Discerning Between the "Easy" and "Hard" Problems of AI Governance

Matti Minkkinen and Matti Mäntymäki

Abstract—**While there is widespread consensus that artificial intelligence (AI) needs to be governed owing to its rapid diffusion and societal implications, the current scholarly discussion on AI governance is dispersed across numerous disciplines and problem domains. This paper clarifies the situation by discerning two problem areas, metaphorically titled the "easy" and "hard" problems of AI governance, using a dialectic theory synthesis approach. The "easy problem" of AI governance concerns how organizations' design, development, and use of AI systems align with laws, values, and norms stemming from legislation, ethics guidelines, and the surrounding society. Organizations can provisionally solve the "easy problem" by implementing appropriate organizational mechanisms to govern data, algorithms, and algorithmic systems. The "hard problem" of AI governance concerns AI as a general-purpose technology that transforms organizations and societies. Rather than a matter to be resolved, the "hard problem" is a sensemaking process regarding socio-technical change. Partial solutions to the "hard problem" may open unforeseen issues. While societies should not lose track of the "hard problem" of AI governance, there is significant value in solving the "easy problem" for two reasons. First, the "easy problem" can be provisionally solved by tackling bias, harm, and transparency issues. Second, solving the "easy problem" helps solve the "hard problem," as responsible organizational AI practices create virtuous rather than vicious cycles.**

Index Terms—**Artificial intelligence, AI, machine intelligence, machine learning, social implications of technology, governance.**

I. Introduction

There is widespread consensus that artificial intelligence (AI) needs to be governed owing to its rapid diffusion and significant organizational and societal implications, as well as its unintended consequences and systemic risks [1], [2], [3]. AI is a broad term that has several definitions. One often-cited definition describes AI as an information system's ability to interpret data, learn, and use learnings to achieve specific goals through adaptation [4]. In contrast, a broader definition conceptualizes AI as a moving frontier of computational advancements that references human intelligence [5]. Recently, these advancements have been visible, for example, in generative AI technologies such as large language models (LLMs) that are used in the ChatGPT chatbot [6]. AI can thus be seen as a research field, a set of technologies, and a more abstract frontier of technological advancement that is always on the horizon. Research on AI dates back to the 1950s, stemming from the quest to develop machines that approach human capabilities in particular domains. Since its early beginnings, AI research and applications have experienced cycles of boom and bust (so-called AI summers and winters), with a recent resurgence of interest owing to more sophisticated algorithms, low-cost graphics processors, and large databases [7]. The current and envisioned AI application areas are numerous and include manufacturing, healthcare, and finance [8], [9].

While AI governance is evidently needed, the current scholarly discussion surrounding AI governance is dispersed across numerous disciplines and problem domains, with authors raising organizational [10], [11], [12] and societal [13], [14] issues. Researchers have conceptualized AI governance as a layered phenomenon concerning software development teams, organizations, industries, and regulators [3], [15], [16]. The scholarly debate on AI governance would benefit from the structuring of the different problem domains.

Our study contributes to structuring AI governance scholarship by providing a dialectic theory synthesis of the AI governance literature. In the 1990s, David Chalmers published the influential article "Facing Up to the Problem of Consciousness," which popularized the distinction between the easy and hard problems of consciousness [17]. The easy problem concerns the ability to react based on information, which can be explained scientifically. The hard problem concerns the phenomenological experience of consciousness, with which science continually struggles [17]. As an analogy, we suggest distinguishing between the *"easy problem" of AI governance* and the *"hard problem" of AI governance*. We use the terms "easy problem" and "hard problem" metaphorically, hence the quotation marks. The "easy problem" is by no means easy, but it is comparatively more straightforward than the "hard problem." Moreover, the separation into two problem domains is a dialectic simplification intended to facilitate discussion, organizations' strategy work, and national and transnational policy planning. In reality, AI governance issues do not fall neatly into two problem areas; rather, they encompass heterogeneous facets, such as software engineering workflows, standardization, diversity among development teams, and unemployment issues due to automation.

It is important to note that what is proposed here is not a distinction between artificial narrow intelligence (AI applied to specific areas) and artificial general intelligence (autonomous AI that can solve problems in many areas) [4]. The distinction between "easy" and "hard" problems already emerges with

Manuscript received 27 January 2022; revised 31 October 2022; accepted 25 March 2023. Date of publication 17 April 2023; date of current version 13 June 2023. This work was supported by the Artificial Intelligence Governance and Auditing (AIGA) Project funded by Business Finland. *(Corresponding author: Matti Minkkinen.)*

The authors are with the Department of Management and Entrepreneurship, Turku School of Economics, University of Turku, 20014 Turku, Finland (e-mail: matti.minkkinen@utu.fi; matti.mantymaki@utu.fi).

Digital Object Identifier 10.1109/TTS.2023.3267382

the current state of artificial narrow intelligence and does not require us to imagine more fully developed artificial general intelligence.

The distinction between different problems is essential because it helps AI researchers and organizations that use AI to approach different problem areas appropriately, devote sufficient attention to both comparatively "easier" and "harder" problems, and contribute to solving both types of problems through a better understanding of their interlinkages. If the distinction between "easy" and "hard" problems is adopted, there is less risk of misunderstanding when using the concept of AI governance in scholarly articles, conference themes, organizational practice, and other contexts. This paper aims to contribute to this objective. We discuss the "easy" and "hard" problems separately before comparing the two problem domains and concluding with the implications.

