorous reporting standards. We also consulted previously published SLRs (Behera *et al.*, 2019; Sharma *et al.*, 2020; Tandon *et al.*, 2020) to refine our protocol and provide information relevant to answering our RQs. We followed established guidelines to validate the transparency of the process and ensure the reproducibility of the dataset (Tranfield *et al.*, 2003).

The study was conducted in three stages: (a) determination of the SLR objectives and search strategy, (b) execution of planned protocols and (c) presentation of the results

and findings. Our ultimate objective was to assimilate and present a structured overview of the state-of-the-art research on AIG, particularly at the organizational level. Figure 1 provides a graphical overview of the SLR process and protocols, detailed in the following subsection.

2.1 Review plan

We commenced the SLR by identifying the relevant search terms and databases. To avoid prematurely fitting AIG into the existing frameworks and to allow for emergent findings resulting from an inductive process, we used a bottom-up approach to generate the search terms rather than starting from the preexisting governance models (e.g. IT governance). Thus, we initially conducted a search on Google Scholar using the terms "artificial intelligence" and "governance" and perused the first 300 results. This was deemed sufficient, as a cursory investigation of subsequent results pages showed little additional value, indicating that the first 300 results provided an adequate saturation of terms. We identified "AI," "black box," "neural networks," "machine learning (ML)" and "deep learning" as possible synonyms for executing the search. Next, we consulted two experts in the field (a senior academic and an industry AI expert) for their opinions on the identified search terms. These experts were deemed to adequately represent the mainstream views on the topic due to their extensive AIG project research and networks. Based on the feedback from these experts, we included "ML" and "ethics" as search terms. The inclusion of the term "ethics" was based on the experts' view that the current literature on AIG draws significantly from AI ethics research. This view was supported by the literature on the problem of translating AI ethics principles and guidelines into effective governance (the so-called principles-to-practices gap) (Morley *et al.*, 2020; Schiff *et al.*, 2021b). A preliminary examination of the AIG literature also confirmed that AI ethics and governance often accompany each other. Thus, the final search included the original and the subsequently identified terms.

We decided that a viable search strategy would include databases that support obtaining a socio-technical perspective on AIG research because AIG draws knowledge and materials from multiple fields (such as ethics, organizational studies and technology). Therefore, we excluded medical science databases and papers from the search due to the sector-specific nature of AI governance questions in the medical domain. Our search then included three electronic databases: Scopus, ISI Web of Science (WoS) and the Association of Computing Machinery (ACM) Digital Library. The WoS and Scopus databases have been used extensively in literature reviews because of their broad coverage of articles published in multiple disciplines, including social sciences, humanities and management. These two databases also comprehensively index the contents of other providers. The third database, ACM, focuses on computer science. Our review strategy aligns with prior SLRs on similar topics (Laato *et al.*, 2022b; Sharma *et al.*, 2020; Tandon *et al.*, 2020). The search was limited to articles published between 2010 and 2021.

Establishing article screening (inclusion and exclusion) and quality evaluation (QE) criteria were the final stages of the review plan. These criteria helped us screen the search results and identify the most relevant articles for addressing our RQs. These criteria were established based on prior SLRs on related topics (Ain *et al.*, 2019; Behera *et al.*, 2019; Tandon *et al.*, 2021). The two consulted experts also validated these criteria and suggested minor modifications to their wording; these were then incorporated. The final screening and QE criteria are detailed in Figure 1.

2.2 Search execution

The database searches, which were conducted in March 2021 and January 2022, resulted in the identification of 1,071 articles from the three databases (see Figure 1 for details). Two of the authors independently screened these results and resolved any conflicts through mutual

Source(s): Authors' own creation

Figure 1.
The review process

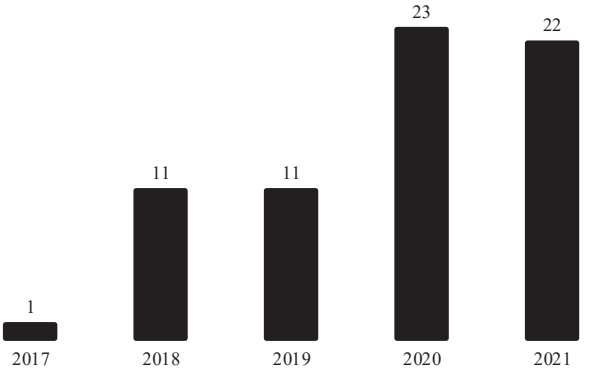
discussion at the end of the article selection process. In cases of unresolved disagreements over whether to include an article, a third author was consulted for the final decision. The screening process identified 82 articles that progressed to the QE process. However, before QE, both backward (references) and forward (citing publications) citation chaining was conducted for these 82 articles. During this process, 50 new articles were identified, of which 35 were removed because they did not meet the article selection criteria. Finally, 97 articles were carried forward for the QE process, which was conducted similarly to the original article screening. The two authors independently scored the articles based on whether they met the five QE criteria (for QE1 to QE4: 2 = Yes and 0 = No, for QE5: +2 if the sum of citations and H Index was >100, +1.5 if the sum of citations and H Index was ≥75 and ≤ 99, +1 sum of citations and H Index was ≥49 and ≤ 74 and 0 if the sum of citations and H Index is < 49). The articles that met the predetermined score threshold of 50% (i.e. 5 with a maximum possible score of 10) were retained in the final dataset. The inter-rater agreement for the article selection and QE process was good, as indicated by Fleiss' kappa values of 0.65 and 0.71 (Landis and Koch, 1977), respectively. After resolving inter-rater conflicts, 68 articles were included in the final dataset and further analyzed to answer our RQs.

2.3 Research profile

We first explored the data profile to generate insights into publication trends by examining year-on-year publication volume, authors, institutions and publication channels. As shown in Figure 2, which depicts the development of publication volumes, research on AIG has been gaining traction since 2017. This observation aligns with de Almeida *et al.*'s (2021) assessment of recent developments in the field.

Table 1 shows the publication outlets for the journal articles (52) and conference papers (16). The AAAI/ACM Conference on AI, Ethics and Society has the highest number of papers (6), followed by Philosophical Transactions of the Royal Society A (4).

Concerning methods, the majority of the publications in the dataset were conceptual (n = 61), with only a few literature reviews (n = 4) or empirical studies (n = 3). The high proportion of conceptual studies may be partly explained by the conceptual nature of the AI ethics scholarship linked to AI governance. The predominance of conceptual papers in AIG research means that the SLR will focus on conceptual consolidation, identifying and



Note(s): The year 2021 includes 4 papers where the online version was published in 2021 and the print version in 2022
Source(s): Authors' own creation

Figure 2.
Publication volume per year

Journals (52 papers)	Conference proceedings (16 papers)
Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences (4)	Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (6)
AI & Society (3)	Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (3)
AI and Ethics (2)	Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (2)
Ethics and Information Technology (2)	Proceedings of the Annual International Conference on Digital Government Research
Information (Switzerland) (2)	Proceedings of the IEEE International Conference on Intelligent Engineering Systems
Journal of the American Medical Informatics Association (2)	Proceedings of the IEEE International Symposium on Technologies for Homeland Security
Minds and Machines (2)	Proceedings of the International Conference on Theory and Practice of Electronic Governance
Philosophy and Technology (2)	Proceedings of the International Joint Conference on Artificial Intelligence
Science and Engineering Ethics (2)	
Telecommunications Policy (2)	
ACM Transactions on Interactive Intelligent Systems	
Asian Bioethics Review	
Asian Journal of Law and Society	
Big Data and Cognitive Computing	
Business Information Review	
Computer Law and Security Review	
Engineering	
Ethics and Human Research	
European Business Organization Law Review	
European Journal of Legal Studies	
Futures	
IEEE Internet Computing	
IEEE Symposium Series on Computational Intelligence	
Information, Communication and Society	
Interdisciplinary Science Reviews	
International Cybersecurity Law Review	
International Journal of Technoethics	
Internet Policy Review	
Journal of Business Research	
Journal of ICT Standardization	
Journal of Information, Communication and Ethics in Society	
Journal of International Humanitarian Legal Studies	
Nature Machine Intelligence	
Policy and Society	
Regulation and Governance	
Social Studies of Science	
Sustainable Development	
Technology in Society	
The RUSI Journal	

Source(s): Authors' own creation

Table 1.
Publication channels

strengthening the conceptual foundations of AIG rather than collating empirical findings. With regard to authorship, Figure 3 shows authors who have published two or more of the articles included in the review. The geographical distribution (based on the institution of the lead author) indicates that the United Kingdom ($n = 21$) and the United States ($n = 13$) are the leading countries in AIG research (Figure 4). The distribution of lead authorships by country aligns with notions pointing to the leading role of Western perspectives and developed countries in AIG research, policies and practices (Schiff *et al.*, 2020; Smuha, 2021). Finally, most (57 out of 68) of the primary affiliations for lead authors were universities

(Figure 5), while the profile also indicates the presence of authors from governmental organizations and companies.

2.4 Data analysis

To analyze the dataset and identify the focal areas of AIG research, two of the authors conducted open and axial coding in iterative rounds (Corbin and Strauss, 2014). The codes and the themes are summarized in Table 2. In the initial rounds, the two authors independently studied the content of the articles, made notes on their focal ideas and then independently assigned labels to their notes to develop the initial open codes. Next, the authors discussed their initial open codes to reach a consensus on the final open codes—those

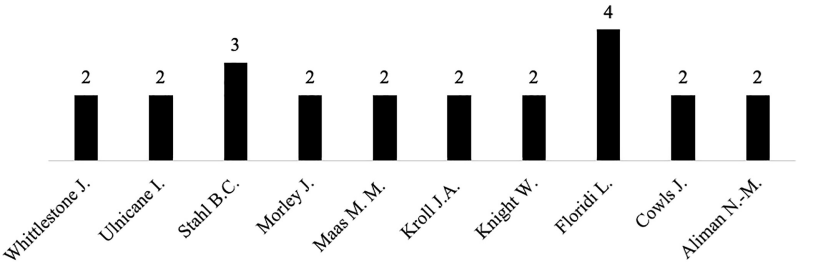


Figure 3.
Authors with two or more articles included in the review

Source(s): Authors' own creation

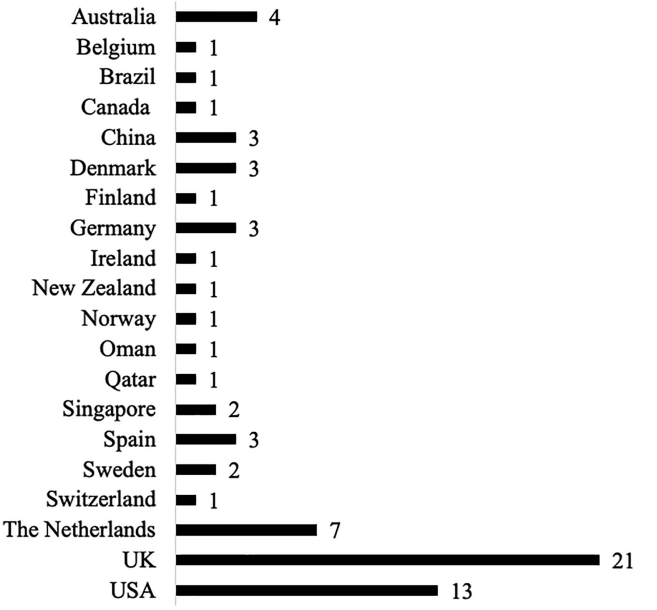
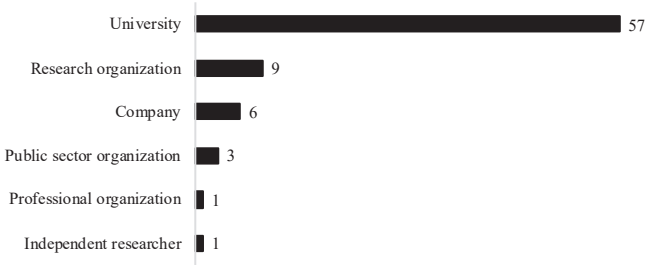


Figure 4.
Lead authorship by country

Note(s): The number of papers exceeds 68 because three authors have reported two affiliations in different countries

Source(s): Authors' own creation



Note(s): * The number of papers exceeds 68 because some authors reported multiple affiliations. Government organization = organizations within national states (e.g. City of New York), Research organization = organizations focused purely on research (e.g. Alibaba-NTU Singapore Joint Research Institute, Alan Turing Institute), Professional organization = Royal Statistical Institute (professional network bodies)
Source(s): Authors' own creation

Figure 5.
Type of lead author affiliation

that best represented the focal content of the studied articles. Examples of these final open codes included "international cooperation," "ethical principles" and "policy to practice."

