Scandinavian Journal of Information Systems

Volume 34 | Issue 2 Article 3

12-31-2022

Responsible AI: Concepts, critical perspectives and an Information Systems research agenda

Polyxeni Vassilakopoulou
University of Agder, polyxenv@uia.no

Elena Parmiggiani
Norwegian University of Science and Technology, parmiggi@ntnu.no

Arisa Shollo
Copenhagen Business School, ash.digi@cbs.dk

Miria Grisot
University of Oslo, miriag@ifi.uio.no

© Scandinavian Journal of Information Systems, 2022 34(2), 89-112

Special issue editorial

Responsible AI

Concepts, critical perspectives and an Information Systems research agenda

Polyxeni Vassilakopoulou
University of Agder, Norway
polyxenv@uia.no

Elena Parmiggiani
Norwegian University of Science & Technology, Norway
parmiggi@ntnu.no

Arisa Shollo
Copenhagen Business School, Denmark
ash.digi@cbs.dk

Miria Grisot
University of Oslo, Norway
miriag@ifi.uio.no

Abstract. Being responsible for Artificial Intelligence (AI) harnessing its power while minimising risks for individuals and society is one of the greatest challenges of our time. A vibrant discourse on Responsible AI is developing across academia, policy making and corporate communications. In this editorial, we demonstrate how the different literature strands intertwine but also diverge and propose a comprehensive definition of Responsible AI as the practice of developing, using and governing AI in a human-centred way to ensure that AI is worthy of being trusted and adheres to fundamental human values. This definition clarifies that Responsible AI is not a specific category of AI artifacts that have special properties or can undertake responsibilities, humans are ultimately responsible for AI, for its consequences and for controlling AI development and use. We explain how the four papers included in this special issue manifest different Responsible AI practices and synthesise their findings into an integrative framework that includes business models, services/prod-

ucts, design processes and data. We suggest that IS Research can contribute socially relevant knowledge about Responsible AI providing insights on how to balance instrumental and humanistic AI outcomes and propose themes for future IS research on Responsible AI.

Key words: Artificial Intelligence, Responsible AI, Trustworthy AI, Ethical AI, Human-Centred AI.

1 Introduction

Being responsible for the power that Artificial Intelligence (AI) brings in business and society is one of the greatest challenges of our time. AI has the potential to promote economic growth and social well-being ultimately helping to achieve global sustainability goals (Pedemonte, 2020) and is already transforming work and everyday life. The ongoing digital transformation fuels AI applications with data accelerating their expansion across domains. Managing AI is unlike information technology management in the past as current AI technologies can be inherently inscrutable, can exhibit autonomous behaviours, and can self-evolve due to their learning capacity (Berente et al., 2021). These unique characteristics call for new research studies on how to be responsible for such self-reliant technologies harnessing their power while minimising risks for individuals and society.

The term AI is evocative and inherently open-ended; it has been part of the public discourse for decades inspiring revolutionary visions including enthusiastic and dystopic ones. AI refers to technological artefacts performing the cognitive functions typically associated with humans, including perceiving and learning (McCarthy et al., 2006; Rai et al., 2019). The recent rise of interest on AI is linked to successes in data-driven modelling and especially Machine Learning enabled by data availability and computational power. These brought us in an era of new kinds of sociotechnical systems, where machines that learn join human learning and create original systemic capabilities: AI-infused metahuman systems (Lyytinen et al., 2021). These new types of complex systems with multiple interconnected human and technological actors show much promise but also raise many concerns. Several examples exist of harm caused because the data used to train the machines were partially or incorrectly representing actual phenomena or were incorrectly pre-processed (Benbya et al., 2021; Teodorescu et al., 2021). Human responsibility for AI is difficult to establish in practice as multiple actors are involved with different roles in AI development and use, deciding when and how to use AI for achieving value targets (Shollo et al., 2022), working with data feeds (Parmiggiani et al., 2022), engaging with AI governance (Schneider et al., 2022) besides developing models and overseeing algorithmic performance.