II. Methodology

Discerning between the "easy" and "hard" problems of AI governance, we provide a theory synthesis [18] of AI governance scholarship using a dialectic approach [19], [20] as a unifying theoretical lens. A theory synthesis is a type of conceptual paper that strives toward conceptual integration across different literature streams by linking disconnected scholarly contributions in a novel way [18]. As AI governance research is currently dispersed across numerous disciplines and communities [1], integration is necessary for this domain. A theory synthesis differs from a literature review because it can build coherence by introducing new theoretical vocabulary, while a literature review remains within the existing conceptual boundaries [18].

To identify the existing AI governance literature for the synthesis, we searched peer-reviewed academic literature on AI governance from five central academic databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library journal publications, and ACM Digital Library conference proceedings. We used a range of synonyms as search terms because AI governance may be discussed using varying terminology. For AI, we used the terms "artificial intelligence," "AI," "machine learning," "deep learning," and "black box." Each of these terms was coupled with the term "governance" using the AND operator. The search was limited to journal and conference publications in English.

To conduct a coherent theory synthesis, we need to limit the body of literature to works that are similar enough to be synthesized. Therefore, we examined titles, abstracts, and full-text documents in borderline cases to screen the relevant literature. The central inclusion criterion was that articles must address the governance of AI systems executed by human actors. This criterion excludes any literature discussing the use of algorithmic systems to govern human behavior. AI governance thus means the governance *of* AI—not governance *by* AI.

We utilized a dialectic lens to make sense of the complex AI governance literature [19], [20]. At its core, a dialectic approach is concerned with historical context and the processes and relationships among entities and problem areas, as opposed to the characteristics of separate entities [20]. From this perspective, socio-technical phenomena evolve continuously owing to dynamics between opposing tendencies rather than linear trends [21]. In our study, the dialectic approach is used as a sensitizing device to simplify the AI governance literature field. Through this lens, we aim to strike a balance between comprehensibility and the risk of oversimplification. The theoretical simplification into two distinct categories is a key limitation of our dialectic theory synthesis approach. Moreover, because dialectics is a research tradition that emphasizes processual development, any specification of a problem domain, such as AI governance, is bound to a particular point in time rather than universally valid.

A dialectic perspective generally works with two opposing entities, sometimes labeled "thesis" and "antithesis," which are ultimately overcome in a subsequent synthesis [21]. While we present the "easy" and "hard" problems of AI governance as dialectically interlinked problem areas, we do not position either as a thesis or antithesis, because making this distinction does not seem fruitful for the analysis. Moreover, we do not aim for a dialectic synthesis between the two. This is because the topic is so novel, and significant work is still required to understand the opposing dialectic poles that are visible in the literature. Therefore, a synthesis in the dialectic sense would be premature. Eventually, a synthesis of AI governance approaches may be achieved, but this is currently speculative and beyond the scope of this article.

III. Results

A. An Overview of the AI Governance Literature

Overall, the literature highlighted the complexities of governing AI [15], [22]. In particular, AI governance includes different dimensions, themes, and perspectives. For example, several studies discussed the legal and regulatory aspects of AI governance [1], [23], [24], [25], [26], [27], which are indirectly relevant to organizations through regulatory compliance and engagement.

Explicit AI governance definitions have begun to emerge in the 2020s. At the organizational level, AI governance is 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" [12, p. 604]. This definition emphasizes the alignment between organizational practices and requirements from the operational environment. However, a notable trend in the AI governance literature is that many authors consider AI governance to intersect with regulation and public policy. For example, Kaminski [28] discusses collaborative governance in the context of algorithmic accountability as a regulatory system consisting of mechanisms that regulate organizational activities. Butcher and Beridze [1, p. 88] define AI governance as "the mechanisms and processes that shape and govern AI, considering regulation as a legislative subset of governance." Approaches at this societal level tend to focus on broad questions, such as workforce substitution, autonomous weapons in warfare, and the general social acceptance of AI [1], [13], [29].

In sum, AI governance is currently an ambiguous term that can refer to complex problem areas that are relevant to governments [13], global governance arrangements [1], [30], or issues upon which individual organizations can act [12], [31]. Our dialectic theory synthesis aims to give structure to this ambiguity. The following sections briefly outline what we call the "easy" and "hard" problems of AI governance before comparing the two problem areas.

B. The "Easy Problem" of AI Governance

A growing literature stream discusses the organizational challenges related to governing AI systems, which we label the "easy problem" of AI governance [11], [12], [31], [32], [33]. The "easy problem" concerns how organizations' use of AI systems aligns with laws, values, and norms stemming from legislation, ethics guidelines, and the surrounding society. This problem area covers aspects such as organizational mechanisms to govern data, algorithms, and algorithmic systems [11], as well as engineering approaches that seek to ensure responsible AI development [34], [35]. In addition to legal compliance and engineering ethics, AI governance is linked to corporate social responsibility and business ethics concerns [36], [37].

This problem domain is becoming increasingly important as organizations strive to adopt AI to enhance their performance in various areas, including analytics and process automation. For example, AI algorithms can be used in the assessment of job candidates, and markets are growing around algorithmic recruitment. Regulation requires job candidates to be treated fairly; this entails auditing hiring algorithms for biases, which has, in turn, led to a nascent algorithmic auditing industry [38], [39]. Similar questions of fairness and transparency could be found in customer service chatbots and many other examples.