In the next rounds, the authors reflected on the similarities and differences in the open codes to develop more comprehensive categories, resulting in an initial set of axial codes. After the authors independently developed the axial codes, they again utilized deliberation to agree on the ones most suited to the article content. The inter-rater reliability was assessed at this stage, with a kappa value of 0.86 indicating solid inter-rater reliability of the axial codes. As examples, the axial codes included "translating policy to practice" and "sociopolitical context." The final proposed themes brought together the findings under four headings (technology, stakeholders and context, regulation and processes) that we will discuss after outlining the conceptualizations of AI governance.

3. Conceptualizations of AI governance and themes in the AI governance literature

In the next sections, we present the conceptualizations of AIG and the themes in the AIG literature to address our RQ1 and RQ2, respectively. Based on these, we turn to the knowledge gaps (the latter part of RQ2) in Section 4.

3.1 Conceptualizations of AI governance

Few of the reviewed papers explicitly define AIG (Morley *et al.*, 2020; Robles Carrillo, 2020; Ulnicane *et al.*, 2021a). Defining AIG is challenging because there is a lack of academic consensus on the definition of AI (Cihon *et al.*, 2020; Robles Carrillo, 2020) and its constituent elements (Larsson and Heintz, 2020; Wu *et al.*, 2020). Rather than the organizational perspective adopted by our study, the definitions found in the dataset (see Table 3) tended to focus on public policy, applied ethics and the value provided by AI (Aliman and Kester, 2019; Butcher and Beridze, 2019; Perry and Uuk, 2019; Al Zadjali, 2020). Butcher and Beridze (2019) discussed global governance and underscored the role of mechanisms, stating that AIG incorporates a variety of solutions and tools. As examples, they mentioned ethical and value-based frameworks that can influence AI development and application, research on AI implications, the formulation of technical solutions and the implementation of legislation.

Proposed theme	Axial codes	Open codes (examples)	Articles
Technology	Data, algorithm, AI	Data	Bu (2021), Raab (2020)
		Algorithm	Barn (2020), Lee (2020), Shah (2018), Yeung (2018)
		AI/system	Buenfil *et al.* (2019), Buhmann and Fieseler (2021), Domanski (2019), Gasser and Almeida (2017), Kroll (2018), Larsson (2020), Larsson and Heintz (2020), Lysaght *et al.* (2019), Reddy *et al.* (2020), Shneiderman (2020), Stilgoe (2018)
		Problems/challenges (e.g. discrimination, human control)	Carter (2020), Liu and Maas (2021), Maas (2018), 2019, Tsamados *et al.* (2022)
Stakeholders and context	Stakeholder	People, individual, media, responsibility, government	Buhmann and Fieseler (2021), Cihon *et al.* (2021), Erdélyi and Goldsmith (2018), Gasser and Almeida (2017), Ibáñez and Olmeda (2022), Lewis *et al.* (2020), Metcalf *et al.* (2021), Orr and Davis (2020), Shneiderman (2020), Tsamados *et al.* (2022), Whittlestone *et al.* (2019), Wieringa (2020)
	Sociopolitical context	Culture, international cooperation, politics, feedback loop/model, alignment	Aliman *et al.* (2019), Aliman and Kester (2019), Cihon *et al.* (2020), Feijóo *et al.* (2020), Hickok (2021), Krijger (2022), ÓhÉigeartaigh *et al.* (2020), Rahwan (2018), Reddy *et al.* (2020), Schiff *et al.* (2020), Stahl *et al.* (2021), Ulnicane *et al.* (2021a), Ulnicane *et al.* (2021b), Al Zadjali (2020), Zhang and Dafoe (2020)
Regulation	Hard regulation	Regulation, law	Erdélyi and Goldsmith (2018), Koulu (2020), Robles Carrillo (2020)
	Soft regulation	Ethical principles, values, guidelines, certifications/ standards/supportive frameworks	de Almeida (2021), Butcher and Beridze (2019), Cath (2018), Gasser and Almeida (2017), Larsson (2020), Lewis *et al.* (2020), Lütge *et al.* (2021), Lysaght *et al.* (2019), Mittelstadt (2019), ÓhÉigeartaigh *et al.* (2020), Robles Carrillo (2020), Roski *et al.* (2021), Shneiderman (2020), Whittlestone *et al.* (2019), Winfield and Jirotka (2018), Yu *et al.* (2018)

Table 2.
Themes and codes in the dataset

(*continued*)

Proposed theme	Axial codes	Open codes (examples)	Articles
Processes	Standards, processes, translation of policy into practice	Oversight/accountability	Lütge *et al.* (2021), Rahwan (2018), Verdiesen *et al.* (2021), Wieringa (2020)
		Auditing	Mökander and Floridi (2021)
		Impact assessment/outcomes	Morley *et al.* (2020), Raab (2020), Smuha (2021), Stix (2021), Truby (2020)
		Translation/implementation/ policy to practice	Floridi *et al.* (2018), Friesen *et al.* (2021), Gasser and Almeida (2017), Kroll (2021), Perry and Uuk (2019), Shneiderman (2020), Zhou *et al.* (2020)
		Processes/making ethical decisions	Aliman and Kester (2019), de Almeida (2021), Boesl and Bode (2019), Häußermann and Lütge (2022), Koulu (2020), Wu *et al.* (2020)

Source(s): Authors' own creation

Table 2.

Aliman and Kester (2019) adopted a more technical standpoint. However, they also emphasized social and ethical values by stating that AIG is focused on addressing the ethical requirements of society, with particular emphasis on aligning human values with AI implementation. Al Zadjali (2020) emphasized the role of governments in relation to AIG, stating that this involves the capacity of government agencies to generate value for all stakeholders through various functions, such as value alignment and performance management. According to Perry and Uuk (2019), AIG is about how humans can best advance AI development, which comprises three main elements: technical landscape (limits and scope of AI), ideal governance (potential pathways for facilitating stakeholder cooperation) and AI politics (political dynamics affecting stakeholders).

3.2 Themes in the AI governance literature

Several recurring themes surfaced from the reviewed literature. The following sections outline four AIG themes identified from our literature review: technology, stakeholders and context, regulation and processes (Figure 6).

3.2.1 Technology. The technology theme refers to data and algorithms that are the foundations of AI technologies, and it also discusses governance challenges related to the technical characteristics of AI systems. The reviewed literature includes multiple studies that develop frameworks for AIG centered on data and algorithms, drawing on areas such as AI safety (Maas, 2018; Macrae, 2019) and responsible innovation (Buhmann and Fieseler, 2021; Ulnicane *et al.*, 2021a).

The technology perspective includes three interlinked levels: data, algorithms and AI systems. AI systems learn and adapt based on *data*. Therefore, it is hardly surprising that data governance is considered a key pillar of AIG (Barn, 2020; Gasser and Almeida, 2017; Koulu, 2020). The studies focusing on data governance discussed issues such as data privacy (Gasser and Almeida, 2017; Raab, 2020), legal protection (Koulu, 2020) and integrity (Carter, 2020), which are integral to practical AIG initiatives (Gasser and Almeida, 2017; Lysaght *et al.*, 2019). Moreover, the role of training data is explicitly discussed in the literature to ensure the appropriate conduct of AI systems. For example, using multiple training datasets can

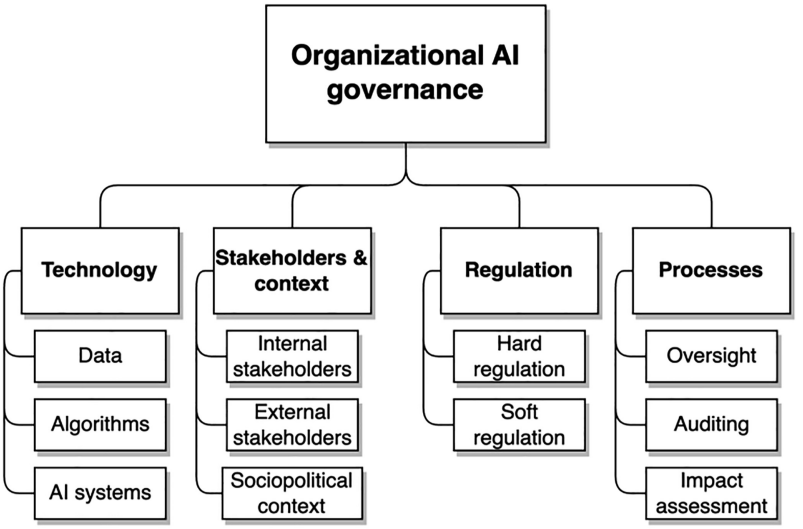
Definition	Source
"In sum, Artificial Intelligence governance is creating optimal value from AI by maintaining a balance between realizing benefits and optimizing risk levels, and resource use. [. . .] the main objective of enforcing AI governance is to ensure government agencies are able to perform risk management, value delivery, strategic alignment, resource management, and performance management to create value for all stakeholders."	Al Zadjali (2020)
"For the governance of artificial intelligent systems which is a field of interest within both AI Safety and AI Ethics at an international level, it becomes crucial to design an appropriate goal specification framework able to encode the ethical and legal requirements within a given societal context." "From the perspective of AI Safety, the governance of AI systems requires solving the value alignment subtask which aims at implementing AI systems such that they are aligned with human ethical values."	Aliman and Kester (2019)
"AI governance can be characterised as a variety of tools, solutions, and levers that influence AI development and applications. Some examples include: promoting norms, ethics and values frameworks (which may take the form of self-regulation from leading tech companies choosing to work on specific projects or not); researching the effects, implications and possible solutions to AI use raising awareness for stakeholders; building technical solutions that deal with certain issues raised by AI technology (such as algorithmic interpretability and explainability, which is the ability to precisely understand how an algorithm has made its decision); and implementing legislative measures and establishing formal regulatory bodies that have jurisdiction to govern AIs-related technologies and fields."	Butcher and Beridze (2019)
"A growing body of the literature covers questions of AI and ethical frameworks, laws to govern the impact of AI and robotics, technical approaches like algorithmic impact assessments, and building trustworthiness through system validation. These three guiding forces in AI governance (law, ethics and technology) can be complementary."	Cath (2018)
"AI governance [. . .] studies how humanity can best navigate the transition to advanced AI systems. This would include the political, military, economic, governance, and ethical considerations and aspects of the problem that advanced AI has on society. AI governance can be further broken down into other components, namely the technical landscape [. . .], ideal governance [. . .], and AI politics."	Perry and Uuk (2019)

Table 3.
Definitions of AI governance

Source(s): Authors' own creation

potentially improve accountability (Shah, 2018), and data available on social media, reports on humanitarian actions and ethical misconduct by public bodies can be used to facilitate the self-learning of AI systems for ethical behavior (Buenfil *et al.*, 2019).

The transparency, explainability and inscrutability of algorithms are significant challenges in AIG (Kroll, 2018; Larsson and Heintz, 2020). Hence, the algorithms and data used to develop and train AI models are discussed in the reviewed studies, with an emphasis on the accountability, transparency (Domanski, 2019; Lysaght *et al.*, 2019), and explainability of AI models and algorithms (Kroll, 2018, 2021). Furthermore, the transparency of algorithms and AI systems has been accorded critical status in existing principles, regulatory guidelines and documents. The "black box" (Cath, 2018; Kroll, 2018; Larsson, 2020) and opacity (Buhmann and Fieseler, 2021; Shneiderman, 2020) problems are perceived by researchers as AIG challenges that require technical solutions. The opacity of AI refers to the inability of humans to understand how AI systems reach their respective outputs. This is a critical challenge, as self-learning AI algorithms generate rules that are not defined by their developers (Buhmann and Fieseler, 2021). However, there is no consensus regarding the scope and understanding of transparency in existing AI ethics guidelines (Larsson, 2020). For example, there is a conceptual difference between the transparency of algorithms and AI



Source(s): Authors' own creation

Figure 6.
Themes in the AI governance literature

(Larsson and Heintz, 2020), meaning the transparency of a technological component (algorithmic transparency) or the overall system (AI transparency).