There is a growing body of literature under the general theme of Responsible AI providing normative guidance for developing and using AI responsibly. The literature on Responsible AI is coming both from academia (Arrieta et al., 2020; Dignum, 2019; Mikalef et al., 2022) and from practice including high-tech companies (Google, 2019; IBM, 2020; Microsoft, 2020) and policy makers (European Commission, 2019; OECD, 2019; US General Services Administration, 2022). This literature includes lengthy documents on many different AI aspects deemed instrumental for Responsible AI (for instance, fairness, privacy, explainability, robustness, accountability, inclusiveness). Some of these aspects relate to processes of developing and using AI while other relate to characteristics of AI artifacts. The guidance provided in this body of literature is susceptible to piecemeal operationalisation and implementation as it is difficult to integrate all the different normative statements (Munn, 2022). The pluralism in normative provisions for Responsible AI can also lead to conflicting demands creating dilemmas and paradoxes (Krijger, 2022). Even more importantly, this literature frequently verges towards a limited, technologically deterministic view of what Responsible AI could mean and how it might work (Greene et al., 2019).

Given the major impact that AI can have, it is important to reflect, discuss and develop critical perspectives on Responsible AI including research on issues of power, ideology and institutional change (Bailey & Barley, 2020). A critical approach implies a perspective that problematises and questions deep-seated assumptions (Orlikowski & Baroudi, 1991) related to social issues such as freedom and social control associated with the impact of information technologies (Myers & Klein, 2011). Information Systems (IS) research is an inherently sociotechnical discipline (Sarker et al., 2019) and is well-positioned to address the crossroads of humanistic, organisational, and technical concerns taking a critical perspective on Responsible AI. The overarching aim of Responsible AI is to ensure societal well-being (Dignum, 2019) preventing loss of control for users and developers as well as bias and discrimination for the involved human beings (Kane et al., 2021). IS researchers have already surfaced the unintended consequences of meshing AI-based and human-based ways of working (Pachidi et al., 2021; van den Broek et al., 2021) proposing approaches for meaningful control of AI in practice (Asatiani et al., 2021). IS research can further develop these insights delineating Responsible AI in a way that balances efficiency-oriented instrumental outcomes with principle-oriented humanistic perspectives in a virtuous circle (Sarker et al., 2019). This special issue aims to contribute in this direction. We do so by promoting a critical, user-oriented, and practice-based approach to Responsible AI.

The papers in this special issue engage with actual practices of designing, using, and living with AI. The critical lens adopted is rooted in a concern about practitioners and

users and their involvement in the processes of taking decisions about and within AI-infused systems. This concern is not new; it sits comfortably in the political sensitivity of Scandinavian research (Bergquist et al., 2018) and the Participatory Design tradition (Simonsen & Robertson, 2013). By taking a critical and practice-oriented approach to Responsible AI, the special issue contributes towards an understanding of the processes of including the skills, interests, and experiences of heterogeneous actors (e.g., developers, clerical workers, managers, policy makers, citizens) into the design and deployment of AI ensuring benefits for all human beings, including future generations.

In this introduction, we consolidate the insights we gained about Responsible AI from handling this special issue. We begin by offering key definitions about Responsible AI and related concepts and an overview of different streams in the related discourse. We continue by discussing implications for IS research and we conclude by providing an overview of the insights contributed by each of the papers included in the special issue.

