

Managing responsible AI in organizations

Kjersti Berg Danilova¹ and Astrid Klotz¹

¹ NHH Norges Handelshøyskole, Helleveien 30, 5045 Bergen

Abstract. This paper focuses on the intersection between responsible artificial intelligence and organizational management. With the rapid advancement of AI, numerous questions emerge. Some of the most serious questions relate to how AI can be implemented and used responsibly in organizations. Managers play an important role in addressing these concerns. Conversely, the implementation of AI also affects managers, eliciting ethical considerations. Reviewing 28 empirical studies, we examine the current state of research in this field.

Keywords: Responsible Artificial Intelligence, Responsible Digitalization, Management, Literature review

1 Introduction

Artificial intelligence is becoming increasingly vital to organizations, playing a significant role in various tasks and business processes [1]. Increasing computational power and the introduction of generative AI have accelerated this trend to the extent that it can be called a ‘continually evolving frontier’ [2, p. 1433]. AI presents both opportunities and risks to individuals, organizations, and society [3, 4, 5, 6].

Harnessing the benefits of AI while mitigating the risks necessitates responsible design, implementation, and use [7]. Responsible AI should be grounded in an ecosystem involving stakeholders at both organizational and societal levels [8]. Where there is more research on work on responsible AI at societal level, studies on responsible AI in organizations are less prevalent [8]. The ultimate accountability for responsible AI in organizations rests with organizational managers, who have central roles to play in the planning and organization of AI and subsequent changes in organizational structures, business models and processes [2, 9]. AI adoption and use also influences managerial roles, practices, and responsibilities, eliciting ethical considerations.

Responding to a call for research on responsible AI [10], this study seeks to investigate the current state of research on the intersection between responsible AI and organizational management. We review 28 empirical research papers each addressing different facets of this topic, and consequently provide an overview that serves as a knowledge base for practitioners and students, and as a foundation for future research.

2 Responsible digitalization and responsible AI

Responsible digitalization can be seen as an umbrella term that encompasses research areas that overall are concerned with questions related to whether or not digital

technology contributes to making the world better [11], and to how elements of technology design, development, and use determine the answers to such questions. Pappas et al. [12, p. 946] talk about responsible digital transformation, which they describe as “integrating digital technology into a business in a way that is ethical, sustainable, and respectful of human values and society. It involves considering the potential consequences of technological change on individuals and communities and taking steps to minimize any negative impacts.” Trier et al. [13, p. 463] refer to digital responsibility, characterized as “efforts of stakeholders such as individuals, corporations or public institutions to contribute to a sustainable, more inclusive, fair, and value-based digital society (or digitalization in general) beyond the legal minimum”. They categorize digital responsibility into topics such as sustainability, inclusion and participation, ethical design and development of digital technologies, data privacy, transparency, fairness, norms and values, and finally accountability – who is accountable for implementation and use [13]. Such values are also reflected in the AIS shared values for digitalization [14], that AIS proposes as a guideline for researchers, students, and practitioners. We adopt the term responsible digitalization as an overarching term for the responsible design, deployment, and use of digital technologies. Responsible digitalization is founded on digital competence [15] and requires that IS researchers be concerned with ethical issues and the question whether technology can contribute to a better world [11].

Artificial intelligence can be defined as the ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals” [16, p. 3]. AI is nevertheless a term encompassing a rather broad range of systems, that might be classified according to their level of intelligence, where most systems being employed in organizations today typically range from mechanical solutions handling simple, repetitive tasks to analytical systems including machine learning and data analytics [17]. Though AI can bring many benefits to organizations, for instance related to efficiency, quality, and decision-making, research has also disclosed serious dark sides to the use of AI in organizations [6]. Responsible design, implementation, and use of AI thus become central topics in IS research [10]; typically discussed in terms of ethical issues such as accountability, transparency, privacy, justice and fairness, safety, non-maleficence, humanity and sustainability [18, 19, 20].