In addition to transparency, AIG deals with fairness issues and potential harm. AI algorithms can potentially create numerous adverse societal consequences (Barn, 2020). Therefore, researchers and the public have raised concerns about the possible misuse, unfairness and discrimination linked to AI systems (Carter, 2020; Kroll, 2018, 2021). The related issue of bias is another key concern in the AIG literature. For instance (Domanski, 2019), posited that bias in AI could arise due to its design, code or data and that each source of bias requires a distinct mitigation or remediation approach.

Turning to *AI systems*, the reviewed literature emphasizes that AIG initiatives and policies should provide clear explanations (including explainability in context) in terms of the designs, operations and evaluations of AI systems (Buhmann and Fieseler, 2021; Kroll, 2018; Morley *et al.*, 2020; Reddy *et al.*, 2020). Such efforts vis-à-vis AIG could verify trustworthiness in AI systems (Kroll, 2018) and balance the interests of stakeholders (such as consumers and citizens) who are affected by an AI system (Larsson and Heintz, 2020). While it is a common perception that AI systems are inscrutable black boxes, Kroll (2018) suggested that these systems are fundamentally understandable and that the current degree of AI inscrutability could result from the existing power dynamics in the field.

3.2.2 Stakeholders and context. The second theme, stakeholders, refers to the actors and actor roles involved in developing, deploying and governing AI systems. In this section, we first list the key stakeholder groups and then consider the roles and responsibilities of different stakeholders. Finally, we discuss cooperation among stakeholders. The list of internal and external stakeholders identified in the AIG literature is presented in Table 4. From the perspective of a particular AI system used in an organization, 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). Overall, Table 4 indicates that the list of stakeholders is heterogeneous and includes numerous potentially overlapping roles.

Stakeholder group	Examples	Sources
Internal stakeholders		
Organizational managers	Chief executive officers, Chief information officers	Shneiderman (2020)
Decision-makers setting AI system specifications	Organizational roles that decide how algorithms are implemented	Orr and Davis (2020), Wieringa (2020)
AI system developers	Software development teams	Ibáñez and Olmeda (2022), Lewis *et al.* (2020), Orr and Davis (2020), Shneiderman (2020), Wieringa (2020)
AI system users	Organizations utilizing AI systems, individual users	Lewis *et al.* (2020), Orr and Davis (2020), Wieringa (2020)
External stakeholders		
Technology companies	Google, Amazon	Schiff *et al.* (2020), Whittlestone *et al.* (2019)
Professional bodies	IEEE, ACM	Cihon *et al.* (2021), Schiff *et al.* (2020), Shneiderman (2020), Whittlestone *et al.* (2019)
Standard-setting bodies	ISO, IEEE	Whittlestone *et al.* (2019)
Decision-makers setting AI system specifications	Organizational roles that decide how algorithms are implemented	Orr and Davis (2020), Wieringa (2020)
AI system developers	Software engineering teams	Ibáñez and Olmeda (2022), Lewis *et al.* (2020), Orr and Davis (2020), Shneiderman (2020), Wieringa (2020)
AI system users	Organizations utilizing AI systems, individual users	Lewis *et al.* (2020), Orr and Davis (2020), Wieringa (2020)
Data providers	Individuals, data brokers	Lewis *et al.* (2020)
Government bodies and regulators, oversight authorities	UK's Centre for Data Ethics and Innovation, European Data Protection Board	Lewis *et al.* (2020), Orr and Davis (2020), Schiff *et al.* (2020), Whittlestone *et al.* (2019)
Intergovernmental bodies	The United Nations Educational, Scientific and Cultural Organization [UNESCO], Council of Europe	Schiff *et al.* (2020)
Researchers	Oxford Internet Institute, Microsoft Research	Whittlestone *et al.* (2019)
Organizational managers	Chief executive officers, Chief information officers	Shneiderman (2020)
Non-governmental organizations and civil society organizations	Electronic Privacy Information Center	Schiff *et al.* (2020), Shneiderman (2020)
Affected stakeholders	Individuals and groups	Lewis *et al.* (2020)
Insurance companies	Allianz, AXA	Shneiderman (2020)
Auditing firms	PwC, KPMG	Shneiderman (2020)
Media	Platforms for public talks, social media	Shneiderman (2020)

Source(s): Authors' own creation

Table 4.
Internal and external stakeholder groups covered in the AI governance literature (examples added by the authors)

Relatively few of the reviewed articles focused explicitly on the roles and responsibilities of stakeholders (Buhmann and Fieseler, 2021; Erdélyi and Goldsmith, 2018), as well as their contributions to the outcomes of AIG, such as economic benefits (Lewis *et al.*, 2020) and the good of society (Tsamados *et al.*, 2022; Ulnicane *et al.*, 2021a). The developers of AI products and services are perceived to balance the complex trade-offs and normative conventions that govern ethical responsibilities in AI deployment (Orr and Davis, 2020). AI development

practitioners play a mediating role by implementing ethics within AI solutions, while the parameters are set by legislation, organizational norms and clients (Orr and Davis, 2020). Moreover, the skills of such practitioners (e.g. information risk management) can contribute significantly to the development of AI- and ML-based technologies (Carter, 2020).

When discussing the dynamics of AIG stakeholders, the reviewed studies suggest that technical expertise and power are mainly found among a small number of actors, particularly in large technology companies with extensive datasets (Lewis *et al.*, 2020). A power imbalance also exists between the actors who set specifications for AI systems (i.e. regulations, organizational norms and clients) and the practitioners who implement them (Orr and Davis, 2020). This can translate into an imbalanced allocation of AIG responsibilities among stakeholders, wherein existing principles may be applied to benefit private organizations, especially those based in more developed economies (Lewis *et al.*, 2020; Ulnicane *et al.*, 2021b).

In addition to the unequal allocation of technical expertise and power, the reviewed literature points to a hierarchy among the different stakeholders of AIG in terms of their degrees of influence (Erdélyi and Goldsmith, 2018; Lewis *et al.*, 2020; Ulnicane *et al.*, 2021b), specific responsibilities (Aliman *et al.*, 2019; Orr and Davis, 2020) and accountability levels (Wieringa, 2020). External stakeholders (such as professional bodies and governments) tend to adopt a more regulatory stance by suggesting, framing and refining the existing guidelines for AIG. By comparison, internal (intraorganizational) stakeholders play more specific roles in translating policies into practice through organization-level mechanisms (Shneiderman, 2020).

Stakeholders operate within a cultural and sociopolitical context (Cihon *et al.*, 2021), which provides an overarching landscape for AIG activities. For example, some commentators note that current AI systems tend to reflect American cultural ideology, particularly the cultural assumptions of large US technology companies (e.g. Cath, 2018). According to some researchers, universal values and rights must be contextualized to particular cultures, and we can see that AI development in the United States, China and Europe has followed different paths, balancing market concerns, individual rights and government (Feijóo *et al.*, 2020; Lewis *et al.*, 2020). For example, China's strong position in global AI development can be attributed to many factors, including its stance toward individual privacy protection and relatively quick adoption of AI-driven technological innovations, compared to Europe and the United States (Feijóo *et al.*, 2020).

In addition to the national cultural context, AIG execution can be affected by specific norms related to an organization, field of research or sector (Maas, 2018; Shneiderman, 2020). The literature identifies safety culture as a vital area of organizational and sectoral norms (Shneiderman, 2020). Additionally, toxic organizational cultures, problematic business cultures (Hickok, 2021), and the cultural divide between scholars from the engineering and humanities fields (Rahwan, 2018) can obstruct AIG efforts.

The literature indicates that multipartite cooperation and dialogue between stakeholders (such as private companies, government bodies, universities and research organizations) are required to harness the potential of AI while meeting the required outcomes of explainability, accountability and transparency (Gasser and Almeida, 2017). This cooperation can take different forms. For example, stakeholder collaboration can occur sectorally, such as in healthcare, wherein academia and healthcare service organizations can build AI-related skills, policies and practices (Reddy *et al.*, 2020).

The extant research has suggested establishing a centralized intergovernmental organization (the International Artificial Intelligence Organization) to institutionalize stakeholder collaboration and regulate AIG internationally (Cihon *et al.*, 2020; Erdélyi and Goldsmith, 2018). The reviewed literature included an examination of the existing collaborative efforts, such as the Partnership on AI (Schiff *et al.*, 2020). Such collaborations among stakeholders can build public trust in AI and negate public perceptions of potential

harm (Carter, 2020). Studies have also promoted global science diplomacy and research cooperation as collaborative approaches (Ulnicane *et al.*, 2021a). However, the political tensions and divergent philosophical traditions between regions (such as the United States, China and Europe) need to be overcome to allow cross-cultural AI cooperation (ÓhÉigeartaigh *et al.*, 2020).

3.2.3 Regulation: hard and soft regulation. The third theme refers to the hard and soft regulation of organizational AIG activities. Interestingly, legislation was discussed in the reviewed literature, even though hard law was excluded from the search scope. While legislative issues warrant a separate literature review, we briefly discuss regulations because they were mentioned in the reviewed literature. The existing regulations for AIG encompass hard law (binding legislation) and soft governance approaches, including standards, certificates, audits and explainable AI systems (Floridi *et al.*, 2018; Kroll, 2018; Lewis *et al.*, 2020; Shneiderman, 2020). The reviewed studies emphasized the criticality of algorithmic (Erdélyi and Goldsmith, 2018) and AI regulation as a part of legislative governance (Butcher and Beridze, 2019). Moreover, hard laws, such as anti-discrimination laws (Shah, 2018) and the EU General Data Protection Regulation (Larsson, 2020), are essential regulatory foundations of AIG.

This review focuses on soft governance mechanisms (such as standards), which have lesser degrees of institutional formality and greater flexibility than hard law (Erdélyi and Goldsmith, 2018; Robles Carrillo, 2020). A recent study highlights the need for industry self-governance mechanisms because organizational self-regulation alone insufficiently mitigates and manages the risks associated with AI deployment (Roski *et al.*, 2021). However, the standardization of practical AIG implementation is in its formative stages Kroll (2021), Larsson and Heintz (2020). The expected benefits of standards include industry-based AIG practices and the reduction of interoperability barriers to commerce (Lewis *et al.*, 2020). The International Organization for Standardization (ISO) and the IEEE are the two most important standardization bodies for AIG. Accordingly, the ISO has launched a working group for AI trustworthiness to address AI transparency and traceability (Kroll, 2021; Larsson and Heintz, 2020). Similarly, the IEEE has instigated the P7000 series of standards, of which P7001 relates to standardizing the transparency of autonomous systems and determining the reasons for its decision-making (Larsson and Heintz, 2020). Such standards can be leveraged by different stakeholders (including certification agencies) to develop measurable transparency levels.

Soft governance also includes AI ethics (de Almeida *et al.*, 2021; Cath, 2018), with commonly discussed principles, such as technical robustness, data security, privacy, reliability, transparency, fairness and accountability (de Almeida *et al.*, 2021; Gasser and Almeida, 2017; Lütge *et al.*, 2021; Winfield and Jirotka, 2018; Wu *et al.*, 2020). Researchers seem to have a consensus that the existing ethical principles should aim to articulate the general human values that AI deployments address and that the principles act as practical guidelines for AI developers (ÓhÉigeartaigh *et al.*, 2020; Whittlestone *et al.*, 2019). In other words, ethical principles help AI developers align human and societal values through the design and governance of associated technologies (Winfield and Jirotka, 2018). In addition to general principles, sector-specific governance frameworks have been suggested with regard to the use of big data in healthcare, drawing on sector-specific ethical guidelines (Lysaght *et al.*, 2019).

Addressing human values and ethical issues in the development and use of AI has drawn intensive attention from public and private organizations, expert groups (such as the EU High-Level Expert Group on AI) and professional bodies (such as the IEEE and ACM). The reviewed studies indicated the existence of over 80 such AI ethics guidelines and policy documents (Mittelstadt, 2019), and this number continues to increase. The extensive discussions on ethical principles seem to converge around four principles of medical ethics

or bioethics: beneficence, non-maleficence, autonomy and justice (Mittelstadt, 2019; Whittlestone *et al.*, 2019). According to some authors, explicability is emerging as an additional ethical AI principle (Floridi *et al.*, 2018).