In regard to their AI governance efforts, organizations deploying AI systems in their operations can receive help from service providers, such as AI auditing firms; however, these organizations ultimately carry the responsibility of ensuring alignment between their AI systems and laws, values, and norms [32]. Accountability issues may be thorny in real cases, but the number of actors is nevertheless limited. The primary actors are AI user organizations; AI system providers; and oversight actors, such as auditors. Questions are focused on specific AI systems used by particular organizations.

Laws, values, and norms evolve; therefore, the "easy problem" of AI governance can be solved only provisionally. However, given a particular stage of AI regulation, organizations and AI systems can reach a sufficient level of compliance, which can be verified by auditing and independent oversight, given that the required governance mechanisms are available [34]. Thus, it is plausible that the "easy problem" of AI governance can be (provisionally) solved.

C. The "Hard Problem" of AI Governance

In contrast, the "hard problem" of AI governance concerns AI as a general-purpose technology [13] that transforms organizations, societies, and the lives of individuals. The transformative potential of AI can be seen on a continuum ranging from narrowly transformative AI (comparable to the invention of the typewriter) to transformative AI (comparable to the internal combustion engine) to radically transformative AI (comparable to the industrial revolution) [40]. While the level of transformation remains to be seen, the point is that AI is an emerging technology with societal effects that extend beyond specific organizations and sectors.

At the societal level, the widespread use of AI raises questions about the future of democracy in the context of sophisticated electoral manipulation, the future of work due to increased automation, and warfare with autonomous lethal robots [13]. For example, misinformation, such as the deepfakes produced by generative adversarial networks, can seriously threaten citizens' trust in media sources [41], [42].

Given the global nature of AI technologies and networks, it is unclear who the responsible governing parties are, particularly if the harms caused by AI systems emerge diffusely. Emergent AI harms can be compared to the argument that many privacy harms, akin to environmental harms, arise from individual practices with no malicious intent [43]. The networked nature of accountability, with different actors, forums, and relationships [44], makes attributing responsibilities complex. While single organizations cannot solve networked ethical problems, such as electoral manipulation, technology platforms currently wield a great deal of power to influence AI governance on a macro scale through platform design, monitoring choices, and lobbying regulators.

Moreover, in the case of the "hard problem" of AI governance, setting the boundaries of problems and solutions is challenging. Can we even discuss one complex problem area, or is the "hard problem" composed of many distinct problem areas, such as democracy, work, and international relations? This question logically leads to the follow-up consideration of whether problems in different domains can be solved horizontally—for example, by introducing a regulation such as the European Union's proposed AI Act [45]—or whether piecemeal (e.g., sectoral) solutions are likely to be more effective.

Assessing alternative options is also challenging because it is difficult to define when we are closer to solving the "hard problem" and whether, at some point, we can consider the problem solved. Perhaps the "hard problem" should not be viewed as a matter to be resolved. Instead, the "hard problem" is like an ongoing sensemaking process concerning socio-technical change. In the coming years, AI technologies and their social implications will continue to evolve, and new global situations, such as financial crises, pandemics, and wars, may induce unpredictable applications and impacts of AI technologies. Hence, the large-scale societal issues around AI are an open category rather than knowable in advance. Moreover, during the sensemaking process of tackling the "hard problem," partial solutions may lead to new unforeseen issues [46].

D. Comparing the "Easy" and "Hard" Problems

Table I presents a simplified structured comparison of the different AI governance problem domains, juxtaposing the two using several characteristics. The "easy" and "hard" problems

TABLE I
CHARACTERISTICS OF "EASY" AND "HARD" PROBLEMS OF AI GOVERNANCE

Characteristic	**"Easy problem"**	**"Hard problem"**
Time horizon	Short-term	Long-term
Conception of AI	AI as abilities of information systems [4]	AI as a frontier of computational advancements [5]
General description of relevant issues	Regulatory compliance, governance mechanisms, engineering tools [3], [47]	Large-scale societal challenges [13]
Actors	Organizational managers, developers, executive sponsors [11], internal review boards [14], [16] suppliers, auditors, insurance companies [16], professional associations (e.g., IEEE, ACM) [16], investors [48]	Supranational legislative bodies, nation-states [30], government agencies [15], [16], new institutions (e.g., European Union oversight agency) [49], global and regional initiatives (e.g., World Economic Forum Global Artificial Intelligence Council), industry-led initiatives (e.g., Partnership on AI) [1], [30], research institutions [16], standardization bodies [50], ecosystems [51], [52]
Potential solutions	Organizational strategies and processes [11], organizational ethics guidelines [32], technical tools and methods (e.g., software engineering workflows, audit trails) [16], [47]	Binding legislation [30], human rights frameworks [53], establishment of new global or regional institutions [49]
Ethical basis	Business ethics, corporate social responsibility [36], [37], [51], deontology [54]	Ethics of emerging technologies, data justice, human rights, technology for good [55]
Example problem domains	Organizational governance processes [11], [12], creation of responsible AI-enabled consumer products [56]	Deepfake technologies' undermining of trust in media [41], [42], frameworks for responsible innovation [57]
Illustrative literature examples	[10]–[12], [31]–[33]	[1], [13], [40], [53]

are compared vis-à-vis the time horizon, the conception of AI, relevant actors, types of issues, solutions, and ethical basis. Some example problem domains and literature examples are also given.