2 Responsible AI concepts and a research agenda

Responsible AI refers to "the development of intelligent systems according to fundamental human principles and values" (Dignum, 2019, p. 6). Academia, policy makers and technology companies have proposed multiple Responsible AI guidelines and principles attending to concerns about the potentially adverse impact of AI on humans and societies. Somewhat recursively, the term Responsible AI is frequently defined through these guidelines and principles that are said to jointly comprise it (Mikalef et al., 2022). For instance, Arrieta and colleagues (2020, p. 83) define Responsible AI as "a series of AI principles to be necessarily met when deploying AI in real applications". As the term is vaguely defined, it is prone to misinterpretations. Responsible AI is sometimes understood as being about entrusting responsibility to AI artifacts. However, Responsible AI is not a way to give machines some kind of responsibility discharging people and organisations (Theodorou & Dignum, 2020), on the contrary, it is about requiring more responsibility from people and organisations. Humans are ultimately responsible for AI, its unintended consequences and for controlling AI development and use (McCoy et al., 2019; Stephanidis et al., 2019; Vassilakopoulou, 2020). AI can be handled as a tool for "enhancing human agency, without removing human responsibility" (Floridi et al., 2018, p. 692). Human responsibility is key for the trajectories AI will take in the coming years; "the machine is us, our processes, an aspect of our embodiment. We can be responsible for machines; they do not dominate or threaten us. We are responsible for boundaries; we are they." (Haraway, 1990, p. 203). Learning machines are an aspect

of our embodiment which we should deeply love and care and accept as our "duty of continuing to care [even] for unwanted consequences" (Latour, 2011).

2.1 Trustworthy, human-centred and ethical AI

Three different streams of research on Responsible AI can be identified. These are linked to different disciplines and academic traditions (Table 1). The first one, draws heavily from computer science proposing approaches for achieving and evaluating specific AI characteristics including explainability, transparency, fairness, reliability, robustness (e.g., Werder et al., 2022; Yang, 2021). These characteristics are treated as requirements to be met in a verifiable way. Scholars in this community frequently use the term 'trustworthy AI' to denote AI that is worthy of being trusted based on evidence for meeting stated requirements (Kaur et al., 2022; Liu et al., 2023). A technical report (ISO/IEC TR 24028:2020) and a recently published standard (ISO/IEC 22989:2022) by the International Organization for Standardization (2020, 2022) establish the terminology and describe key concepts for trustworthiness in AI. An adjacent community engaged in Responsible AI research draws from human computer interaction and human-centred design (Lee et al., 2020; Shneiderman, 2021). Scholars in this community frequently use the term 'human-centred AI' (Shneiderman, 2020; Xu, 2019). Human-centred AI refers to AI amplifying and augmenting human abilities while preserving human control. Work in this stream covers the whole AI lifecycle from conceptualisation to deployment including concerns about arranging systems of software and human actors (for instance human-in-the-loop arrangements). Finally, a third vibrant community engaged in Responsible AI research draws from ethics and philosophy (Eitel-Porter, 2021; Zhu et al., 2022). The work on ethical AI can be paralleled to the work on medical ethics which emerged in the 1960s, although significant differences exist between medicine and AI development (Mittelstadt, 2019). Scholars in this community frequently use the term 'ethical AI' to denote AI that adheres to fundamental human values (for instance, privacy and non-discrimination) and point to the importance of ethical considerations and deliberations in determining legitimate and illegitimate uses of AI, identifying risks and assessing ethical implications.

Reference discipline:	*Responsible AI viewed as:*	*Related concept:*	*Definition:*	*Selected references:*
Computer Science	A set of requirements to be met in a verifiable way	Trustworthy AI	AI worthy of being trusted based on evidence for meeting stated requirements	Werder et al., 2022; Yang, 2021
Human-Computer Interaction	A design approach	Human-Centred AI	AI amplifying and augmenting human abilities while preserving human control	Lee et al., 2020; Shneiderman, 2021
Philosophy and Ethics	Assessment of AI practices and use purposes in the context of moral duty	Ethical AI	AI adhering to human values and ethical considerations determining legitimate and illegitimate use	Eitel-Porter, 2021; Zhu et al., 2022

Table 1. Three different streams of research on Responsible AI

2.2 Comprehensive definition for responsible AI and an agenda for IS research

Responsible AI is a term found in academic writings across different disciplines. The term is also widely used in policy documents, in corporate communications and also in the context of public service delivery (European Commission, 2019; Google, 2019; IBM, 2020; Microsoft, 2020; OECD, 2019; Schmager et al., 2023; US General Services Administration, 2022; Wilson & Van Der Velden, 2022). Especially big technology companies, after being exposed to public criticism, responded by developing and promoting guidelines and frameworks for Responsible AI. However, exactly what Responsible AI means does vary by academic discipline and industry. There is a vibrant, complex discourse on Responsible AI developing on many levels, with the academic,

policy and corporate strands intertwining. The vagueness and multiplicity in Responsible AI conceptualisations fragments efforts and can be counterproductive.