The literature on ethical guidelines focused on technically addressing the inscrutability of evidence, unfair outcomes (Barn, 2020) and algorithmic opacity (Tsamados *et al.*, 2022). This emphasis on enforcing ethics in AI technologies can be attributed to the assumption that most AIG problems are due to the technology itself (Koulu, 2020). However, the recent literature has also started to adopt a more business-oriented (Häußermann and Lütge, 2022) and socio-legal approach (Larsson, 2020) to incorporating ethics into AI.

Overall, ethical principles and guidelines drive AIG to address human rights concerns, such as bias, discrimination and privacy (de Almeida *et al.*, 2021; Bu, 2021). Ethical values can be characterized as the foundation of AIG and the desired output because they represent the human values that should be embedded in governed AI technologies. However, the landscape of ethical principles is complicated because ethical guidelines are drafted in different geographic regions; by actors from various sectors; and with distinct motivations, such as competitive advantage and signaling leadership (Schiff *et al.*, 2020).

3.2.4 Processes: oversight, auditing and impact assessment. The fourth theme comprises the processes, procedures and practices through which AI development and use are governed. Process-oriented approaches have been suggested in the literature to facilitate AIG at the technical (Wu *et al.*, 2020), team, organizational and industry levels (Shneiderman, 2020). With regard to the technical aspects, federated learning and blockchains can be used to ensure data security and privacy (Wu *et al.*, 2020), while data validation is presented as a viable mechanism for addressing the inscrutability of the evidence affecting AI systems (Tsamados *et al.*, 2022).

The literature on core organizational AIG processes is still scant, but the studies document three primary mechanisms that support organizational AIG: oversight, auditing, and risk and impact assessments. Especially in Europe, policymakers suggest involving human control and oversight in AI development processes through approaches like human-on-the-loop, human-in-command and human-in-the-loop (HITL) (de Almeida *et al.*, 2021; Koulu, 2020). Here, HITL is about utilizing individual or group judgment when optimizing AI systems, such as using human workers to label data for training ML algorithms (Rahwan, 2018).

Scholars have proposed that alignment with societal expectations is necessary for AI ethics (Friesen *et al.*, 2021; Ulnicane *et al.*, 2021b). As a result, the HITL approach has been expanded to the community-in-the-loop (Häußermann and Lütge, 2022) and society-in-the-loop (SITL) frameworks (de Almeida *et al.*, 2021; Rahwan, 2018). These frameworks build on social contract theory and focus on incorporating societal values into AI systems that affect broader societal outcomes, such as societal well-being (Lütge *et al.*, 2021). Moreover, CITL emphasizes stakeholder deliberation to resolve conflicts and agree on the rules (and trade-offs) required when deploying AI systems (Häußermann and Lütge, 2022). By comparison, SITL seems to focus more on tools for quantifiably measuring human values (Häußermann and Lütge, 2022; Rahwan, 2018) and provides a mechanism for monitoring compliance (Rahwan, 2018).

Complementing organizational oversight, the literature also acknowledged the need for a broader level of governance, such as a global oversight body that can curate principles and norms for AI systems (Gasser and Almeida, 2017; Perry and Uuk, 2019). For example, Floridi *et al.* (2018) suggested the establishment of an AI oversight body at the European level (analogous to the European Medicine Agency). This oversight actor would be responsible for supervising AI systems and products. It could also facilitate AIG execution in high-risk areas, such as lethal autonomous weapons systems, transport and general AI safety (Butcher and Beridze, 2019; Schiff *et al.*, 2020).

Moving from oversight to auditing, researchers have suggested that ethics-based auditing could promote AIG alongside other processes (Mökander and Floridi, 2021). Auditing enables practitioners to limit adverse outcomes through continual examinations and risk assessment processes (Truby, 2020). Internal assessments are part of a broader oversight system that strengthens the capacity for self-regulated quality checks (Shneiderman, 2020). Moreover, some companies have recognized the importance of ethics-based auditing, with external auditing organizations, accounting and professional services firms contributing to developing ethics-based auditing frameworks (Mökander and Floridi, 2021).

Risk and impact assessments, in turn, identify who is impacted by AI systems and how (Wieringa, 2020). One example of an impact assessment is the Algorithmic Impact Assessment (AIA) (Metcalf *et al.*, 2021), which is employed by the Canadian government (Stix, 2021). AIAs exemplify the existing approach of evaluating the ethical impacts and risks of a technical or research project before permission is granted for its execution (Raab, 2020). Other assessments with established self-regulatory mechanisms include the Privacy Impact Assessment (Morley *et al.*, 2020; Raab, 2020; Wieringa, 2020), Data Protection Impact Assessment (DPIA) (Raab, 2020; Wieringa, 2020) and Social Impact Assessment (Raab, 2020). The EU High-Level Expert Group on AI has also proposed the "Trustworthy AI Assessment," which builds on the DPIA and emphasizes the assessment of AI impacts on human rights, including for children and people with disabilities (Smuha, 2021).

4. Knowledge gaps

Addressing RQ2 on themes and knowledge gaps, our review highlights four gaps in the current knowledge of AIG: the limited understanding of the implementation of AIG in organizations, lack of attention to the AIG context, uncertain effectiveness of ethical principles and regulation, and insufficient operationalization of AIG processes. These gaps, including their related themes and future research agendas, are summarized in Table 5 and discussed in the following sections. The gaps were derived by contrasting the first research question on the conceptualization of organizational AIG with the themes in the literature. As explained in the methodology section, conceptual papers far outnumbered empirical research. As a result, there were numerous conceptual discussions on principles and issues but little empirical elaboration. This gives the gaps a similar form: We know what the issues are but do not know enough about them, especially empirically. Where relevant, broader literature within the human-centric AI research stream but outside the SLR sample was used to elaborate on the gaps.

As a general observation, we note the challenges in understanding AIG maturity and evaluating the current and future effectiveness of AIG activities, amplified by technological development around AI. Moreover, a significant part of the reviewed AIG literature focuses on finding and implementing technical solutions to improve existing AI systems (Aliman *et al.*, 2019; Domanski, 2019; Stilgoe, 2018). These technical solutions are designed to, for example, protect data privacy (Carter, 2020) and security (Shneiderman, 2020), improve AI system learning by obtaining data from public sources (Buenfil *et al.*, 2019) and leverage diverse datasets to mitigate possible biases (Shneiderman, 2020).

4.1 Limited understanding of the implementation of AI governance in organizations

The first gap in the AIG literature relates to the organizational implementation of AIG. Empirical research providing guidance on AIG efforts has remained scant. Pioneering conceptual research has advocated the importance of examining the systemic aspects of AIG (e.g. Kaminski, 2019), and empirical studies (e.g. Mayer *et al.*, 2021; Seppälä *et al.*, 2021) have acknowledged the importance of integrating AIG aspects into technical AI development and operations activities.

Table 5. Summary of themes, gaps and agendas for future research

Theme	Gaps		Future agendas
Technology	Limited understanding of AIG implementation	Lack of attention to systemic aspects of AIG	Integration of AIG into software development and IT operations (DevOps & MLOps)
		Limited understanding of AI system design implications of transparency and explainability	Governance by design, automatization of AIG activities
Stakeholders and context	Lack of attention to AIG context	Limited consideration of the sociocultural and political environment	Consideration of AIG mechanisms and processes according to stakeholders' political power and cooperation, organizational work culture, and norms
		Limited attention to the formal acknowledgment of stakeholder roles and responsibilities	Examination of the stakeholder ecology or network; collaborative governance; inter-stakeholder trust; stakeholders' degree of involvement, trust and active role in AIG development and execution
Regulation: hard and soft regulation	Uncertain effectiveness of ethical principles and regulations	Perceived inconsistency in the effectiveness of some regulatory measures	Emphasis on the translation of policies and regulatory measures into actionable AIG mechanisms Establishment of responsibility and stakeholder compliance mechanisms to resolve conflicts in the perceived effectiveness of AIG measures
		Concerns regarding the suitability of current principles and their success at organizational-level AIG deployment	Broadening of the scope of ethical principles by considering related fields, such as Internet and digital governance, corporate social responsibility and societal values (such as democracy) AI ethics and competence training in organizations to increase employee awareness of relevant ethical and AIG issues
Processes: oversight, auditing, and impact assessment	Insufficient operationalization of AIG processes	Lack of consideration of organizational processes and factors regarding the ethical alignment and efficacy of AIG	Empirical research identifying the organizational risks and factors influencing the effective execution of AI Development of concrete mechanisms for addressing identified risks (through governance initiatives) and the investigation of their efficacy
		Unclear sufficiency of the existing process-oriented approaches to AIG	Empirical research directed at how these processes are deployed at the organizational level and the inherent complexities while targeting individual actors (i.e. the people responsible for enacting AIG initiatives)

Source(s): Authors' own creation

In addition, AIG research has discussed opacity and inscrutability as technical challenges stemming from the key attributes of AI algorithms (Kroll, 2018; Larsson and Heintz, 2020). Technical challenges, then, create a need for technical explainable AI (XAI) solutions (Barredo Arrieta *et al.*, 2020; Laato *et al.*, 2022b). As Laato *et al.* (2022b) pointed out, explainability is one of the key design issues related to AIG. Importantly, when implemented and deployed in organizations, AI algorithms are typically components of IS and thus operate in interaction with other AI algorithms and system components. Hence, transparency and explainability need to be considered in technical implementation at both the algorithm and AI system levels.

Considering the embeddedness of AI in IS, the AIG literature includes little discussion on how AIG could be incorporated into AI system design. In general, there is still minimal overlap between the engineering-oriented AI literature and the ethics and regulation literature. Thus, the translation of socially defined requirements into organizationally and technically implemented means of governing AI systems remains a central challenge.

4.2 Lack of attention to the context of AI governance

The second gap is the scant consideration of stakeholders and AIG activities' broader social and political contexts. The heterogeneity of the stakeholders involved in AI system development, operations and governance makes it challenging to identify stakeholder roles (de Almeida *et al.*, 2021). Addressing this challenge is crucial because conflicts related to stakeholders, such as differing interests or an inadequate understanding of roles and responsibilities, can have ethical implications for AI systems (Mökander and Floridi, 2021; Orr and Davis, 2020). For example, according to Cihon *et al.* (2021), considering multiple stakeholders can influence coordination and collaborative efforts toward AIG implementation from a corporate governance perspective.

Our findings reveal that environmental (e.g. social, political and cultural) factors have received little attention in AIG research. This is a significant gap, as efficient AI deployment depends on how well organizations manage the trade-offs between ethical and societal values and organizational objectives, such as profit maximization (Krijger, 2022). Moreover, establishing effective governance solutions typically requires considering environmental factors, such as sociocultural and political dynamics (Cihon *et al.*, 2021; Tsamados *et al.*, 2022). Such contextual and cultural differences can affect the understanding, governance and practical applications of AI technologies (OhÉigeartaigh *et al.*, 2020; Whittlestone *et al.*, 2019). The scarcity of discourse on social and political factors may be attributed to a lack of awareness of how these factors shape AI's impact on humans (Robles Carrillo, 2020). The lack of discussion on the role of culture is somewhat unexpected, as AI as a socio-technical phenomenon is innately embedded in culture and subject to being viewed from different perspectives (Hickok, 2021; Wieringa, 2020).

Regarding the political environment, few studies discuss the implications of existing disparities in power, expertise and cooperation among AI stakeholders. These issues are important in the current economic situation because the competitive advantages of being an AI leader have been recognized (Feijóo *et al.*, 2020). Moreover, while AI development has been posited as critical for societal good (Floridi *et al.*, 2018), the literature pays limited attention to promoting positive societal outcomes, such as sustainable development, financial inclusion and the prevention of corruption (Truby, 2020). The lack of attention to AIG's social and political context gap highlights the need for approaches like SITL and HITL to align the ethical, social and technical perspectives and reduce potential conflicts in the AI development value chain.