The first characteristic, time horizon, is implicit rather than explicit in the literature. Arguably, short- and long-term concerns are involved in both problem areas. However, organizational AI governance tends to be more engaged with the present and the near future—for example, with regulatory compliance and stakeholder pressures here and now, rather than speculating what law and ethical issues could look like in the longer term. In contrast, the "hard problem" takes a longer-term view and relates to how the law should be changed, what new institutions are needed, how to ensure meaningful employment for future generations, and other longer-term concerns.

The conception of AI is also an implicit characteristic that goes together with the time horizon because different aspects of the broad AI phenomenon are relevant at different time-frames. In the short term, the "AI" that is governed is related to information systems with particular abilities, such as machine learning, data interpretation, and adaptation [4]. Over the longer term, governing AI means the governance of an ever-evolving frontier of computation that seeks to attain humanlike capabilities and, possibly, eventually surpass them [5]. From this latter perspective, it is not enough to govern what is possible today (e.g., with machine learning). To deal with the "hard problem," governance must also tackle the broader social implications of machines continuously approaching human-like capabilities. In sum, with respect to the "easy problem," AI is conceived of as a particular set of things (information systems), while for the "hard problem," AI is a more abstract dynamic frontier.

In tandem with the time horizon and conception of AI, the "easy problem" issues tend to be about complying with current legislation and stakeholder pressures and developing the required organizational mechanisms and technical tools. In contrast, the "hard problem" pertains to large-scale societal challenges, such as the future of work as a result of automation and even existential risks related to autonomous weapons [13].

The relevant actors also show the different scopes of the problem areas. The actors for the "easy problem" contain the types of actors necessary for practically implementing AI governance. These include organizational managers and development teams, as well as oversight bodies, such as review boards, investors [48], and professional associations that help organizations. The "hard problem," in turn, is more diffuse, and its solution requires both established actors (e.g., nation-states) and proposed new actors (e.g., new institutions). Numerous actors, such as policy makers, regulators, auditors, standardization bodies such as the International Organization for Standardization (ISO) and the IEEE P7000 groups [50], professional associations, and consumers, are affected by problems and try to mitigate and manage them. In this context, researchers and EU institutions promote an ecosystem approach whereby networked actors attempt to jointly resolve the issues of responsible AI [51], [52]. In responsible AI ecosystems, actors such as auditing firms, regulators, and technology developers could contribute to the shared societal goal of responsible AI with some degree of coordination. At the time of writing, such ecosystems are in a nascent state, and the question of who should orchestrate them is unresolved; however, over time, they may prove to be a mainstay in inter-organizational and global AI governance.

The types of actors are closely linked to the types of solutions that are appropriate for the divergent problem areas. For the "easy problem," solutions operate at two levels: organizational (strategies, processes, guidelines) and technical (tools and methods). Organizational solutions include, for example, appointing an AI governance officer and drafting organizational AI strategies. In contrast, technical tools and methods include, for example, explainable machine learning tools, such as the Local Interpretable Model-agnostic Explanations (LIME) method, as well as broader machine learning pipelines [47]. In contrast, the "hard problem" requires binding regulation; global frameworks, such as human rights frameworks; and potentially new institutions, such as an "International AI Agency." Attempts at regulatory frameworks, such as the EU objective of trustworthy AI [58], and standardization efforts by the ISO and IEEE P7000 groups, are relevant at this level. The "hard problem" of AI governance is thus partly a problem of institutional design, while the "easy problem" is more about organizational or technical design.

Concerning the ethical basis of AI governance, the current approaches to the "easy problem" primarily build on deontology—that is, the obligation to adhere to principles and rules—although virtue ethics has also been promoted as a promising approach [54]. Within the domain of the "hard problem," the issues are linked to the ethics of emerging technologies [59] and similar ethical foresight analyses [60], which anticipate and assess the potential implications of new and emerging technologies and seek to influence their development. Links can also be drawn to the data justice [61], human rights [53], and technology for good [55] discourses. Both the "easy" and "hard" problems are connected to responsible research and innovation [62], which seeks to steer technological development toward socially acceptable goals.

IV. Discussion and Conclusion

This paper has explored the "easy" and "hard" problems of AI governance and clarified the distinction between these two problem areas using a dialectic theory synthesis of the AI governance literature. Our study naturally has some limitations. Synthesizing a heterogeneous set of literature into two problem areas requires a balance to be struck between comprehensibility and the risk of oversimplification. While we cannot do justice to the rich discussions in the AI governance literature streams, we argue that the AI governance field presently needs such broad distinctions to help position the debates and research institutions within it. In fact, it could possibly be beneficial if more nuanced terminology were developed and "AI governance" carried fewer meanings, but at present, this seems unlikely. Therefore, we consider our main contribution to be the distinction between the "easy" and "hard" problems as a sensemaking and communication device for the various scholarly and practitioner communities working on AI governance.

However, the dialectic theoretical lens allows us to go further than distinguishing between two problem domains. Thus far, we have contrasted the two problem areas as separate entities, but a dialectic approach is concerned with the interrelations and dynamics between different entities. How do the "easy" and "hard" problems of AI governance interact with each other? While this question warrants further analysis beyond the scope of this paper, we present a heuristic model of intertwining developments in the two problem areas, visualized as two loops and a further loop connecting them (Fig. 1).