AI can be employed, framed, and perceived as a technical tool or as a coworker, depending on the type of AI system, the tasks it can perform, and the degree of interaction between the AI and employees and/or customers [21]. As coworkers, humans and AI might work separately and even compete, alternatively complement each other or work interdependently [22]. AI can automate tasks and thus replace human activities, alternatively augment human value by enhancing our ability to perform different tasks with higher efficiency or quality [23]. In organizations today, the varying employ of AI in different contexts often implies that automation and augmentation are intertwined and interdependent [24]. Sabherwal and Grover [25] propose that for the use of AI to benefit society, the use should meet three conditions. First, AI systems should predominantly be used for enhancing rather than automating or replacing human activity, as enhancement of human value has positive effect on society, whereas replacing humans has a negative effect. Second, AI-generated output must be rooted in reality and not

produce false and harmful content. Third, human strengths such as human reasoning and human values must be preserved, keeping humans in the loop [26].

3 Methods

We conducted a systematic literature review based on the recommendations of Okoli & Schabram [27], and Snyder [28]. We selected the keywords for the searches based on terms used in recent organizational AI literature [29, 30]. The keywords employed were leader*, manager*, governance and human-AI-interaction, in combination with “artificial intelligence” or AI.

The search was conducted in May 2024. The search was performed in the Scopus database where we searched the keywords in the article’s title, keywords, and abstract. Considering that we view AI with an information systems lens, we considered the following subject areas: Computer Science; Engineering; Social Science; Business, Management and Accounting; Decision Sciences; Economics, Econometrics, and Finance; and Psychology [27]. We limited the output to journal papers published in English. The search yielded 4780 results, after duplicates were removed. The papers retrieved from the search strings were filtered based on the Chartered Association of Business Schools journal list (Academic Journal Guide, 2021). We selected papers that were published in journals that scored 2, 3, 4, and 4* [31]. Subsequently, 586 papers remained. We screened these for the following inclusion criteria:

- Only empirical studies should be included;
- the topic of the study should lie in the intersection of responsible AI and organizational leadership;
- the study should be related to responsible AI in traditional organizations (papers on algorithmic management of gig work were thus excluded).

After an initial screening of titles and abstract, 114 papers were retained for closer perusal. We read these thoroughly, applying the inclusion and exclusion criteria. Subsequently, 20 papers remained.

In July 2024, we chose to perform an additional search on the search term ‘algorithmic manag*’, following the same procedure, as this term appeared in some of the papers. The initial search yielded 721 results, but after screening for quality and inclusion/exclusion-criteria, only 4 of these ended up being included in the review, as most papers on algorithmic management dealt with gig work through digital platforms and thus did not match the inclusion criteria of our study.

Finally, we scanned the reference lists of the included papers, and conducted a forward search through Google Scholar. We also examined published literature reviews on AI for further references [e.g. 32, 33, 34]. These searches resulted in the inclusion of four additional papers. In total, the review is thus based on 28 papers.

We extracted relevant data from the papers (title, authors, year, journal, method, results) and organized it in an Excel spreadsheet. We conducted a thematic analysis, resulting in four overarching themes, that we present in the next section.

4 Findings

How managers make sense of AI and related ethical concerns impacts AI-adoption. In a survey of 269 managers, Cao et al. [35] studied how managers' attitude and intention to use AI for decision-making is influenced by both positive and 'dark side' variables. They found that managers' attitude towards using AI is negatively influenced by perceived risks and threats, such as the risk that AI might make discriminating or harmful decisions. Such fears were however not significantly related to their intention to use AI. Managers also had personal concerns that the use of AI in decision making might affect their own career and reduce their possibility to learn from their own experience and strengthen their own decision-making abilities. These concerns also negatively affect their intention to use AI for decision-making. Similarly, in a survey of 417 HR managers, Suseno et al. [36] found that managers with positive beliefs about AI were more likely to implement AI, while managers expressing anxiety over AI and changes following the adoption of AI were less likely to do so. Furthermore, HR-practices such as clear job descriptions, training, reward systems and participation might attenuate managers' anxiety and thus strengthen their readiness for AI adoption.

Koponen et al. [18] conducted a case study of a leading financial services provider in Scandinavia where AI could take the form of a technical tool or of a coworker. The study indicated that middle-managers play an important role in balancing different concerns, both ethical and technical and related to job design and task performance. In the case organization, ethical issues related to GDPR represented a worry to the middle-managers, who questioned whether AI-systems were designed to handle such concerns. Still, managers' concerns can depend on the nature of their role and responsibilities. A survey of 73 CIOs in Spanish public city councils indicated a dominant technical focus on AI, while their attention to ethical and human challenges seemed limited [37].