4.3 Uncertain effectiveness of ethical principles and regulation

Uncertainty regarding the effectiveness of ethical principles is the third knowledge gap in the AIG literature. Ensuring the ethical use of AI for societal advancement is continually raised

by scholars (ÓhÉigeartaigh *et al.*, 2020), and research on algorithmic ethics has gained traction since 2016 (Schiff *et al.*, 2020; Tsamados *et al.*, 2022). However, there is an ongoing debate on whether the existing ethical guidelines provide sufficient support in practice (Mittelstadt, 2019; Morley *et al.*, 2020) and whether they improve or exacerbate the inherent risks and errors associated with AI (Maas, 2019). There are also substantial differences in the interpretation and practical application of the existing ethical principles (Ulnicane *et al.*, 2021b), which are argued to be procedurally weak (Larsson, 2020) and insufficient for improving AI systems (Hickok, 2021; Reddy *et al.*, 2020). The difficulty of applying ethical principles has been attributed to multiple factors. These include their relative youth and inability to impact governmental policies (Stix, 2021), the influence of existing legislation and the need to understand the context and main audience of AI applications (Morley *et al.*, 2020). These debates are particularly salient for sectors directly impacting human lives, such as healthcare.

Another critical question about AI ethics is the convergence toward principles inherited from bioethics (Zhou *et al.*, 2020). The suitability of applying bioethics principles to AI ethics has been questioned, as the two fields differ significantly (Mittelstadt, 2019; cf. Schiff *et al.*, 2021a). Importantly, AI development lacks crucial characteristics compared to medicine, including professional history, fiduciary duties and proven methods to translate guidelines into practice (Mittelstadt, 2019). This gap suggests the need to adopt a broader societal perspective on ethics. For example, this could be achieved through the consideration of digital ethics (Tsamados *et al.*, 2022), computing ethics and responsible innovation (Ulnicane *et al.*, 2021a).

The applicability challenges of AI ethics have triggered debates on commercial organizations that engage in ethics washing (Koulu, 2020) instead of incorporating ethics in organizational culture and all AI development stages (Zhou *et al.*, 2020). However, empirical studies on the effectiveness of AI ethics are largely missing. In particular, few studies have investigated how professional culture and the heterogeneity of AI stakeholders (Orr and Davis, 2020) affect the application of AI ethics. Current approaches include published ethics documents and organizational ethics training (Ibáñez and Olmeda, 2022; Winfield and Jirotka, 2018), but their effectiveness is uncertain. Scholars have suggested that the scope of principles should include societal values (such as democracy) to counterbalance the dominant commercial interests (Hickok, 2021).

From a regulatory standpoint, some studies consider ethical problems inherent to the technology and promote regulations directed at product liability and data governance (Koulu, 2020). However, practitioners still lack clarity about the applicability of current regulations (Reddy *et al.*, 2020). Overly stringent regulation has also been criticized for stifling innovation and restricting the advancement of nation-states in the AI race (Butcher and Beridze, 2019; Feijóo *et al.*, 2020). In addition to regulation, effective auditing techniques are needed (Truby, 2020), including those with a sector-specific focus (Ibáñez and Olmeda, 2022). Auditing requires professionals who can ensure that AI systems comply with standards (Rahwan, 2018) because, unlike hard law, there is no responsibility mechanism to ensure compliance with ethical principles (Robles Carrillo, 2020).

4.4 Insufficient operationalization of AIG processes

The fourth gap pertains to understanding organization-level AIG processes and the factors affecting their efficacy. While the reviewed studies indicate that processes are integral to AIG, relatively little research has addressed the organizational deployment of these processes. The lack of discussion on organizational AIG is in stark contrast to IT governance and data governance, which include extensive scholarly discussions and frameworks, such as Control Objectives for Information and Related Technologies [COBIT] (cf. Mäntymäki *et al.*, 2022a).

Despite some scholarly attention to organizational AIG processes (Shneiderman, 2020), a lack of understanding persists.

The insufficient operationalization of AIG processes is due to the nascent nature of the field and the limited understanding of stakeholder roles. Although AIG tools and processes (such as ethics-based auditing and impact assessments) have been proposed (Mökander and Floridi, 2021; Raab, 2020), stakeholder roles remain unclear, causing organizational challenges (Mökander and Floridi, 2021; Orr and Davis, 2020). In addition to internal stakeholders, the inputs of various affected stakeholders and community members are required to maintain accountable AIG (Metcalf *et al.*, 2021). While studies have proposed the need to adopt a deliberative approach to foster public governance and responsible innovation (Buhmann and Fieseler, 2021; Morley *et al.*, 2020; Tsamados *et al.*, 2022), it is often unclear who should be involved in this discourse. Further, researchers have warned that impact assessments might become abstract exercises rather than efforts to continually evaluate the real-world problems that accompany AI deployment (Metcalf *et al.*, 2021).

More research is needed to clarify organizational AIG processes and actor roles. Future initiatives should develop formal procedures for operationalizing AI ethics, followed by developers and the overall organization (Ibáñez and Olmeda, 2022). AI is a socio-technical construct, and its governance evolves with the interactions between users and machines (Orr and Davis, 2020). However, scholars have discussed the paucity of methods to operationalize accountability within socio-technical systems (Verdiesen *et al.*, 2021). There is, thus, a need for processes that connect the design, production and implementation stages of AI development with its governance initiatives.

5. Discussion and conclusion

In summary, section 3.1 addressed our RQ1 (the conceptualizations of AIG in the literature), and section 3.2 tackled the first part of RQ2 (main themes in the AIG research). The subsequent section 4, in turn, addressed the latter part of RQ2 (knowledge gaps in AIG research). In the following, we turn to our final research question (RQ3) on future agendas that can be identified based on the literature.

AI has been recognized as a transformative technology, but research on AIG is still in the early stages (Butcher and Beridze, 2019; Perry and Uuk, 2019). Therefore, it is essential for future research to address the knowledge gaps identified in the preceding sections. This will contribute to the development of functional AIG frameworks that facilitate the translation of existing policy into practice. Accordingly, we summarize the insights derived from the literature and the research gaps into a framework of four future AIG agendas. This is followed by a discussion of the limitations of our study and its research and practical implications.

5.1 Future agendas: the four agendas of organizational AI governance

Synthesizing the previous discussion on themes and knowledge gaps, we present an organizational AIG framework with four agendas (technical, stakeholder and contextual, regulatory, and process) as a basis for future research efforts. The organizational perspective is crucial to AIG because AIG research is currently fragmented into subthemes and lacks a common unit of analysis and an integrative view of key processes. Taking an organizational view allows for identifying AIG roles and mechanisms, which strengthens the agency of organizations and individuals. The organizational view also paves the way from conceptual and principle-based approaches to empirical research. Our focus on the organization as the unit for AIG deployment is also based on our findings that indicate an underrepresentation of organizational factors in AIG research. The four interrelated

agendas, in turn, provide an umbrella for organizational processes in relation to technology, stakeholders, and regulation.

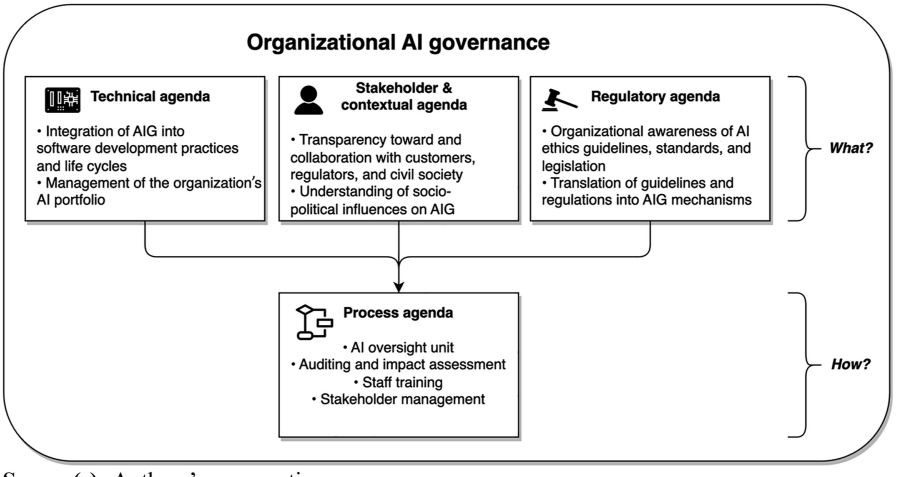
Drawing on the AIG literature published after our search period, we adopt the following working definition of organizational AI governance:

> AI governance is 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; fulfills legal requirements; and meets principles of ethical AI followed by the organization. (Mäntymäki *et al.*, 2022a)

In light of this definition and the previously identified themes and knowledge gaps, we distill organizational AIG into four agendas: technical, stakeholder and contextual, regulatory, and process (see Figure 7). The working definition originates from a source external to our inductive literature review, but the definition and themes exhibit strikingly similar elements: technological tools (technology); principles of ethical AI (stakeholders and context); legal requirements (regulation); and rules, practices and processes (processes). In the following sections, we outline the four agendas from the organizational and research perspectives. While the technical, stakeholder and contextual, and regulatory agendas largely focus on the aims of organizational AIG (i.e. *what*), the process agenda deals with the means to achieve the aims (i.e. *how*).

Our intention is not to present a full-fledged operational AIG framework like, for example, COBIT provides for IT governance. The literature review demonstrated that it is too early to summarize the heterogeneous literature into an operational framework. Moreover, AIG includes factors, such as AI ethics guidelines and broad impact assessments, which differentiate it from traditional IT governance. Hence, we envision AIG as a complement to, rather than a substitute for, IT governance and data governance. AIG should ultimately be linked to other governance processes in organizations. We encourage researchers to develop and refine this generic framework to pursue specific research goals.

5.1.1 Technical agenda. The *technical agenda* of AIG refers to the processes that govern the data, algorithms and algorithmic systems in practice, including their integration into the software development life cycles, the management of the entire AI portfolio and the automation of AIG. The literature includes discussions of technical tools and engineering



Source(s): Authors' own creation

Figure 7.
The four agendas of AI governance

methods for ensuring fairness and transparency (Domanski, 2019; Lysaght *et al.*, 2019). However, organizations also need to integrate these methods into a comprehensive understanding of how they manage their AI portfolios. In particular, organizations need to integrate AIG practices into current software development life cycle models, such as DevOps (Virmani, 2015) to focus on ethical considerations throughout algorithmic life cycles (Ibáñez and Olmeda, 2022; Laato *et al.*, 2022a). Promoting the technical agenda also requires effective dialogue between AI developers and employees responsible for legal compliance and stakeholder engagement.

Within the technical agenda, future research should examine how to incorporate AIG aspects into AI system design. To this end, future research could, for example, evaluate the applicability of different ethical guidelines for AI system design, such as the IEEE's Ethically Aligned Design (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019). In addition to AI system design, future research could explore incorporating AIG into AI system development and operations (DevOps and MLOps) (cf. Laato *et al.*, 2022a; Seppälä *et al.*, 2021). Moreover, future research could examine how to automate AIG activities.

5.1.2 Stakeholder and contextual agenda. The *stakeholder and contextual agenda* refers to stakeholder collaboration to ensure that AI use and development meet their requirements. AIG initiatives should be supported by open and transparent collaboration with external stakeholders. For example, consideration of the ethical concerns of external stakeholders (e.g. civil society and clients) would strengthen the acceptability of choices made by internal stakeholders, such as AI developers and team leaders. External stakeholders' continuous scrutiny of organizational AIG processes could also clarify the long-term challenges for AIG (Liu and Maas, 2021) and its organizational and societal impacts (Häußermann and Lütge, 2022). For instance, the inclusion of stakeholders as intermediaries in high-impact areas (i.e. clinicians in medicine) could improve their commitment to and trust in using AI. This close collaboration would assist in building an AI ecosystem that monitors AI development stages and their integration with inter- and intraorganizational processes to prevent fallouts and improve outcomes (Reddy *et al.*, 2020).

These increased interactions between external and internal stakeholders could advance AIG protocol development by building consensus-driven standards (Butcher and Beridze, 2019). Such collaborative efforts could also be implemented internationally (e.g. between governments and international research organizations). Inclusive collaboration and increased dialogue between stakeholders at the organizational, national and international levels could also lead to the development of a polycentric perspective on AIG (Cihon *et al.*, 2020) and the establishment of an "intelligent discourse regime" to address policy issues and ethical problems associated with AI systems, such as its black box nature (Domanski, 2019).