Drawing from literature across disciplines we propose the following comprehensive definition:

> Responsible AI is the practice of developing, using and governing AI in a human-centred way to ensure that AI is worthy of being trusted and adheres to fundamental human values.

This definition makes it clear that Responsible AI is not a specific category of AI artifacts that have special properties or can undertake responsibilities. It is rather a term that points to complex practices that entail: 1) identifying desirable and undesirable applications of AI technologies, 2) defining desirable and undesirable characteristics of these technologies with relevance to specific contexts and use purposes and, 3) instilling responsibility when organising work for these technologies (as designers, developers, managers, policy makers and regulators) and with these technologies.

IS Research can contribute socially relevant knowledge about Responsible AI providing insights on how to balance on the sociotechnical axis of cohesion between instrumental and humanistic AI outcomes (Sarker et al., 2019). Researchers in our field are well-positioned for studying phenomena at the intersection of information systems, organisations and society. Responsible AI is not a philosophical concept nor a formulaic set of requirements. Even more importantly, it should not become a rhetorical tool for ethics-washing (Bietti, 2020) to conveniently and uncritically facilitate business opportunities associated with AI. It can rather be a concerted effort to harness the power of AI for the benefit of societies while minimising risks. IS academics can not only contribute to knowledge, but also, play a key role in educating on Responsible AI (Grøder et al., 2022) the next generation of practitioners.

There are multiple different avenues that IS research on Responsible AI can take. Research can be performed on the situated and contextual aspects of AI technology use, on AI technology production processes, on the macrosocial and institutional mechanisms. At the use level, research is needed to better understand how we can achieve synergies between humans and machines seeking modalities that allow humans to maintain meaningful control and at the same time enjoy the benefits of trustful technologies (but without viewing machines as moral agents). At the technology production level, more studies on the actual work of professionals with different roles in AI design, deployment and monitoring are needed, especially studies investigating the real-world tensions, conflicting demands and dilemmas and their resolutions. Research

is also needed for understanding how power structures shape AI and how AI establishes or reinforces power structures, who gets to benefit and who may be harmed. Such value-related questions need to be answered before we can produce technical solutions and human-friendly designs.

3 Articles in this special issue

The four articles in this special issue provide a diverse set of studies that explore different aspects of responsible AI research in the IS discipline. Each article has a truly sociotechnical perspective and contributes not only to the literature on responsible AI, but also explores responsible AI in relation in the development, use and design processes, the core focus of the Journal. We use the definition of responsible AI previously developed to analyse the papers in this special issue and capture insights on where and how responsible AI manifests in organisations.

In "Responsible Artificial Intelligence Systems. Critical considerations for business models design" Zimmer and colleagues (2022) argue that for companies to build Responsible AI, it is important to have a business model where the value of Responsible AI is clear. The paper offers considerations for how to design such business models. By focusing on business models, they emphasise the practices of defining desirable and undesirable characteristics of AI systems with relevance to specific value propositions. The authors argue that organisations need to create business models for Responsible AI systems, rather than incorporate AI systems into (responsible) business models. Thus, they develop the perspective of designing Responsible AI business models based on the value proposition of Responsible AI systems. Specifically, the paper addresses the challenge of designing a value proposition that solves the tension between commercial interests and social interests in AI/technological innovation and turns Responsible AI into a competitive advantage for organisations. The paper is based on empirical data from industry experts from companies participating in a joint research project on AI governance and auditing. The paper examines design elements and development approaches for RAI business models by focusing on elements such as value proposition, potential customers, key partners and key activities.