Cuellar et al. [24] found that exposure to information on AI regulation has an impact on managers' perceptions and actions. An online survey experiment involving more than 1200 managers in American businesses showed that when managers are exposed to information on AI regulation, their focus on ethical issues is strengthened, and they are more willing to spend money on developing an AI strategy and on hiring more managers to oversee ethical adoption of AI. The managers had a clear perception of being accountable for the AI-related ethical issues in their organization. But as their awareness of ethical issues increased due to exposure to AI regulations, also their willingness to adopt AI decreased. Managers were prone to using AI in fewer business processes and to slow down the speed of adoption. Moreover, the findings indicated that managers find it easier to handle issues related to safety concerns (as these were perceived as more concrete and thus more easily manageable) than issues such as biases, discrimination and lack of transparency, that might be harder to handle.

Einola et al. [38] studied the introduction of an RPA programmed to automate advertising space bookings in a Nordic media consultancy. The robot was framed as a colleague, given a name, and referred to by managers and employees in human terms. This humanization of the robot also reflected certain expectations, that the robot failed to meet, to the frustration of its human co-workers. In the study it was clear that managerial and employee sensemaking related to the robot, benefits, challenges and time

horizon differed largely and became a source and an amplifier of tension, as managers' and employees did not comprehend each other's perspectives. Although anthropomorphizing an AI-tool can have positive consequences such as increasing customer satisfaction and making customers more forgiving of mistakes [39], in a leadership context, such practices can evidently have more negative outcomes [38, 40]. In another paper by Einola and Khoreva [21], the researchers elaborate on the same case. Their main conclusion is that there is no clear distinction between automation and augmentation; in the complex landscape of organizations these are "“everywhere” and constitutive of each other" (p. 128). The authors suggest that humans and AI become interdependent, and exert mutual influence, shaping each other's choices, actions, and interactions in an iterative process and a 'constantly changing puzzle' (p. 130). Therefore, they conclude that leaders cannot leave AI implementation up to technical experts but must themselves be heavily involved in the process.

The use of AI can influence managers' ethical behavior positively or negatively.

Kipp et al. [41] found that leaders using AI as support in processes producing financial reports are less likely to engage in earnings management, unethical behavior related to financial reporting for their private gain or for their organization to look better [42]. These findings were based on an experiment with 146 managers, exploring how the use of AI influences their moral reasoning and subsequent financial reporting decisions. The study shows that managers are more careful in their financial reporting when the financial information supporting the decision has been gathered by an AI agent than by a human assistant, though less conservative with a more autonomous AI assistant than with a less autonomous. The researchers conclude that these findings are explained by managers' perception of control and of the possibility of diffusing responsibility.

Conversely, AI can impact managers' ethical behavior negatively. Monod et al. [43] conducted a case study of a large Chinese firm introducing an AI sales assistant designed to support salespeople in their work, to boost customer support, sales, and customer satisfaction. Over time, managers realized that the AI tool also could enhance their own work, by using it to surveil and control how employees performed their job. Managers used the tool to gauge employees' attitude, by analyzing their speech, and based on the analysis even to send employees immediate warnings during conversations with customers. The tool was also used to flag employees with lower performance and attitude, displaying a frequently updated overview on managers' dashboards. Employees were thus under continuous surveillance. Increasingly, a secret objective of the managers became employing AI to replace (preferably all) human employees to reduce costs. At the same time, only the managers were fully informed of the actual affordances of the system. To make employees use the tool, employees were provided with limited information on the actual functionality and the negative consequences it held for them and were led to believe that the tool was only designed to assist them. The findings of this study also point to a gap between the intended and unintended consequences of implementing an AI tool and indicate that a development process where both managers and employees had been more closely involved might have helped mitigate this gap.

The use of AI to automate or augment managerial work affects employees, often negatively. Whereas most studies on algorithmic management - the use of AI in managerial processes, such as HRM and decision-making - are experimental or look at gig work [44], a study by Kinowska and Sienkiewi [45] represents one of the few studies that investigates algorithmic management in traditional work contexts. The study was done by analyzing data from the European Company Survey, involving interviews with more than 20.000 senior managers. The researchers found that algorithmic management moderately influences workplace well-being through negative impact on job autonomy and rewards practices. The impact was less significant for larger organizations.

Another negative consequence of introducing AI as a boss involves a potential loss of status. Across five studies involving surveys and experimental methods, Jago et al. [46] found that people perceived that being algorithmically managed compared to having a human manager leads to a lower social status in the organization because it signals a simplification and devaluation of the job.