While the reviewed articles mentioned the roles of internal stakeholders—particularly employees, such as analytics translators (Carter, 2020) and clinicians (Lysaght *et al.*, 2019)—there is little discussion of their roles in AIG. In future research, the entire ecology or network of AIG stakeholders should be studied to understand the roles and functions of stakeholder groups. Developing and operating AI systems is typically a collaborative effort that includes experts such as data scientists, system developers and management (Laato *et al.*, 2021). Hence, future research could examine the roles and interplay of internal stakeholders in AIG efforts. Moreover, technical AI development is seldom a purely in-house effort. Rather, it requires input and resources from external stakeholders, such as AI vendors and consultancies (Minkkinen *et al.*, 2021; Seppälä *et al.*, 2021). Therefore, future research could investigate AIG challenges throughout the AI value chain, including the roles of external stakeholders (Carter, 2020; Truby, 2020). It is especially critical to understand how less-discussed stakeholders (such as investors, the media and third-party suppliers) influence organizational AIG development and success (Buhmann and Fieseler, 2021; Deloitte, 2021; Minkkinen *et al.*, 2022b; Statista, 2021b; Tsamados *et al.*, 2022).

In addition, the reviewed literature exhibits limited consideration of the sociocultural and political environment surrounding AIG. Accordingly, future research should consider AIG mechanisms and processes in relation to political power and sectoral and organizational norms, as these underlying factors may promote or obstruct AIG efforts. We also propose investigating how power concentration among stakeholders could affect AIG initiatives. For instance, a possible aim of future research would be to study the ability of stakeholders (such as civil society) to trust AI development and the efficacy of AIG mechanisms (Zhang and Dafoe, 2020). Investigations should also be directed at examining stakeholder ecology (or networks), including the commensurate degree of involvement in ethical and governance mechanisms (Orr and Davis, 2020).

More broadly, we suggest *collaborative governance* as an AIG research approach to broaden the focus from organizations to networks and ecosystems. Fostering a collaborative network requires identifying internal and external stakeholders and clarifying their roles and responsibilities during the AI system life cycle. Collaborative AIG processes could be improved by adopting approaches from research areas such as Internet and nanotechnology governance (Gasser and Almeida, 2017) and by integrating a corporate social responsibility perspective (Zhou *et al.*, 2020) within AIG mechanisms.

5.1.3 Regulatory agenda. The *regulatory agenda* refers to aligning organizational AIG processes with ethical guidelines and legal requirements. As a baseline, effective organizational AIG requires organizational managers and relevant staff members to be aware of ethical principles, guidelines, standards and legislation. Ensuring awareness of soft and hard regulations will likely require staff training in organizations. The review indicates that legislation is a fundamental facilitator of AIG even though hard law was excluded from the scope of our study. In particular, the role of international law is crucial given the global nature of AI advancements (Aliman and Kester, 2019; Robles Carrillo, 2020; Truby, 2020). The existing legislation, such as the GDPR, is already shaping AIG initiatives and AI development (Viljanen and Parviainen, 2022). A further legislative development is the European AI Act proposed in April 2021 (European Commission, 2021). However, because the regulatory landscape is still developing, there is little knowledge of the effectiveness of hard and soft regulatory measures.

To address the knowledge gap in regulation, future research should examine how AIG policies and regulatory measures translate into actionable AIG mechanisms and how conflicts in the perceived effectiveness of AIG measures can be resolved. In addition to the predominant principles derived from bioethics (Mittelstadt, 2019), AI ethics principles could be adapted based on other fields, such as Internet and digital governance and corporate social responsibility.

5.1.4 Process agenda. The process agenda outlines the means to achieving the other agendas, operating as the glue between the technical, stakeholder and contextual, and regulatory agendas. In particular, senior management should be committed to promoting AIG processes to ensure that organizational culture and strategy support (rather than hinder) AIG efforts.

To effectively promote AIG, organizations that deploy AI should consider establishing an *AI oversight unit* (AOU) to facilitate senior management's commitment to AIG and the practical implementation of the AIG agendas. The AOU role is comparable to the internal data protection officer role required by the EU's GDPR. An organization's data protection officer ensures that the processing of personal data complies with data protection regulations (European Data Protection Supervisor, 2022). The AOU would enable and oversee AIG processes, ensure operational standards are met, and dispense advice. Moreover, it could initiate staff training (e.g. on AI ethics) and stakeholder management connected to AIG. The AOU could also be the first point of contact to address issues raised in the literature, such as detailed rules for good algorithms (Lee, 2020); the level of AIG operationalization

(Schiff *et al.*, 2020); and the identification of the time, stages, and levels of human oversight required in AIG (Koulu, 2020). Our proposal for an organizational AOU is grounded in research that highlights the critical role of professionals who can interrogate algorithms (Rahwan, 2018), as well as institutional review boards (Friesen *et al.*, 2021) and internal review boards to oversee problems and plans (Shneiderman, 2020). However, it is also important to consider the inherent problems that such review boards may encounter, such as a lack of transparency and representativeness (Friesen *et al.*, 2021).

The current AIG research includes little discussion on organizational AIG processes. To address this gap, empirical research is needed on organizational risks, AIG effectiveness factors and risk management mechanisms (Roski *et al.*, 2021). Furthermore, because organizational AIG outcomes are largely unexamined, empirical research is required to understand how organizations deploy AIG processes, such as oversight, auditing and impact assessment, and how individuals enact AIG practices. A more robust understanding can advance organizational and stakeholder consensus on the essential AIG issues and the processes organizations should implement (Boesl and Bode, 2019; Lewis *et al.*, 2020; Whittlestone *et al.*, 2019).

Empirical research is also needed on the factors that support and hinder the success of organizational AIG implementation (Cihon *et al.*, 2020). For example, this could be achieved through surveys and interviews or by adopting a case study approach. Surveys and network analyses could be conducted to understand how external stakeholders (e.g. civil society members) view an organization's AIG efforts, as such stakeholders catalyze AIG advancement (Cihon *et al.*, 2020). Organizational ethics-based auditing could also be examined using a case study approach (Mökander and Floridi, 2021). Moreover, studies can be conducted by independent research organizations, university research groups, or the organization's research and development unit. When conducted at regular intervals, such examinations will also help organizations understand the effects of their AIG initiatives over a specific period, thereby creating a feedback loop.

Regarding success factors, future research should consider how an organization's AIG efforts can be affected by changing operating environments and human resource practices (Domanski, 2019; Hickok, 2021). The resilience of organizational AIG mechanisms and the organization's AIG capabilities should be tested by considering the changing contextual environment. For instance, one such environmental change could be the emergence of the new AI Act considered by the EU, which organizations must implement when it becomes enforceable.

5.2 Limitations

Understandably, our study has some limitations. As with any SLR, the search strategy and the inclusion and exclusion criteria determine the pool of analyzed literature. Including more search databases and additional governance keywords (e.g. from IT governance) could have resulted in a more extensive and comprehensive literature sample. However, as highlighted in the current review, AIG research is in a formative stage, and the use of terminology varies. Thus, applying a broader search strategy and more liberal inclusion and exclusion criteria would have resulted in a larger and more heterogeneous pool of studies.

The search terms notwithstanding, it is surprising that IT governance perspectives were largely absent from the dataset, considering that AI algorithms are typically part of IS. This omission may be due to the formative stage of the AIG discussion, especially at the organizational level. It may also stem partly from the explicit ethics focus adopted in this paper, starting from the so-called principles-to-practice gap (Schiff *et al.*, 2021b).

Moreover, only a subset of the reviewed articles provided explicit definitions of AIG, and these differed from each other. Varying emphases were placed on public policy, ethics and

other elements, making it challenging to synthesize this heterogeneous field of literature. While this conceptual heterogeneity is characteristic of formative areas of inquiry, AIG research could benefit from convergence in terminology and definitions. To address this issue, we have adopted an initial working definition of organizational AIG (Mäntymäki *et al.*, 2022a), but further conceptual synthesis and unifying definitions are required.

Finally, we note that most of the reviewed papers (61 out of 68) were conceptual rather than empirical. This naturally limits the generalizability of the findings because current research provides few empirical insights into the application of AIG. However, the field is rapidly developing, and the number of empirical studies will likely increase in the coming years, calling for subsequent literature reviews once empirical AIG research is more mature.

5.3 Research and practical implications

The aims of this review were to understand how organization-level AIG has been conceptualized (RQ1), identify the key themes and knowledge gaps in AIG research (RQ2) and provide directions for future research (RQ3). Summing up our findings, our SLR has four primary implications for AIG research. First, we established boundaries for conceptualizing organizational AIG, which the review confirmed was unclear (RQ1). We further distilled the themes in the AIG literature and adopted a working definition of organizational AI governance as a starting point for further conceptual clarification and empirical work (RQ2). Second, we outlined the critical knowledge gaps to advance the field of AIG research (continuing RQ2). The limited understanding of AIG implementation, lack of attention to context, questions over the effectiveness of ethical principles and regulations, and insufficient operationalization of processes have made AIG research challenging. Accordingly, the scope of contextual investigations should be broadened. Our understanding of AIG mechanisms' characteristics, effectiveness and determinants should be more detailed, calling for in-depth empirical research. Third, we provided a foundation for research into organizational AIG processes and structures by delineating the AIG concept, complete with four agendas: technical, stakeholder and contextual, regulatory, and process (RQ3). Following this overview, each agenda area could be further investigated and operationalized in subsequent AIG research. In the long term, we expect our framework to facilitate the convergence of terms and themes, particularly in organization-centered AIG research. The fourth and most general implication is that our review points toward a collaborative governance approach to AIG that considers organizations and their AIG practices in the context of stakeholder networks and sociopolitical environments. The shift to collaborative governance has both research and practical implications, requiring organizations to be more transparent about their AI operations and researchers to study the entire ecosystem of AI development and use.

In terms of practical implications, a crucial question is how AIG will be integrated into the overall governance systems of organizations. Thus far, the AIG literature has included little explicit discussion of IT governance and data governance. However, organizational AIG does not exist in a vacuum, and it will most likely draw on existing IT governance concepts and frameworks such as COBIT (Mäntymäki *et al.*, 2022a). It is unclear to what extent IT governance frameworks can accommodate the features of AI technologies and issues, such as learning systems and the emphasis on ethical implications. Irrespective of how AIG and IT governance may be linked in future governance arrangements, organizations need to consider the sufficiency of their IT governance and data governance in light of the themes introduced in this paper.

As a concrete recommendation, we propose the establishment of an organizational AOU to oversee AIG processes and facilitate stakeholder interactions, thereby providing a clear focal point for ensuring the responsible use of AI. While the AOU offers a first point of contact regarding AIG, it does not erase the fact that embedding AIG practices into organizations

requires the commitment of senior management and the adoption of change management to ensure practical implementation instead of checkbox ethics. We further highlight the need for staff training and competencies related to AIG and AI ethics. Increased individual awareness is foundational for AIG because the lack of clarity over AIG issues could diminish their salience for individuals. As indicated by the stakeholder and contextual agenda, we emphasize stakeholder management as an essential aspect of organizational AIG. As organizations implement AI systems on an increasing scale, the ecology of affected stakeholders increases in parallel, resulting in greater complexity of perspectives to be considered. This necessitates processes, competencies and organizational cultures that enable managing stakeholder requirements and ensuring the acceptance of AI technologies.

Although public demand for ethical AI continues to grow, if AI technologies are to benefit individuals, then organizations, society and stakeholders need to be able to trust the technologies and the organizations using them. Academic research should keep pace with the demand and lead discussions on sufficiently broad and practicable AIG approaches.

References

Ain, N.U., Vaia, G., DeLone, W.H. and Waheed, M. (2019), "Two decades of research on business intelligence system adoption, utilization and success – a systematic literature review", *Decision Support Systems*, Vol. 125, doi: 10.1016/j.dss.2019.113113.

Al Zadjali, H. (2020), "Building the right AI governance model in Oman", *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance*, pp. 116-119, doi: 10.1145/3428502.3428516.

Aliman, N.-M. and Kester, L. (2019), "Extending socio-technological reality for ethics in artificial intelligent systems", *Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality*, IEEE, pp. 275-282, doi: 10.1109/AIVR46125.2019.00064.

Aliman, N.-M., Kester, L. and Werkhoven, P. (2019), "XR for augmented utilitarianism", *Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality*, IEEE, pp. 283-285, doi: 10.1109/AIVR46125.2019.00065.