In "Strengthening Human Autonomy in the era of autonomous technology", Soma and colleagues (2022) look at Responsible AI in relation to human autonomy, and argue that, since data delimit how autonomous technologies operate, humans should understand and intervene on data to contribute to Responsible AI. AI is a "datanomous technology" where data practices take centre stage and must serve as an entry point to developing responsible AI systems. In particular, the authors discuss the notion of

human autonomy and what it means to be autonomous when it comes to the use of simple autonomous technology, such as autonomous robots that are entering our everyday life (e.g., vacuum cleaner robots, smart insulin pumps). The authors reflect on how these autonomous technologies affect human autonomy. The paper builds on an understanding of human autonomy as relational and situated, going beyond understanding autonomy as a dichotomy (i.e., an individual can be autonomous or not). In a relational and situated view, human autonomy is an emerging property of the situation and circumstances of an individual, including the technologies in the situation. However, while autonomous humans have situational awareness, autonomous technologies operate in response to data and the data available to them, not in response to a situation. The authors propose to conceptualise this as "datanomous technology" to stress the role of data in how a technology operates: these technologies are governed by data (and limited by data). The authors further argue that human autonomy does not depend on controlling technology per se, but rather on understanding how to improve the conditions for datanomy, to increase the chances that the technology operates as intended.

The third paper, entitled "Exploring tensions in Responsible AI in practice. An interview study on AI practices in and for Swedish Public Organizations", by Figueras and colleagues (2022), examines how practitioners perceive Responsible AI. They identify tensions in relation to how ethical principles are interpreted and enacted in design processes. The study concludes that AI practitioners should have more space to reflect on ethical issues throughout the design process in order to design solutions that are responsible. The authors take a view on ethics and design as inseparable activities, arguing that ethical awareness needs to be pervasive in the whole design process and make the responsibility shared among the different involved stakeholders. Based on this, the paper develops the notion of 'ethos tension' to indicate situations where individual, team or organisational ethos are misaligned. The study identifies tensions in several aspects of AI practices in relation to how principles are interpreted and enacted and shows that understanding tensions in practice is crucial for understanding how these affect the design of technology. Overall, the authors advocate for encouraging and giving more space to reflecting and discussing ethical considerations and values in design processes.

In "How can I help you? A chatbot's answers to citizens' information needs" by Verne and colleagues (2022) the authors look at Responsible AI by examining how a chatbot interacts with citizens. The paper is based on a study of conversations between citizens and a chatbot in the context of public welfare services and focuses on how well the chatbot responds to citizens' inquiries. The paper shows that the chatbot responses are influenced by the hidden working of the technology such as training data and predictions rules. The authors argue that technological transparency and accountability are

important aspects of Responsible AI. In addition, they argue for considering responsibility as an attribute of the overall service, and not just of the technology.

Reflecting on the papers of this special issue, we observe that being responsible for AI solutions manifests in a plethora of practices at different levels. Although these studies only scratch the surface of Responsible AI practices, all together they indicate the need for a more integrative approach to Responsible AI. In Figure 1 we synthesise the findings of the four papers into an integrative framework for Responsible AI. According to this framework, being responsible requires on an organisational level considering a Responsible AI business model and a Responsible AI value proposition; on a product/service level considering the behaviour of the intelligent product/service during deployment and operation; on the level of inputs considering the selection and quality of the data that are fed to the system and on a process level, considering responsibility-sharing among those who shape AI systems throughout the design and development process.