Through an experiment investigating trust, Höddinghaus et al. [47] found that employees overall see human leaders as more trustworthy than AI leaders, though AI was seen as having higher integrity and (perhaps surprisingly) higher transparency, and human leaders as more benevolent and adaptable. Integrity was considered particularly important in situations when the power distance was more evident, such as in the case of disciplinary decision. Also, Bankins et al. [48] found that employees preferred human managers making decisions, as these were perceived as more competent, fair, and able to show respect and empathy than AI managers. However, employees' evaluation of the fairness of the decision seemed based on the outcome rather than the decision process. When the outcome was positive for the employee, AI was seen as more just and trustworthy than when the outcome was negative. The researchers suggest that these findings can be explained by employees' limited understanding of AI and of how human versus AI decision-making is done, and by a tendency to explain negative outcomes from an 'anti-AI'-perspective.

In an experiment investigating how personnel selection supported by algorithms is perceived [49], 215 participants taking the role of an HR-manager were asked to evaluate applicant preselection, one gender-balanced, the other with a majority of male candidates, and one human-supported preselection and one algorithm-based recommendation. The study found that algorithm-based recommendations were perceived as more consistent and less biased than human decision-support. Still, participants preferred human recommendations as these were perceived as more fair and more trustworthy to uphold moral values. On a positive note, through four experiments involving 1700 participants, Lanz et al. [50] found that employees are less likely to adhere to unethical instructions given by an AI supervisor than by a human one.

Tomprou and Lee [51] found that how employees perceive employer commitments might depend on whether promises given during the recruitment process are delivered by a human or an AI manager. In situations where the organization failed to deliver outcomes promised in the recruitment process, employees seemed to be more accepting if the outcomes had been promised by an AI, thus implying that they perceived these promises as less trustworthy and binding from the start. Employees were then also less likely to quit. These findings pertained to relational commitments (socio-emotional

support), but not to transactional promises (salary, bonuses), that could be expected to be upheld no matter if they were delivered by a human or by AI.

Based on a field experiment in a fintech firm, Qin et al. [52] found that AI can outperform human managers in providing feedback to employees. In the experiment, 196 of 654 employees tasked with making calls to collect overdue payments from borrowers were given regular feedback from AI during a month, whereas the remaining employees received feedback from human managers. AI-feedback was based on audio analytics involving speech-to-text technology and machine learning aiming to evaluate calls compared to different scripts. Human managers had to listen through the calls and evaluate these manually. The study found that employees perceived AI feedback as both more accurate and more fair; free from personal biases against individual employees, as unlike the human managers, the AI had no relationships to the employees. The group receiving feedback from AI ended up making fewer mistakes and achieving higher performance. The study also found, however, that strengthening interpersonal relationship between managers and employees would be the most effective way for human managers of competing with the AI, as the relationship also affected employees' perception of fairness and their belief that the managers hold good intent towards them. This indicates that managers' social and interpersonal skills have become even more important.

In two experiments involving about three hundred employees in total, Pei et al. [53] found that receiving negative feedback from AI rather than from a human leader increased employees' motivation to learn and improved their job performance. The study points to higher quality feedback and a lower interpersonal threat (lower fear of losing face). Likewise, Tong et al. [54] concluded that AI feedback supports employees' learning and performance better than feedback from human managers, due to higher quality in terms of breadth and depth. In a field experiment in which 265 employees in a financial services company were randomly assigned to receive feedback from AI or from a human manager, results showed a higher positive effect from AI feedback on employee performance. Nevertheless, the study also demonstrated that there can be a negative disclosure effect, meaning that employees negative perceptions of AI feedback, lack of trust and fear of being replaced might yield a negative effect on learning and performance. This effect was smaller for employees with longer tenure.

Yam et al. [40] discovered that if an AI tool is humanized and given a name, it may be perceived as more agentic and abusive when it provides negative feedback to employees, and that humanizing AI makes employees more likely to retaliate.

The use of AI increases the challenges and complexities of managerial work. Several studies indicate that introducing AI might increase the challenges and complexity of managerial work, such as ensuring appropriate skills in the organizations, building relations, and leading teams and ensuring collaboration.

Analyzing a public set of US occupational data from 2006 compared with 2016, Huang et al. [55] conclude that we are entering a 'feeling economy' where interpersonal and empathetic tasks will become more important for managers and employees, while AI handles many of the analytical tasks. The authors note that this shift has implications for how managers should think of AI as part of their team and for how they

subsequently design jobs for human versus AI workers, and that in recruitment processes of managers as well as employees, people skills should be recognized as crucial.