Barn, B.S. (2020), "Mapping the public debate on ethical concerns: algorithms in mainstream media", *Journal of Information, Communication and Ethics in Society*, Vol. 18 No. 1, pp. 38-53, doi: 10.1108/JICES-04-2019-0039.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020), "Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, Vol. 58, pp. 82-115, doi: 10.1016/j.inffus.2019.12.012.

Behera, R.K., Bala, P.K. and Dhir, A. (2019), "The emerging role of cognitive computing in healthcare: a systematic literature review", *International Journal of Medical Informatics*, Vol. 129, pp. 154-166, doi: 10.1016/j.ijmedinf.2019.04.024.

Berente, N., Gu, B., Recker, J. and Santhanam, R. (2021), "Managing artificial intelligence", *MIS Quarterly*, Vol. 45 No. 3, pp. 1433-1450, doi: 10.25300/MISQ/2021/16274.

Boesl, D.B.O. and Bode, M. (2019), "Signaling sustainable robotics – a concept to implement the idea of robotic governance", *Proceedings of the IEEE 23rd International Conference on Intelligent Engineering Systems*, pp. 143-146, doi: 10.1109/INES46365.2019.9109458.

Bu, Q. (2021), "The global governance on automated facial recognition (AFR): ethical and legal opportunities and privacy challenges", *International Cybersecurity Law Review*, Vol. 2 No. 1, pp. 113-145, doi: 10.1365/s43439-021-00022-x.

Buenfil, J., Arnold, R., Abruzzo, B. and Korpela, C. (2019), "Artificial intelligence ethics: governance through social media", *Proceedings of the 2019 IEEE International Symposium on Technologies for Homeland Security*. doi: 10.1109/HST47167.2019.9032907.

Buhmann, A. and Fieseler, C. (2021), "Towards a deliberative framework for responsible innovation in artificial intelligence", *Technology in Society*, Vol. 64, doi: 10.1016/j.techsoc.2020.101475.

Butcher, J. and Beridze, I. (2019), "What is the state of artificial intelligence governance globally?", *The RUSI Journal*, Vol. 164 Nos 5-6, pp. 88-96, doi: 10.1080/03071847.2019.1694260.

Carter, D. (2020), "Regulation and ethics in artificial intelligence and machine learning technologies: where are we now? Who is responsible? Can the information professional play a role?", *Business Information Review*, Vol. 37 No. 2, pp. 60-68, doi: 10.1177/0266382120923962.

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*, Vol. 376, p. 2133, doi: 10.1098/rsta.2018.0080.

Cihon, P., Maas, M.M. and Kemp, L. (2020), "Should artificial intelligence governance be centralised? Design lessons from history", *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 228-234, doi: 10.1145/3375627.3375857.

Cihon, P., Schuett, J. and Baum, S.D. (2021), "Corporate governance of artificial intelligence in the public interest", *Information (Switzerland)*, Vol. 12, p. 7, doi: 10.3390/info12070275.

Corbin, J. and Strauss, A. (2014), *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 4th ed., Sage Publications, Los Angeles, CA.

de Almeida, P.G.R., dos Santos, C.D. and Farias, J.S. (2021), "Artificial intelligence regulation: a framework for governance", *Ethics and Information Technology*, Vol. 23 No. 3, pp. 505-525, doi: 10.1007/s10676-021-09593-z.

Deloitte (2021), "AI governance survey: a shift in the ways companies use and invest in ai brought about by changes in the social landscape", available at: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/gx-risk-ai-governance-survey.pdf (accessed 20 June 2022).

Dignum, V. (2020), "Responsibility and artificial intelligence", in Dubber, M.D., Pasquale, F. and Das, S. (Eds), *The Oxford Handbook of Ethics of AI*, Oxford University Press, pp. 213-231. doi: 10.1093/oxfordhb/9780190067397.013.12.

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.

Erdélyi, O.J. and Goldsmith, J. (2018), "Regulating artificial intelligence: proposal for a global solution", *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 95-101, doi: 10.1145/3278721.3278731.

European Commission (2020), "White paper on artificial Intelligence – a European approach to excellence and trust", available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065 (accessed 1 June 2022).

European Commission (2021), "Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts", available at: https://ec.europa.eu/newsroom/dae/redirection/document/75788 (accessed 4 May 2021).

European Data Protection Supervisor (2022), "Data protection officer (DPO)", available at: https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en (accessed 18 August 2022).

Feijóo, C., Kwon, Y., Bauer, J.M., Bohlin, E., Howell, B., Jain, R., Potgieter, P., Vu, K., Whalley, J. and Xia, J. (2020), "Harnessing artificial intelligence (AI) to increase wellbeing for all: the case for a new technology diplomacy", *Telecommunications Policy*, Vol. 44, p. 6, doi: 10.1016/j.telpol.2020.101988.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. and Srikumar, M. (2020), "Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI", available at: https://ssrn.com/abstract=3518482 (accessed 3 June 2022).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2018), "AI4People—an ethical

framework for a good AI society: opportunities, risks, principles, and recommendations", *Minds and Machines*, Vol. 28 No. 4, pp. 689-707, doi: 10.1007/s11023-018-9482-5.

Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., Lorimer, J., Véliz, C., Hallowell, N., Graham, M., Chan, M.S., Davies, H. and Sallamuddin, T. (2021), "Governing AI-driven health research: are IRBs up to the task?", *Ethics and Human Research*, Vol. 43 No. 2, pp. 35-42, doi: 10.1002/eahr.500085.

Gasser, U. and Almeida, V.A.F. (2017), "A layered model for AI governance", *IEEE Internet Computing*, Vol. 21 No. 6, pp. 58-62, doi: 10.1109/MIC.2017.4180835.

Häußermann, J.J. and Lütge, C. (2022), "Community-in-the-loop: towards pluralistic value creation in AI, or—why AI needs business ethics", *AI and Ethics*, Vol. 2 No. 2, pp. 341-362, doi: 10.1007/s43681-021-00047-2.

Hagendorff, T. (2020), "The ethics of AI ethics – an evaluation of guidelines", *Minds and Machines*, Vol. 30 No. 1, pp. 99-120, doi: 10.1007/s11023-020-09517-8.

Hickok, M. (2021), "Lessons learned from AI ethics principles for future actions", *AI and Ethics*, Vol. 1 No. 1, pp. 41-47, doi: 10.1007/s43681-020-00008-1.

Huang, G., Huang, G.-B., Song, S. and You, K. (2015), "Trends in extreme learning machines: a review", *Neural Networks*, Vol. 61, pp. 32-48, doi: 10.1016/j.neunet.2014.10.001.

Ibáñez, J.C. and Olmeda, M.V. (2022), "Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study", *AI and Society*, Vol. 37, p. 4, doi: 10.1007/s00146-021-01267-0.

Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Vol. 1 No. 9, pp. 389-399, doi: 10.1038/s42256-019-0088-2.

Kaminski, M.E. (2019), "Binary governance: lessons from the GDPR's approach to algorithmic accountability", *Southern California Law Review*, Vol. 92 No. 6, pp. 1529-1616.

Kaplan, A. and Haenlein, M. (2019), "Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business Horizons*, Vol. 62 No. 1, pp. 15-25, doi: 10.1016/j.bushor.2018.08.004.

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J. and Linkman, S. (2009), "Systematic literature reviews in software engineering – a systematic literature review", *Information and Software Technology*, Vol. 51 No. 1, pp. 7-15, doi: 10.1016/j.infsof.2008.09.009.

Koniakou, V. (2023), "From the 'rush to ethics' to the 'race for governance' in Artificial Intelligence", *Information Systems Frontiers*, Vol. 25 No. 1, pp. 71-102, doi: 10.1007/s10796-022-10300-6.

Koulu, R. (2020), "Human control over automation: EU policy and AI ethics", *European Journal of Legal Studies*, Vol. 12 No. 1, pp. 9-46, doi: 10.2924/EJLS.2oi9.oi9.

KPMG (2021), "The shape of AI governance to come", available at: https://home.kpmg/xx/en/home/insights/2020/12/the-shape-of-ai-governance-to-come.html (accessed 20 May 2022).

Krijger, J. (2022), "Enter the metrics: critical theory and organizational operationalization of AI ethics", *AI and Society*, Vol. 37 No. 4, pp. 1427-1437, doi: 10.1007/s00146-021-01256-3.

Kroll, J.A. (2018), "The fallacy of inscrutability", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 376, doi: 10.1098/rsta.2018.0084.

Kroll, J.A. (2021), "Outlining traceability: a principle for operationalizing accountability in computing systems", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 758-771, doi: 10.1145/3442188.3445937.

Laato, S., Mäntymäki, M., Birkstedt, T., Islam, A.K.M.N. and Hyrynsalmi, S. (2021), "Digital transformation of software development: implications for the future of work", in Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Pappas, I. and Mäntymäki, M. (Eds), *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*, Springer International Publishing, pp. 609-621, doi: 10.1007/978-3-030-85447-8_50.

Laato, S., Birkstedt, T., Mäntymäki, M., Minkkinen, M. and Mikkonen, T. (2022a), "AI governance in the system development life cycle: insights on responsible machine learning engineering", *Proceedings of the 1st Conference on AI Engineering – Software Engineering for AI*, ACM, New York.

Laato, S., Tiainen, M., Islam, A.K.M.N. and Mäntymäki, M. (2022b), "How to explain AI systems to end users: a systematic literature review and research agenda", *Internet Research*, Vol. 32 No. 7, doi: 10.1108/INTR-08-2021-0600.

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174, doi: 10.2307/2529310.

Larsson, S. (2020), "On the governance of artificial intelligence through ethics guidelines", *Asian Journal of Law and Society*, Vol. 7 No. 3, pp. 437-451, doi: 10.1017/als.2020.19.

Larsson, S. and Heintz, F. (2020), "Transparency in artificial intelligence", *Internet Policy Review*, Vol. 9 No. 2, doi: 10.14763/2020.2.1469.

Lee, J. (2020), "Access to finance for artificial intelligence regulation in the financial services industry", *European Business Organization Law Review*, Vol. 21 No. 4, pp. 731-757, doi: 10.1007/s40804-020-00200-0.

Lewis, D., Hogan, L., Filip, D. and Wall, P.J. (2020), "Global challenges in the standardization of ethics for trustworthy AI", *Journal of ICT Standardization*, Vol. 8 No. 2, pp. 123-150, doi: 10.13052/jicts2245-800X.823.

Liu, H.Y. and Maas, M.M. (2021), "'Solving for X?' Towards a problem-finding framework to ground long-term governance strategies for artificial intelligence", *Futures*, Vol. 126, doi: 10.1016/j.futures.2020.102672.

Lu, Y. (2019), "Artificial intelligence: a survey on evolution, models, applications and future trends", *Journal of Management Analytics*, Vol. 6 No. 1, pp. 1-29, doi: 10.1080/23270012.2019.1570365.

Lütge, C., Poszler, F., Acosta, A.J., Danks, D., Gottehrer, G., Mihet-Popa, L. and Naseer, A. (2021), "AI4people: ethical guidelines for the automotive sector-fundamental requirements and practical recommendations", *International Journal of Technoethics*, Vol. 12 No. 1, pp. 101-125, doi: 10.4018/IJT.20210101.oa2.

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.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. and Viljanen, M. (2022a), "Defining organizational AI governance", *AI and Ethics*, Vol. 2 No. 4, pp. 603-609, doi: 10.1007/s43681-022-00143-x.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. and Viljanen, M. (2022b), "Putting AI Ethics into practice: the hourglass model of organizational AI governance", available at: https://arxiv.org/abs/2206.00335 (accessed 3 September 2022).

Maas, M.M. (2018), "Regulating for 'normal AI accidents'", *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 223-228, doi: 10.1145/3278721.3278766.

Maas, M.M. (2019), "Innovation-proof global governance for military artificial intelligence? How I learned to stop worrying, and love the bot", *Journal of International Humanitarian Legal Studies*, Vol. 10 No. 1, pp. 129-157, doi: 10.1163/18781527-01001006.

Macrae, C. (2019), "Governing the safety of artificial intelligence in healthcare", *BMJ Quality and Safety*, Vol. 28 No. 6, pp. 495-498, doi: 10.1136/bmjqs-2019-009484.