The "easy" and "hard" problems of AI governance can be illustrated as two loops that depict the continuous management of organizational and societal AI governance challenges, respectively. Each loop illustrates the iterative search for solutions, which are different for the two problem domains, as discussed in the previous section. However, an additional loop between the two problem domains can be identified where they contribute to each other. In particular, provisional solutions to the "hard problem" provide new regulations, institutions, and refined problem definitions for organizations to deal with. In turn, the "easy problem" provides emergent solutions as organizations innovate new AI governance mechanisms that are contextually appropriate in their sectors. Over time, these governance efforts can, in the best case, lead to broader solutions that are more than the sum of their parts. For example, technical bias mitigation tools developed by technology companies, such as AI Fairness 360 [63], can contribute to making the policy goal of fairer AI more feasible. Even if this aggregation does not take place, the continuous management of the "easy problem" provides refined and more contextual problem definitions to help understand the "hard problem" from a bottom-up perspective. As both problem domains refine each other's problem definition, the broader loop between them is a dialectic development toward a more systemic understanding of AI governance, driven by both large-scale societal problems and more concrete organizational problems. New generations of AI technologies, such as generative AI in chatbots such as ChatGPT [6], necessitate this continuous loop because they introduce new ethical issues, such as the effective spreading of misinformation in the case of generative AI [64].

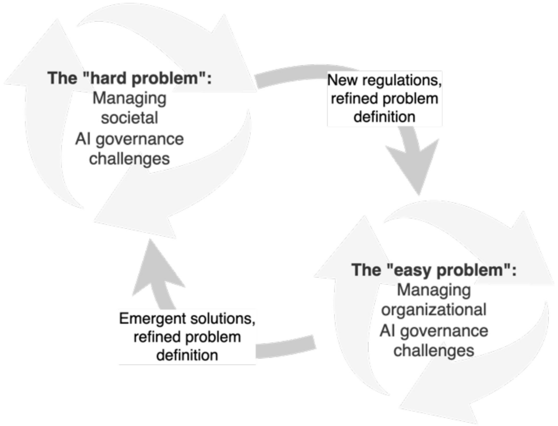

Fig. 1. Connections between the "easy" and "hard" problems of AI governance.

Solving the "hard problems" is crucial in the long run, as these are large-scale challenges threatening preferable socioeconomic development. Nevertheless, there are two key reasons why there is significant value in devoting at least equal scholarly attention to the comparatively easier ones. First, the "easy problem" can plausibly be solved for each development stage of AI and AI regulation, whereas the "hard problem" is more like an umbrella term for an ongoing process toward an ideal that may never be reached but deserves our efforts. Keeping our eyes solely on the "North Star"—that is, the "hard problem"—carries the significant risk of losing track of the current concrete issues, such as biased systems and harm toward particular groups, while navigating AI development and use.

Second, understanding and resolving the "easy problem" helps identify linkages between the problem areas and thus manage the "hard problem." Each piecemeal solution that allows organizations to govern their AI systems and ensure their alignment with laws, values, and norms can create a small virtuous cycle that may ultimately solve more significant problems than was initially anticipated. Conversely, if "easy" AI governance problems are not solved satisfactorily, they may accumulate to form more complex emergent problems. For example, organizations lacking quality assurance mechanisms and offering irresponsible image manipulation products may exacerbate the problems that deepfakes pose vis-à-vis trust [41]. Indeed, portraying the two problem areas as unconnected would be irresponsible, and more work should be devoted to understanding the complex connections between them.

Thus, researchers, companies, and public organizations should take both "easy" and "hard" problems and their interconnections seriously. In other words, we advocate a more systemic approach to AI governance, and our recommendation is twofold. First, neither problem area should be neglected when companies, public organizations, investors, researchers, and funding bodies devote resources to research and practical solutions because progress in both domains is ultimately interlinked. Second, for the same reason, researchers within a particular problem domain should also position their work in relation to other AI governance domains to promote interdisciplinary dialogue. In practice, this could mean including broader implications in organizational AI governance research and, conversely, exploring concrete implementation questions in research devoted to societal AI governance. These recommendations would strengthen the mutual refinement of the problem definitions shown in Fig. 1 and, thus, potentially enable researchers and organizations to concurrently address both "easy" and "hard" problems.

More broadly, societies should not lose track of the "North Star"—that is, ensuring the long-term responsible design, development, and use of AI technologies. It would be shortsighted to focus solely on the present problem of aligning AI system use with current laws, values, and norms at the societal level. As previously mentioned, the actor and accountability networks are complex, with formal regulators, self-regulatory mechanisms, inter-organizational networks, and end users playing their distinct parts [65].

The problems of which AI actors should devote attention to which problem area and how different actors should collaborate require extensive scholarly discussion that extends beyond the scope of this paper. However, two general conclusions can be mentioned. First, a clearer understanding of current AI governance problems helps organizations and policy makers define pathways and take steps toward effective AI governance. Second, problems should be raised and tackled at an appropriate level of governance (development team, organizational, national, or transnational), depending on their complexity.

References

[1] J. Butcher and I. Beridze, "What is the state of artificial intelligence governance globally?" *RUSI J.*, vol. 164, nos. 5–6, pp. 88–96, Sep. 2019, doi: 10.1080/03071847.2019.1694260.

[2] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2.