Rinta-Kahila et al. [56] studied erosion and eventual restoring of employee skills in a European accountant firm that introduced an AI-based system in a shared-services center targeting a broad client base. The system automated activities including cognitive tasks that had often been challenging and burdensome for the employees. Satisfaction with the system was initially high, as employees experienced that it performed well and freed up time. Overreliance on the system, however, gradually led to erosion of employees' skills, which to begin with did not trouble anyone, as the system delivered consistently. For the organization, the system remained a black box, as they lacked understanding of the system's procedures and activities. Leaving the job to the system however left few incentives for employees to maintain their own domain knowledge. Also, the reliable performance of the system limited employees' incentives to maintain a mindful eye on the output of the system but started to accept and use it without reflection or critical examination. The overreliance on the system and ensuing skill erosion of the employees was not known to the managers but became apparent when the management decided on changes in the IS architecture involving the replacement of the system with a centralized system lacking the same functionality. Subsequently, management had to enforce skill maintenance policies and practices supporting a more mindful relationship to the process and an increase in employees' domain knowledge.

Similarly, Lebowitz et al. [57] warn that to avoid wrongful or harmful AI-founded decision-making, leaders must be able to question the foundation for AI outputs and understand the strengths and limitations of AI as well as the foundation and strengths of human decisions. At the same time, relying on AI output in decision-making processes can limit learning and lead to skill erosion among human experts. Their findings are based on a field-study of managers in the health-sector evaluating machine-learning based AI tools making diagnostics of illnesses. The AI tools did not perform according to expectations, and managers eventually realized that the poor performance could be explained by that the tools were trained on 'know-what'; known facts and rules, but lacked input resembling the know-how, tacit knowledge and intuition of human experts – strengths that AI designers had not able to tap into. Lebowitz et al. conclude that the final decision should always remain with humans, and that leaders must be responsible for ensuring human aspects of the evaluation.

Introducing AI as a team member might necessitate that leaders take extra steps to maintain collaboration within teams. You and Robert [58] found that there is a risk that team members bond with AI to the extent of forming subgroups within the team to the detriment of the overall teamwork quality. Leadership practices promoting communication between the human team members, team building, and building a culture emphasizing the value of human effort and collaboration were found to reduce such risks. Also, leaders should ensure training that would help human team members understand the attributes and limitations of AI.

In a case study of a financial services provider in Scandinavia, Koponen et al. [18] saw that employees experienced a loss of human interaction due to AI adoption. In particular, AI adoption led to a loss of interaction with customers. This increased the need for manager emotional intelligence, to provide social support to employees who

were uncertain or fearful of working with AI or of being replaced. In this organization, humanizing the AI tool was employed as a means of facilitating relations between human employees and their AI 'coworkers'. The introduction of AI can affect not only employees' tasks and responsibilities, but also their professional role identity, a change that might be hard to handle [59]. Increased interaction with AI leads to social disconnection at work, as AI does not offer the kind of feedback and interaction (smiles, facial expressions, and other verbal and nonverbal cues). Though employees might try to compensate by reaching out to colleagues, negative consequences still include increased feeling of loneliness and has impact on employee wellbeing also after working hours (insomnia, alcohol consumption), that leaders need to be aware of [60].

Schulte-Derne and Gnewuch [61] found that operationalizing AI ethics principles was important for managers to appropriately handle challenges related to employees' negative feelings. Though many organizations and public entities have established guidelines and principles, the authors note that these are often vague and challenging to put into practice. In a longitudinal case study of a German energy service provider, they found that principles must be operationalized through concrete measures and practices related to all the central principles: responsibility, justice and fairness, transparency, non-maleficence, privacy [26]. Recommendations include discussing and establishing responsibility, that managers must take steps to understand and address employee concerns and tensions, enabling open communication, striving for a fair distribution of benefits and costs among employees, and allowing employees to see AI in action, to get a deeper comprehension of how AI works.