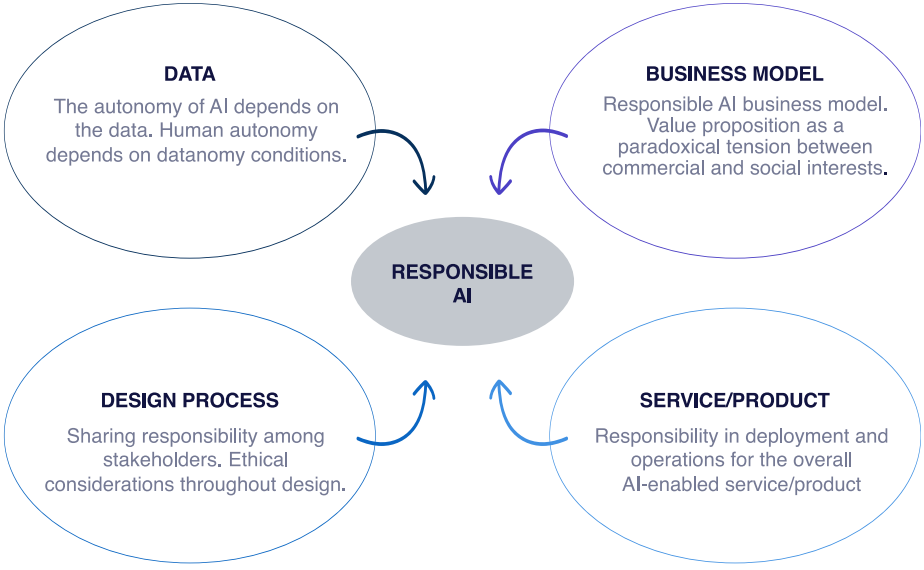

Figure 1. Integrative framework for Responsible AI

In conclusion, we hope that this special issue on Responsible AI will serve as an inspiration to colleagues in IS around the world. This special issue is only one step towards contributing to a more nuanced, practice-oriented, and critical perspective on the social sustainability of complex, opaque, and self-learning technologies such as AI.

References

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

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, *22*(2), 325-352.

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization*, *30*(2), 100286.

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), 281-303.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly*, *45*(3), 1433-1450.

Bergquist, M., Henriksen, H. Z., Ojala, A., & Vassilakopoulou, P. (2018). SJIS Mission. Topical Areas and Research Approaches. *Scandinavian Journal of Information Systems*, *30*(2), 3-4.

Bietti, E. (2020). *From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy* ACM Conference on fairness, accountability, and transparency (FAccT 2020),

Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Springer Nature.

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

European Commission. (2019). *Ethics guidelines for trustworthy AI*. Retrieved 30 November from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., & Rossi, F. (2018). AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707.

Google. (2019). *Responsible AI practices*. Retrieved 30 November from https://ai.google/responsibilities/responsible-ai-practices/

Greene, D., Hoffmann, A. L., & Stark, L. (2019). *Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning* 52nd Hawaii International Conference on System Sciences (HICSS-52),

Grøder, C. H., Schmager, S., Parmiggiani, E., Vassilakopoulou, P., Pappas, I., & Papavlasopoulou, S. (2022). *Educating about Responsible AI in IS: Designing a course based on Experiential Learning* International Conference on Informationn Systems (ICIS 2022),

Haraway, D. (1990). A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s. In L. Nicholson (Ed.), *Feminism/Postmodernism (Thinking Gender)* (pp. 190-233). Routledge.

IBM. (2020). *AI ethics (IBM's multidisciplinary, multidimensional approach helping advance responsible AI)*. Retrieved 30 November from https://www.ibm.com/artificial-intelligence/ethics

International Organization for Standardization. (2020). *ISO/IEC TR 24028:2020 Information technology—Artificial intelligence—Overview of trustworthiness in artificial intelligence*. https://www.iso.org/standard/77608.html

International Organization for Standardization. (2022). *ISO/IEC 22989:2022 Information technology—Artificial intelligence—Artificial intelligence concepts and terminology*. https://www.iso.org/standard/74296.html

Kane, G. C., Young, A. G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS quarterly*, *45*(1), 371-396.

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

Krijger, J. (2022). Enter the metrics: critical theory and organizational operationalization of AI ethics. *Ai & Society*, *37*(4), 1427-1437.

Latour, B. (2011). Love your monsters. *Breakthrough Journal*, *2*(11), 21-28.

Lee, M. K., Grgić-Hlača, N., Tschantz, M. C., Binns, R., Weller, A., Carney, M., & Inkpen, K. (2020). *Human-centered approaches to fair and responsible AI* Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems,

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A., & Tang, J. (2023). Trustworthy AI: a computational perspective. *ACM Transactions on Intelligent Systems and Technology*, *14*(1), 1-59.