[3] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Putting AI ethics into practice: The hourglass model of organizational AI governance," Jun. 2022, *arXiv:2206.00335*.

[4] A. Kaplan and M. Haenlein, "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence," *Bus. Horizons*, vol. 62, no. 1, pp. 15–25, Jan./Feb. 2019, doi: 10.1016/j.bushor.2018.08.004.

[5] N. Berente, B. Gu, J. Recker, and R. Santhanam, "Managing artificial intelligence," *MIS Quart.*, vol. 45, no. 3, pp. 1433–1450, Sep. 2021.

[6] M. Jovanović and M. Campbell, "Generative artificial intelligence: Trends and prospects," *Computer*, vol. 55, no. 10, pp. 107–112, Oct. 2022, doi: 10.1109/MC.2022.3192720.

[7] C. Collins, D. Dennehy, K. Conboy, and P. Mikalef, "Artificial intelligence in information systems research: A systematic literature review and research agenda," *Int. J. Inf. Manag.*, vol. 60, Oct. 2021, Art. no. 102383, doi: 10.1016/j.ijinfomgt.2021.102383.

[8] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *J. Ind. Inf. Integr.*, vol. 23, Sep. 2021, Art. no. 100224, doi: 10.1016/j.jii.2021.100224.

[9] P. Kumar, Y. K. Dwivedi, and A. Anand, "Responsible artificial intelligence (AI) for value formation and market performance in healthcare: The mediating role of patient's cognitive engagement," *Inf. Syst. Front.*, to be published, doi: 10.1007/s10796-021-10136-6.

[10] R. Benjamins, A. Barbado, and D. Sierra, "Responsible AI by design in practice," Sep. 2019, *arXiv:1909.12838v2*.

[11] J. Schneider, R. Abraham, C. Meske, and J. Vom Brocke, "Artificial intelligence governance for businesses," *Inf. Syst. Manag.*, to be published, doi: 10.1080/10580530.2022.2085825.

[12] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance," *AI Ethics*, vol. 2, pp. 603–609, Feb. 2022, doi: 10.1007/s43681-022-00143-x.

[13] A. Dafoe, *AI Governance: A Research Agenda*, Future Humanity Inst., Univ. Oxford, Oxford, U.K., 2018.

[14] L. Floridi et al., "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds Mach.*, vol. 28, no. 4, pp. 689–707, Dec. 2018, doi: 10.1007/s11023-018-9482-5.

[15] U. Gasser and V. A. F. Almeida, "A layered model for AI governance," *IEEE Internet Comput.*, vol. 21, no. 6, pp. 58–62, Nov./Dec. 2017, doi: 10.1109/MIC.2017.4180835.

[16] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 4, pp. 1–31, Dec. 2020, doi: 10.1145/3419764.

[17] D. J. Chalmers, "Facing up to the problem of consciousness," *J. Consciousness Stud.*, vol. 2, no. 3, pp. 200–219, 1995.

[18] E. Jaakkola, "Designing conceptual articles: Four approaches," *AMS Rev.*, vol. 10, no. 1, pp. 18–26, Jun. 2020, doi: 10.1007/s13162-020-00161-0.

[19] J. K. Benson, "Organizations: A dialectical view," *Admin. Sci. Quart.*, vol. 22, no. 1, pp. 1–21, Mar. 1977.

[20] M. Farjoun, "Contradictions, dialectics, and paradoxes," in *The SAGE Handbook of Process Organization Studies*, A. Langley and H. Tsoukas, Eds. London, U.K.: SAGE Publ. Ltd., 2016, doi: 10.4135/9781473957954.

[21] A. H. Van de Ven and M. S. Poole, "Explaining development and change in organizations," *Acad. Manag. Rev.*, vol. 20, no. 3, pp. 510–540, 1995, doi: 10.2307/258786.

[22] L. Floridi, "Soft ethics, the governance of the digital and the general data protection regulation," *Phil. Trans. Roy. Soc. A, Math. Phys. Eng. Sci.*, vol. 376, no. 2133, Nov. 2018, Art. no. 20180081, doi: 10.1098/rsta.2018.0081.

[23] D. B. O. Boesl, M. Bode, and S. Greisel, "Drafting a robot manifesto—New insights from the robotics community gathered at the European robotics forum 2018," in *Proc. 27th IEEE Int. Symp. Robot Human Interact. Commun. (RO-MAN)*, Aug. 2018, pp. 448–451, doi: 10.1109/ROMAN.2018.8525699.

[24] D. Doneda and V. A. F. Almeida, "What is algorithm governance?" *IEEE Internet Comput.*, vol. 20, no. 4, pp. 60–63, Jul./Aug. 2016, doi: 10.1109/MIC.2016.79.

[25] O. J. Erdélyi and J. Goldsmith, "Regulating artificial intelligence: Proposal for a global solution," in *Proc. AAAI/ACM Conf. AI Ethics Soc.*, Dec. 2018, pp. 95–101, doi: 10.1145/3278721.3278731.

[26] M. E. Kaminski and G. Malgieri, "Multi-layered explanations from algorithmic impact assessments in the GDPR," in *Proc. Conf. Fairness Accountability Transparency*, Jan. 2020, pp. 68–79, doi: 10.1145/3351095.3372875.

[27] H. Zhang and L. Gao, "Shaping the governance framework towards the artificial intelligence from the responsible research and innovation," in *Proc. IEEE Int. Conf. Adv. Robot. Soc. Impacts (ARSO)*, Oct. 2019, pp. 213–218, doi: 10.1109/ARSO46408.2019.8948762.