Though the introduction of AI might increase the complexities of managerial work and increase the need for leadership training, a study by Newstead et al. [62] shows that AI might impact leadership training and education negatively, due to subtle biases in the output of generative AI used for such purposes. Newstead et al. studied 50 AI-generated narratives of more than 40.000 words, where the chatbot was prompted to describe characteristics of good and bad leadership (in general, for male leaders, and for female leaders). Upon analysis of the narratives, the researchers concluded that when the prompts did not specify anything related to gender, all the examples of leadership proposed by the chatbot involved male leaders, thus not accounting for female leaders at all. When prompts asked the chatbot to characterize good and bad leadership for each gender, the output varied largely. Female leaders were more likely to be described as bad, though the reasoning was often weaker for women than for men, meaning the chatbot seemingly had a more lenient perspective towards men. Wording also differed largely in the descriptions of good and bad leadership across genders, making good female leaders sound less impressive than male, and implying that women struggle more exerting leadership. For instance, bad leadership among men was typically ascribed to tyrannic behavior, while bad leadership among women was linked to incompetence. Based on their findings, the researchers assert that leadership training and education should use AI-generated output on leadership with care and consciousness, and that a way of disclosing biases might even be to use AI-chatbots to identify them.

5 Discussion

Responsible digitalization and responsible AI are fundamental topics for organizations today [6, 10]. While a range of actors contribute to responsible AI [8, 63], in organizations, managers hold key roles [2]. In this study, we reviewed 28 empirical research papers on the intersection between responsible AI and organizational management.

Our review identifies four overarching themes in the research. First, ethical concerns are central to managers' considerations regarding the implementation of AI, though managers' awareness of such issues may vary. The studies reviewed highlight the value of regulations in increasing awareness [18] and regulating unethical behavior, such as undisclosed use of AI or hidden intentions [56]. Second, the use of AI might strengthen ethical behavior among managers but may also create new ethical issues for managers. Third, our review of papers dealing with algorithmic management in traditional organizations reveals significant ethical pros and cons associated with AI use in managerial processes. Managers should thus evaluate and carefully determine when and how to integrate AI in managerial processes, and predominantly seek to use AI to enhance rather than to replace human managers' strengths and practices [23, 25]. Fourth, introducing AI in the organization increases the challenges and complexities of managerial work, particularly as protecting employee wellbeing necessitates a conscious and responsible management approach. The responsible implementation and use of AI may thus require a shift in managerial focus, with a stronger emphasis on leading employees and lesser focus on administrative tasks, which could be supported or automated through AI [64].

By investigating the current state of research on the intersection between responsible AI and organizational management, we provide a knowledge base that is valuable to organizational managers as well as to students that in the future might become managers or find themselves in other roles accountable for AI implementation and use. Furthermore, our review lays the groundwork for subsequent research.

Our review reveals that the research in this area is rather limited. Considering the significance of the topic and the rapid pace of technological development, there is clearly a need for more research. We conclude this review with a call for further research on managing responsible AI. Potential topics for future research could include:

- how accountability for responsible AI should be distributed between managers in different roles and in different positions;
- which skills future managers need to lead organizations where AI plays a prominent part and where managers are tasked with ensuring that AI design, implementation and use is done according to principles of responsible AI;
- how AI-regulations concretely should be implemented and utilized for supporting and promoting responsible AI;
- how managers might concretely handle issues such as risks of social disconnect, competency loss, and employees' fear of losing jobs or status;
- as many of the studies published are experimental, further empirical studies on issues and possible solutions related to responsible AI are warranted.