Lyytinen, K., Nickerson, J. V., & King, J. L. (2021). Metahuman systems= humans+ machines that learn. *Journal of Information Technology*, *36*(4), 427-445.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

McCoy, L., Burkell, J., Card, D., Davis, B., Gichoya, J., LePage, S., & Madras, D. (2019). *On Meaningful Human Control in High-Stakes Machine-Human Partnerships* UCLA School of Law, The Program on Understanding Law, Science, and Evidence (PULSE), California Digital Library, University of California.,

Microsoft. (2020). *Responsible AI (policies, practices, and tools that make up a framework for Responsible AI by Design)*. Retrieved 30 November from https://www.microsoft.com/en-us/ai/responsible-ai

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

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501-507.

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics*, 1-9.

Myers, M. D., & Klein, H. K. (2011). A set of principles for conducting critical research in information systems. *MIS quarterly*, *35*(1), 17-36.

OECD. (2019). *Artificial Intelligence Principles*. Retrieved 30 November from https://oecd.ai/en/ai-principles

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research*, *2*(1), 1-28.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science*, *32*(1), 18-41.

Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the Backrooms of Data Science. *Journal of the Association for Information Systems*, *23*(1), 139-164.

Pedemonte, C. (2020). *AI for Sustainability: an overview of AI and the SDGs to contribute to the European policy-making*. Retrieved 30 November from https://ec.europa.eu/futurium/en/system/files/ged/vincent-pedemonte_ai-for-sustainability_0.pdf

Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's comments: next-generation digital platforms: toward human–AI hybrids. *MIS quarterly*, *43*(1), iii-x.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *MIS quarterly*, *43*(3), 695-720.

Schmager, S., Grøder, C. H., Parmiggiani, E., Pappas, I., & Vassilakopoulou, P. (2023). *What do citizens think of AI adoption in public services? Exploratory research on citizen attitudes through a social contract lens* 56th Hawaii International Conference on System Sciences (HICSS-56),

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2022). Artificial intelligence governance for businesses. *Information Systems Management*, 1-21.

Shneiderman, B. (2020). Human-centered artificial intelligence: reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

Shneiderman, B. (2021). Responsible AI: Bridging from ethics to practice. *Communications of the ACM*, *64*(8), 32-35.

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. *The Journal of Strategic Information Systems*, *31*(3), 101734.

Simonsen, J., & Robertson, T. (2013). *Routledge international handbook of participatory design*. Routledge

Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., & Fu, L. P. (2019). Seven HCI Grand Challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229-1269.

Teodorescu, M. H., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of fairness ib automation require a deeper understanding of human-ML augmentation *MIS quarterly*, *45*(3), 1483-1500.

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

US General Services Administration. (2022). *Responsible and Trustworthy AI Implementation*. Retrieved 30 November from https://coe.gsa.gov/coe/ai-guide-for-government/responsible-ai-implementation/index.html

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS quarterly*, *45*(3), 1557-1580.

Vassilakopoulou, P. (2020). Sociotechnical Approach for Accountability by Design in AI Systems. European Conference on Information Systems (ECIS 2020),

Werder, K., Ramesh, B., & Zhang, R. (2022). Establishing Data Provenance for Responsible Artificial Intelligence Systems. *ACM Transactions on Management Information Systems (TMIS)*, *13*(2), 1-23.

Wilson, C., & Van Der Velden, M. (2022). Sustainable AI: An integrated model to guide public sector decision-making. *Technology in Society*, *68*, 101926.

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *Interactions*, *26*(4), 42-46.

Yang, Q. (2021). Toward responsible ai: An overview of federated learning for user-centered privacy-preserving computing. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *11*(3-4), 1-22.

Zhu, L., Xu, X., Lu, Q., Governatori, G., & Whittle, J. (2022). AI and Ethics—Operationalizing Responsible AI. In C. Fang & Z. Jianlong (Eds.), *Humanity Driven AI* (pp. 15-33). Springer.