[28] M. E. Kaminski, "Binary governance: Lessons from the GDPR's approach to algorithmic accountability," *Southern California Law Rev.*, vol. 92, no. 6, pp. 1529–1616, 2019.

[29] B. W. Wirtz, J. C. Weyerer, and B. J. Sturm, “The dark sides of artificial intelligence: An integrated AI governance framework for public administration,” *Int. J. Public Admin.*, vol. 43, no. 9, pp. 818–829, Jul. 2020, doi: 10.1080/01900692.2020.1749851.

[30] L. Schmitt, “Mapping global AI governance: A nascent regime in a fragmented landscape,” *AI Ethics*, vol. 2, pp. 303–314, Aug. 2021, doi: 10.1007/s43681-021-00083-y.

[31] R. Benjamins, “A choices framework for the responsible use of AI,” *AI Ethics*, vol. 1, no. 1, pp. 49–53, Feb. 2021, doi: 10.1007/s43681-020-00012-5.

[32] A. Seppälä, T. Birkstedt, and M. Mäntymäki, “From ethical AI principles to governed AI,” in *Proc. 42nd Int. Conf. Inf. Syst. (ICIS)*, Austin, TX, USA, 2021, pp. 1–10. [Online]. Available: https://aisel.aisnet.org/icis2021/ai_business/ai_business/10

[33] R. Eitel-Porter, “Beyond the promise: Implementing ethical AI,” *AI Ethics*, vol. 1, no. 1, pp. 73–80, Feb. 2021, doi: 10.1007/s43681-020-00011-6.

[34] M. Brundage et al., “Toward trustworthy AI development: Mechanisms for supporting verifiable claims,” Apr. 2020, *arXiv:2004.07213*.

[35] V. Vakkuri, K.-K. Kemell, and P. Abrahamsson, “ECCOLA—A method for implementing ethically aligned AI systems,” presented at the 46th Euromicro Conf. Softw. Eng. Adv. Appl. (SEAA), Aug. 2020, pp. 195–204, doi: 10.1109/SEAA51224.2020.00043.

[36] A. Buhmann, J. Paßmann, and C. Fieseler, “Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse,” *J. Bus. Ethics*, vol. 163, no. 2, pp. 265–280, May 2020, doi: 10.1007/s10551-019-04226-4.

[37] K. Martin, “Ethical implications and accountability of algorithms,” *J. Bus. Ethics*, vol. 160, no. 4, pp. 835–850, Dec. 2019, doi: 10.1007/s10551-018-3921-3.

[38] M. Sloane. “The algorithmic auditing trap.” OneZero. Mar. 2021. Accessed: Oct. 26, 2022. [Online]. Available: https://onezero.medium.com/the-algorithmic-auditing-trap-9a6f2d4d461d

[39] M. Minkkinen, J. Laine, and M. Mäntymäki, “Continuous auditing of artificial intelligence: A conceptualization and assessment of tools and frameworks,” *DISO*, vol. 1, no. 3, p. 21, Oct. 2022, doi: 10.1007/s44206-022-00022-2.

[40] R. Gruetzemacher and J. Whittlestone, “The transformative potential of artificial intelligence,” *Futures*, vol. 135, Jan. 2022, Art. no. 102884, doi: 10.1016/j.futures.2021.102884.

[41] M. Mustak, J. Salminen, M. Mäntymäki, A. Rahman, and Y. K. Dwivedi, “Deepfakes: Deceptions, mitigations, and opportunities,” *J. Bus. Res.*, vol. 154, Jan. 2023, Art. no. 113368, doi: 10.1016/j.jbusres.2022.113368.

[42] M. Westerlund, “The emergence of deepfake technology: A review,” *Technol. Innov. Manag. Rev.*, vol. 9, no. 11, pp. 40–53, 2019. [Online]. Available: http://doi.org/10.22215/timreview/1282

[43] D. J. Solove, *Understanding Privacy*. Cambridge, MA, USA: Harvard Univ. Press, 2008.

[44] M. Wieringa, “What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability,” in *Proc. Conf. Fairness Accountability Transparency*, Jan. 2020, pp. 1–18, doi: 10.1145/3351095.3372833.

[45] “Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts COM/2021/206 final.” European Commission. Apr. 2021. Accessed: May 4, 2021. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence

[46] S. Ney and M. Verweij, “Messy institutions for wicked problems: How to generate clumsy solutions?” *Environ. Plan. C, Government Policy*, vol. 33, no. 6, pp. 1679–1696, 2015, doi: 10.1177/0263774x15614450.

[47] S. Laato, T. Birkstedt, M. Mäntymäki, M. Minkkinen, and T. Mikkonen, “AI governance in the system development life cycle: Insights on responsible machine learning engineering,” in *Proc. 1st Conf. AI Eng. Softw. Eng. AI*, 2022, pp. 113–123. [Online]. Available: https://doi.org/10.1145/3522664.3528598

[48] M. Minkkinen, A. Niukkanen, and M. Mäntymäki, “What about investors? ESG analyses as tools for ethics-based AI auditing,” *AI Soc.*, to be published, doi: 10.1007/s00146-022-01415-0.