References

1. Brynjolfsson, E. & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), pp.1530-1534.
2. Berente, N., Gu, B., Recker, J. & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly*, 45(3), pp. 1433-1450.
3. Davenport, T.H. & Kirby, J. (2016). Just how smart are smart machines? *MIT Sloan Management Review*, 57(3) pp.21-25.
4. Jarrahi, M. H. (2018). "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making." *Business Horizons* 61, no. 4, pp. 577-586.
5. Mayer, A.S., Strich, F. & Fiedler, M. (2020). Unintended Consequences of Introducing AI Systems for Decision Making. *MIS Quarterly Executive*, 19(4), pp. 239-257.
6. 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), pp.257-268.
7. European Commission (2024). EU Artificial Intelligence Act in Europe. <https://artificialintelligenceact.eu>, Retrieved September 25, 2024.
8. Minkkinen, M., Zimmer, M.P. & Mäntymäki, M. (2023). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*, 25(1), pp.103-121.
9. Jorzik, P., Yigit, A., Kanbach, D.K., Kraus, S. & Dabić, M. (2023). Artificial intelligence-enabled business model innovation: Competencies and roles of top management. *IEEE transactions on engineering management*, 71, pp.7044-7056.
10. Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y.K., Mäntymäki, M. & Pappas, I.O. (2023). Artificial intelligence (AI) and information systems: perspectives to responsible AI. *Information Systems Frontiers*, 25(1), pp.1-7.
11. Walsham, G. (2012). Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *Journal of Information Technology*, 27(2), pp.87-93.
12. Pappas, I.O., Mikalef, P., Dwivedi, Y.K., Jaccheri, L. & Krogstie, J. (2023). Responsible digital transformation for a sustainable society. *Information Systems Frontiers*, 25(3), pp.945-953.
13. Trier, M., Kundisch, D., Beverungen, D., Müller, O., Schryen, G., Mirbabaie, M. & Trang, S. (2023). Digital Responsibility: A Multilevel Framework for Responsible Digitalization. *Business & Information Systems Engineering*, 65(4), pp.463-474.
14. AIS (2024). AIS Shared Values for Digitalization. <https://aisnet.org/page/SharedValuesforDigitalization>
15. Iden, J., Danilova, K.B., & Osmundsen, K. (2022). *Digitaliseringsledelse*. Fagbokforlaget.
16. Mikalef, P. & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3), p.103434.
17. Huang, M.H. & Rust, R.T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), pp.155-172.
18. Cuéllar, M.F., Larsen, B., Lee, Y.S. & Webb, M. (2024). Does information about AI regulation change manager evaluation of ethical concerns and intent to adopt AI?. *The Journal of Law, Economics, and Organization*, 40(1), pp.34-75.

19. Heyder, T., Passlack, N. & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, 32(3), p.101772.
20. Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, 1(9), pp.389-399.
21. Koponen, J., Julkunen, S., Laajalahti, A., Turunen, M. & Spitzberg, B. (2023). Work characteristics needed by middle managers when leading AI-integrated service teams. *Journal of Service Research*, p.10946705231220462.
22. Sowa, K., Przegalinska, A. & Ciechanowski, L. (2021). Cobots in knowledge work: Human-AI collaboration in managerial professions. *Journal of Business Research*, 125, pp.135-142.
23. Raisch, S. & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1), pp.192-210.
24. Einola, K. & Khoreva, V., 2023. Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. *Human Resource Management*, 62(1), pp.117-135.
25. Sabherwal, R. & Grover, V. (2024). The societal impacts of generative artificial intelligence: A balanced perspective. *Journal of the Association for Information Systems*, 25(1), pp.13-22.
26. Grönsund, T. & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), p.101614.
27. Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*.
28. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339.
29. Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, 45(2), 159–182.
30. Cortellazzo, L., Bruni, E., & Zampieri, R. (2019). The Role of Leadership in a Digitalized World: A Review. *Frontiers in Psychology*, 10, 1938.
31. Rowe, F. (2014). What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems*, 23(3), pp.241-255.
32. Lippert, I. (2024). Artificial Intelligence and The Future of Managerial Work: A Theoretical Review of Managerial Roles. *ECIS 2024 Proceedings*.
33. Lee, M.C., Scheepers, H., Lui, A.K. & Ngai, E.W. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. *Information & Management*, 60(5), p.103816.
34. Malik, A., Budhwar, P. & Kazmi, B.A. (2023). Artificial intelligence (AI)-assisted HRM: Towards an extended strategic framework. *Human Resource Management Review*, 33(1), p.100940.
35. Cao, G., Duan, Y., Edwards, J.S. & Dwivedi, Y.K. (2021). Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. *Technovation*, 106, p.102312.
36. Suseno, Y., Chang, C., Hudik, M., & Fang, E. S. (2022). Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: the moderating