Mayer, A.-S., Haimerl, A., Strich, F. and Fiedler, M. (2021), "How corporations encourage the implementation of AI ethics", 29*th* European Conference on Information Systems.

Mazurek, G. and Małagocka, K. (2019), "Perception of privacy and data protection in the context of the development of artificial intelligence", *Journal of Management Analytics*, Vol. 6 No. 4, pp. 344-364.

Metcalf, J., Moss, E., Watkins, E.A., Singh, R. and Elish, M.C. (2021), "Algorithmic impact assessments and accountability: the co-construction of impacts", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 735-746, doi: 10.1145/3442188.3445935.

Minkkinen, M. and Mäntymäki, M. (2023), "Discerning between the 'easy' and 'hard' problems of AI governance", *IEEE Transactions on Technology and Society*, Vol. 4 No. 2, pp. 188-194, doi: 10.1109/TTS.2023.3267382.

Minkkinen, M., Zimmer, M.P. and Mäntymäki, M. (2021), "Towards ecosystems for responsible AI: expectations on sociotechnical systems, agendas, and networks in EU documents", Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Pappas, I. and Mäntymäki, M. (Ed.s), *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*, Springer International Publishing, pp. 220-232, doi: 10.1007/978-3-030-85447-8_20.

Minkkinen, M., Laine, J. and Mäntymäki, M. (2022a), "Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks", *Digital Society*, Vol. 1 No. 3, p. 21, doi: 10.1007/s44206-022-00022-2.

Minkkinen, M., Niukkanen, A. and Mäntymäki, M. (2022b), "What about investors? ESG analyses as tools for ethics-based AI auditing", *AI and Society*. doi: 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*, Vol. 25 No. 1, pp. 103-121, doi: 10.1007/s10796-022-10269-2.

Mittelstadt, B. (2019), "Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, Vol. 1 No. 11, pp. 501-507, doi: 10.1038/s42256-019-0114-4.

Mökander, J. and Floridi, L. (2021), "Ethics-based auditing to develop trustworthy AI", *Minds and Machines*, Vol. 31 No. 2, pp. 323-327, doi: 10.1007/s11023-021-09557-8.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L.A. and PRISMA-P Group (2015), "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement", *Systematic Reviews*, Vol. 4 No. 1, doi: 10.1186/2046-4053-4-1.

Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (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*, Vol. 26 No. 4, pp. 2141-2168, doi: 10.1007/s11948-019-00165-5.

ÓhÉigeartaigh, S.S., Whittlestone, J., Liu, Y., Zeng, Y. and Liu, Z. (2020), "Overcoming barriers to cross-cultural cooperation in AI ethics and governance", *Philosophy and Technology*, Vol. 33, pp. 571-593, doi: 10.1007/s13347-020-00402-x.

Orr, W. and Davis, J.L. (2020), "Attributions of ethical responsibility by Artificial Intelligence practitioners", *Information, Communication and Society*, Vol. 23 No. 5, pp. 719-735, doi: 10.1080/1369118X.2020.1713842.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and Moher, D. (2021), "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *BMJ*, Vol. 372, p. n71, doi: 10.1136/bmj.n71.

Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P. and Krogstie, J. (2023), "Toward AI governance: identifying best practices and potential barriers and outcomes", *Information Systems Frontiers*, Vol. 25 No. 1, pp. 123-141, doi: 10.1007/s10796-022-10251-y.

Perry, B. and Uuk, R. (2019), "AI governance and the policymaking process: key considerations for reducing AI risk", *Big Data and Cognitive Computing*, Vol. 3 No. 2, doi: 10.3390/bdcc3020026.

Raab, C.D. (2020), "Information privacy, impact assessment, and the place of ethics", *Computer Law and Security Review*, Vol. 37, doi: 10.1016/j.clsr.2020.105404.

Rahwan, I. (2018), "Society-in-the-loop: programming the algorithmic social contract", *Ethics and Information Technology*, Vol. 20 No. 1, pp. 5-14, doi: 10.1007/s10676-017-9430-8.

Reddy, S., Allan, S., Coghlan, S. and Cooper, P. (2020), "A governance model for the application of AI in health care", *Journal of the American Medical Informatics Association*, Vol. 27 No. 3, pp. 491-497, doi: 10.1093/jamia/ocz192.

Robles Carrillo, M. (2020), "Artificial intelligence: from ethics to law", *Telecommunications Policy*, Vol. 44, p. 6, doi: 10.1016/j.telpol.2020.101937.

Roski, J., Maier, E.J., Vigilante, K., Kane, E.A. and Matheny, M.E. (2021), "Enhancing trust in AI through industry self-governance", *Journal of the American Medical Informatics Association*, Vol. 28 No. 7, pp. 1582-1590, doi: 10.1093/jamia/ocab065.

Schiff, D., Biddle, J., Borenstein, J. and Laas, K. (2020), "What's next for AI ethics, policy, and governance? A global overview", *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 153-158, doi: 10.1145/3375627.3375804.

Schiff, D., Borenstein, J., Biddle, J. and Laas, K. (2021a), "AI ethics in the public, private, and NGO sectors: a review of a global document collection", *IEEE Transactions on Technology and Society*, Vol. 2 No. 1, pp. 31-42, doi: 10.1109/TTS.2021.3052127.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A. and Lennon, M. (2021b), "Explaining the principles to practices gap in AI", *IEEE Technology and Society Magazine*, Vol. 40 No. 2, pp. 81-94, doi: 10.1109/MTS.2021.3056286.

Seppälä, A., Birkstedt, T. and Mäntymäki, M. (2021), "From ethical AI principles to governed AI", *Proceedings of the International Conference on Information Systems*, 2021.

Shah, H. (2018), "Algorithmic accountability", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 376, doi: 10.1098/rsta.2017.0362.

Sharma, G.D., Yadav, A. and Chopra, R. (2020), "Artificial intelligence and effective governance: a review, critique and research agenda", *Sustainable Futures*, Vol. 2, doi: 10.1016/j.sftr.2019.100004.

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*, Vol. 10 No. 4, p. 31, doi: 10.1145/3419764.

Sigfrids, A., Nieminen, M., Leikas, J. and Pikkuaho, P. (2022), "How should public administrations foster the ethical development and use of artificial intelligence? A review of proposals for developing governance of AI", *Frontiers in Human Dynamics*, Vol. 4, doi: 10.3389/fhumd.2022.858108.

Sigov, A., Ratkin, L., Ivanov, L. and Xu, L. (2022), "Emerging enabling technologies for Industry 4.0 and beyond", *Information Systems Frontiers*. doi: 10.1007/s10796-021-10213-w.

Smuha, N.A. (2021), "Beyond a human rights-based approach to AI governance: promise, pitfalls, plea", *Philosophy and Technology*, Vol. 34, pp. 91-104, doi: 10.1007/s13347-020-00403-w.

Stahl, B.C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M. and Wright, D. (2021), "Artificial intelligence for human flourishing – beyond principles for machine learning", *Journal of Business Research*, Vol. 124, pp. 374-388, doi: 10.1016/j.jbusres.2020.11.030.

Statista (2021a), "Statista digital market outlook in-depth: artificial intelligence 2020", available at: https://www.statista.com/study/50485/artificial-intelligence/ (accessed 15 May 2022).

Statista (2021b), "Artificial intelligence (AI)", available at: https://www.statista.com/study/38609/artificial-intelligence-ai-statista-dossier/ (accessed 15 May 2022).

Stilgoe, J. (2018), "Machine learning, social learning and the governance of self-driving cars", *Social Studies of Science*, Vol. 48 No. 1, pp. 25-56, doi: 10.1177/0306312717741687.

Stix, C. (2021), "Actionable principles for artificial intelligence policy: three pathways", *Science and Engineering Ethics*, Vol. 27 No. 1, doi: 10.1007/s11948-020-00277-3.

Tandon, A., Dhir, A., Islam, N. and Mäntymäki, M. (2020), "Blockchain in healthcare: a systematic literature review, synthesizing framework and future research agenda", *Computers in Industry*, Vol. 122, doi: 10.1016/j.compind.2020.103290.

Tandon, A., Dhir, A. and Mäntymäki, M. (2021), "Jealousy due to social media? A systematic literature review and framework of social media-induced jealousy", *Internet Research*, Vol. 31 No. 5, pp. 1541-1582, doi: 10.1108/INTR-02-2020-0103.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019), "Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems, first edition", available at: https://standards.ieee.org/industry-connections/ec/ead-v1/ (accessed 12 June 2022).

Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", *British Journal of Management*, Vol. 14, pp. 207-222.

Truby, J. (2020), "Governing artificial intelligence to benefit the UN sustainable development goals", *Sustainable Development*, Vol. 28 No. 4, pp. 946-959, doi: 10.1002/sd.2048.

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M. and Floridi, L. (2022), "The ethics of algorithms: key problems and solutions", *AI and Society*, Vol. 37 No. 1, pp. 215-230, doi: 10.1007/s00146-021-01154-8.

Ulnicane, I., Eke, D.O., Knight, W., Ogoh, G. and Stahl, B.C. (2021a), "Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies", *Interdisciplinary Science Reviews*, Vol. 46 Nos 1-2, pp. 71-93, doi: 10.1080/03080188.2020.1840220.

Ulnicane, I., Knight, W., Leach, T., Stahl, B.C. and Wanjiku, W.G. (2021b), "Framing governance for a contested emerging technology: insights from AI policy", *Policy and Society*, Vol. 40 No. 2, pp. 158-177, doi: 10.1080/14494035.2020.1855800.

Verdiesen, I., Tubella, A.A. and Dignum, V. (2021), "Integrating comprehensive human oversight in drone deployment: a conceptual framework applied to the case of military surveillance drones", *Information (Switzerland)*, Vol. 12 No. 9, doi: 10.3390/info12090385.

Viljanen, M. and Parviainen, H. (2022), "AI applications and regulation: mapping the regulatory strata", *Frontiers in Computer Science*, Vol. 3, doi: 10.3389/fcomp.2021.779957.

Virmani, M. (2015), "Understanding DevOps and bridging the gap from continuous integration to continuous delivery", *Proceedings of the Fifth International Conference on the Innovative Computing Technology*, IEEE, pp. 78-82, doi: 10.1109/INTECH.2015.7173368.

Webster, J. and Watson, R.T. (2002), "Analyzing the past to prepare for the future: writing a literature review", *MIS Quarterly*, Vol. 26 No. 2, pp. xiii-xxiii, doi: 10.1.1.104.6570.

Whittlestone, J., Alexandrova, A., Nyrup, R. and Cave, S. (2019), "The role and limits of principles in AI ethics: towards a focus on tensions", *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 195-200, doi: 10.1145/3306618.3314289.

Wieringa, M. (2020), "What to account for when accounting for algorithms", *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, January 27-30, doi: 10.1145/3351095.3372833.

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*, Vol. 376, p. 2133, doi: 10.1098/rsta.2018.0085.

Wu, W., Huang, T. and Gong, K. (2020), "Ethical principles and governance technology development of AI in China", *Engineering*, Vol. 6 No. 3, pp. 302-309, doi: 10.1016/j.eng.2019.12.015.

Yeung, K. (2018), "Algorithmic regulation: a critical interrogation", *Regulation and Governance*, Vol. 12 No. 4, pp. 505-523, doi: 10.1111/rego.12158.

Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V.R. and Yang, Q. (2018), "Building ethics into artificial intelligence", *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 5527-5533, doi: 10.24963/ijcai.2018/779.

Zhang, B. and Dafoe, A. (2020), "U.S. public opinion on the governance of artificial intelligence", *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 187-193, doi: 10.1145/3375627.3375827.

Zhang, C. and Lu, Y. (2021), "Study on artificial intelligence: the state of the art and future prospects", *Journal of Industrial Information Integration*, Vol. 23, doi: 10.1016/j.jii.2021.100224.

Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S. and Savage, S. (2020), "A survey on ethical principles of AI and implementations", *2020 IEEE Symposium Series on Computational Intelligence*, pp. 3010-3017, doi: 10.1109/SSCI47803.2020.9308437.

Zimmer, M.P., Minkkinen, M. and Mäntymäki, M. (2022), "Responsible artificial intelligence systems: critical considerations for business model design", *Scandinavian Journal of Information Systems*, Vol. 34 No. 2, 4.

Corresponding author
Matti Mäntymäki can be contacted at: matti.mantymaki@utu.fi