[49] C. Stix, “Foundations for the future: Institution building for the purpose of artificial intelligence governance,” *AI Ethics*, vol. 2, pp. 463–476, Sep. 2021, doi: 10.1007/s43681-021-00093-w.

[50] P. Cihon, *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*, Future Humanity Inst., Oxford, U.K., Apr. 2019.

[51] B. C. Stahl, *Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies* (SpringerBriefs in Research and Innovation Governance). Cham, Switzerland: Springer Int., 2021, doi: 10.1007/978-3-030-69978-9.

[52] M. Minkkinen, M. P. Zimmer, and M. Mäntymäki, “Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame,” *Inf. Syst. Front.*, vol. 25, pp. 103–121, Feb. 2023, doi: 10.1007/s10796-022-10269-2.

[53] N. A. Smuha, “Beyond a human rights-based approach to AI governance: Promise, pitfalls, plea,” *Philos. Technol.*, vol. 34, pp. 91–104, May 2020, doi: 10.1007/s13347-020-00403-w.

[54] T. Hagendorff, “The ethics of AI ethics: An evaluation of guidelines,” *Minds Mach.*, vol. 30, no. 1, pp. 99–120, Mar. 2020, doi: 10.1007/s11023-020-09517-8.

[55] K. Michael, S. Kobran, R. Abbas, and S. Hamdoun, “Privacy, data rights and cybersecurity: Technology for good in the achievement of sustainable development goals,” in *Proc. IEEE Int. Symp. Technol. Soc. (ISTAS)*, Nov. 2019, pp. 1–13, doi: 10.1109/ISTAS48451.2019.8937956.

[56] S. Du and C. Xie, “Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities,” *J. Bus. Res.*, vol. 129, pp. 961–974, May 2021, doi: 10.1016/j.jbusres.2020.08.024.

[57] A. Buhmann and C. Fieseler, “Towards a deliberative framework for responsible innovation in artificial intelligence,” *Technol. Soc.*, vol. 64, Feb. 2021, Art. no. 101475, doi: 10.1016/j.techsoc.2020.101475.

[58] A. Renda, “Europe: Toward a policy framework for trustworthy AI,” in *The Oxford Handbook of Ethics of AI*, M. D. Dubber, F. Pasquale, and S. Das, Eds. Oxford, U.K.: Oxford Univ. Press, 2020, pp. 649–666, doi: 10.1093/oxfordhb/9780190067397.013.41.

[59] P. A. E. Brey, “Anticipating ethical issues in emerging IT,” *Ethics Inf. Technol.*, vol. 14, no. 4, pp. 305–317, May 2012.

[60] L. Floridi and A. Strait, “Ethical foresight analysis: What it is and why it is needed?” *Minds Mach.*, vol. 30, no. 1, pp. 77–97, Mar. 2020, doi: 10.1007/s11023-020-09521-y.

[61] L. Taylor, “What is data justice? The case for connecting digital rights and freedoms globally,” *Big Data Soc.*, vol. 4, no. 2, pp. 1–14, 2017, doi: 10.1177/2053951717736335.

[62] R. von Schomberg, “A vision of responsible research and innovation,” in *Responsible Innovation*, R. Owen, J. Bessant, and M. Heintz, Eds. Chichester, U.K.: Wiley, 2013, pp. 51–74, doi: 10.1002/9781118551424.ch3.

[63] R. K. E. Bellamy et al., “AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias,” *IBM J. Res. Develop.*, vol. 63, nos. 4–5, pp. 1–15, Jul.–Sep. 2019, doi: 10.1147/JRD.2019.2942287.

[64] T. Y. Zhuo, Y. Huang, C. Chen, and Z. Xing, “Exploring AI ethics of ChatGPT: A diagnostic analysis,” 2023, *arXiv:2301.12867*.

[65] R. Clarke, “Regulatory alternatives for AI,” *Comput. Law Security Rev.*, vol. 35, no. 4, pp. 398–409, Aug. 2019, doi: 10.1016/j.clsr.2019.04.008.

Matti Minkkinen received the B.Sc. degree (sociology) from the London School of Economics in 2007, the M.A. degree (European thought) from University College London, the M.A. degree (futures studies) from the University of Turku in 2013, and the Ph.D. degree (futures studies) from the University of Turku in 2020.

He has worked as a Project Researcher and a Senior Researcher in projects covering responsible AI, food, water, and energy security, and public policy foresight and as a University Teacher in futures studies and foresight. He is currently working as a Postdoctoral Researcher of Information Systems Science with the University of Turku.

Dr. Minkkinen is a member of the Association of Professional Futurists (APF) and the winner of the APF’s Most Significant Futures Works Award in 2019.

Matti Mäntymäki received the D.Sc. degree (Econ. & Bus. Adm.) from the University of Turku, Finland.

He is a Professor of Information Systems Science with the University of Turku. He holds an Adjunct Professorship of Information Systems with the University of Oulu, Finland. His research has been published in journals, such as Information Systems Journal, Journal of Systems and Software, Technological Forecasting and Social Change, Journal of Business Research, Information Systems Frontiers, Information Technology and People, and International Journal of Information Management, and Communications of the Association for Information Systems. His research focuses on the societal, organizational, and psycho-social implications of IT.

Dr. Mäntymäki has served as the Chair of IFIP WG6.11 ‘Communication Aspects of the E-World’ from 2012 to 2018 and currently serves as the Vice-Chair. He is a member of the Association for Information Systems.