- role of high-performance work systems. *International Journal of Human Resource Management*, 33(6), 1209–1236.
37. Criado, J.I. & de Zarate-Alcarazo, L.O. (2022). Technological frames, CIOs, and Artificial Intelligence in public administration: A socio-cognitive exploratory study in Spanish local governments. *Government Information Quarterly*, 39(3), p.101688.
 38. Einola, K., Khoreva, V. & Tienari, J. (2023). A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace. *Human Relations*, p.00187267231206328.
 39. Yam, K.C., Bigman, Y.E., Tang, P.M., Ilies, R., De Cremer, D., Soh, H. & Gray, K. (2021). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology*, 106(10), p.1557.
 40. Yam, K.C., Goh, E.Y., Fehr, R., Lee, R., Soh, H. & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology*, 102, p.104360.
 41. Kipp, P.C., Curtis, M.B. & Li, Z. (2020). The attenuating effect of intelligent agents and agent autonomy on managers' ability to diffuse responsibility for and engage in earnings management. *Accounting Horizons*, 34(4), pp.143-164.
 42. Schipper, K. (1989). Commentary on earnings management. *Accounting Horizons*, 3(4).
 43. Monod, E., Mayer, A.S., Straub, D., Joyce, E. & Qi, J. (2024). From worker empowerment to managerial control: The devolution of AI tools' intended positive implementation to their negative consequences. *Information & Organization*, 34(1), p.100498.
 44. Parent-Rochelleau, X. & Parker, S.K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*, 32(3), p.100838.
 45. Kinowska, H. & Sienkiewicz, Ł.J. (2023). Influence of algorithmic management practices on workplace well-being—evidence from European organisations. *Information Technology & People*, 36(8), pp.21-42.
 46. Jago, A.S., Raveendhran, R., Fast, N. & Gratch, J. (2024). Algorithmic management diminishes status: An unintended consequence of using machines to perform social roles. *Journal of Experimental Social Psychology*, 110, p.104553.
 47. Höddinghaus, M., Sondern, D. & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, 116, p.106635.
 48. Bankins, S., Formosa, P., Griep, Y. & Richards, D. (2022). AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, 24(3), pp.857-875.
 49. Feldkamp, T., Langer, M., Wies, L. & König, C.J. (2024). Justice, trust, and moral judgments when personnel selection is supported by algorithms. *European Journal of Work and Organizational Psychology*, 33(2), pp.130-145.
 50. Lanz, L., Briker, R. & Gerpott, F.H. (2024). Employees adhere more to unethical instructions from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics*, 189(3), pp.625-646.
 51. Tomprou, M. & Lee, M.K. (2022). Employment relationships in algorithmic management: A psychological contract perspective. *Computers in Human Behavior*, 126, p.106997.

52. Qin, S., Jia, N., Luo, X., Liao, C. & Huang, Z. (2023). Perceived fairness of human managers compared with artificial intelligence in employee performance evaluation. *Journal of Management Information Systems*, 40(4), pp.1039-1070.
53. Pei, J., Wang, H., Peng, Q. & Liu, S. (2024). Saving face: Leveraging artificial intelligence-based negative feedback to enhance employee job performance. *Human Resource Management*, 63(5), pp.775-790.
54. Tong, S., Jia, N., Luo, X. & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), pp.1600-1631.
55. Huang, M.H., Rust, R. & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence. *California Management Review*, 61(4), pp.43-65.
56. Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W. & Ruissalo, J., (2023). The vicious circles of skill erosion: a case study of cognitive automation. *Journal of the Association for Information Systems*, 24(5), pp.1378-1412.
57. Lebovitz, S., Levina, N. & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating ai tools based on experts' know-what. *MIS quarterly*, 45(3).
58. You, S. & Robert, L.P. (2023). Subgroup formation in human–robot teams: A multi-study mixed-method approach with implications for theory and practice. *Journal of the Association for Information Science and Technology*, 74(3), pp.323-338.
59. Strich, F., Mayer, A.S. & Fiedler, M. (2021). What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems*, 22(2), p.9.
60. Tang, P.M., Koopman, J., Mai, K.M., De Cremer, D., Zhang, J.H., Reynders, P., Ng, C.T.S. & Chen, I. (2023). No person is an island: Unpacking the work and after-work consequences of interacting with artificial intelligence. *Journal of Applied Psychology*, 108(11), 1766–1789.
61. Schulte-Derne, D. & Gnewuch, U. (2024). Translating AI Ethics Principles into Practice to Support Robotic Process Automation Implementation. *MIS Quarterly Executive*, 23(2), p.6.
62. Newstead, T., Eager, B. & Wilson, S. (2023). How AI can perpetuate—or help mitigate—gender bias in leadership. *Organizational Dynamics*, 52(4), p.100998.
63. Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P. & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers*, 25(1), pp.123-141.
64. Kolbjørnsrud, V., Amico, R. & Thomas, R.J. (2016). How artificial intelligence will redefine management. *Harvard Business Review*, 2(1), pp.3-10.