



Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence

Thomas Herrmann¹ · Sabine Pfeiffer²

Received: 21 June 2021 / Accepted: 24 January 2022 / Published online: 18 February 2022
© The Author(s) 2022

Abstract

The human-centered AI approach posits a future in which the work done by humans and machines will become ever more interactive and integrated. This article takes human-centered AI one step further. It argues that the integration of human and machine intelligence is achievable only if human organizations—not just individual human workers—are kept “in the loop.” We support this argument with evidence of two case studies in the area of predictive maintenance, by which we show how organizational practices are needed and shape the use of AI/ML. Specifically, organizational processes and outputs such as decision-making workflows, etc. directly influence how AI/ML affects the workplace, and they are crucial for answering our first and second research questions, which address the pre-conditions for keeping humans in the loop and for supporting continuous and reliable functioning of AI-based socio-technical processes. From the empirical cases, we extrapolate a concept of “keeping the organization in the loop” that integrates four different kinds of loops: AI use, AI customization, AI-supported original tasks, and taking contextual changes into account. The analysis culminates in a systematic framework of keeping the organization in the loop look based on interacting organizational practices.

Keywords Artificial intelligence · Machine learning · Human-centered AI · Organizational practices · Socio-technical design · Predictive maintenance · Digital work

1 Introduction

With new technologies such as machine learning, the question of the division of labor between man and machine is being raised anew. There is an ongoing discussion of how the advance of artificial intelligence (AI), especially machine learning (ML), will change the work of employees (Brödner 2018; Brynjolfsson and Mitchell 2017; Webb 2020). There are serious concerns about potential employment loss and also, more generally, about whether businesses and other

organizations are losing their connection to human intelligence and creativity.

Numerous theoretical approaches try to capture how the division of labor between humans and machines is changing, looking at the degree of respective autonomy and at the distribution of decision-making competencies. In both engineering and the social sciences, these approaches focus on technology on the one hand and on human individuals as users on the other. They widely neglect the fact that both the use of technology and the decisions generated in this interplay of humans and technology are embedded in human organizations. Organizations—be they commercial enterprises or public institutions—are subject to their own logic, integrated into complex external environments, and divided internally by competition among departments with divergent interests. No matter whether decisions are technically or humanly generated, they must be negotiated and processed within the organization. This study shows how organizational embeddedness can be incorporated into the design concepts of AI-driven human–machine decisions.

This article builds on the “human-centered artificial intelligence” perspective, which shows how and why humans

✉ Thomas Herrmann
thomas.herrmann@rub.de

Sabine Pfeiffer
sabine.pfeiffer@fau.de

¹ Information and Technology Management, Institut für Arbeitswissenschaft, Ruhr-Universität Bochum, 44780 Bochum, Germany

² Chair of Sociology (Technology – Labour – Society), Nuremberg Campus of Technology (NCT), Friedrich-Alexander-Universität Erlangen-Nürnberg, 90429 Nürnberg, Germany

are and for the foreseeable future will continue to be “in the loop” in organizations that use AI. Most of these approaches share a socio-technical perspective, which refers to the mutual, crossover influences of social and technological factors in organizations. Socio-technical design emphasizes the *relevance* of the organization and argues that the dimensions of human individuals, technology, and organization *should* be integrated (Baxter and Sommerville 2011; Eason 1989; Kirsch et al. 1995; Mumford 2006). Authors promoting the human-centered AI perspective have not yet done so systematically.

The suggestion that we need to go beyond the micro-level injunction of “keeping the human in the loop” was advanced already by Rahwan (2018), who proposes keeping “society in the loop” e.g., by implementing negotiation mechanisms between stakeholders with different values. There is no reason this idea could not apply also to the meso level, i.e., to the organization in which AI is applied. By “organization”, we refer to all the organizational practices of task management in which AI is utilized, the organizational actors responsible for evaluating and controlling AI, and the organizational structures, logics, and hierarchies in which all these are embedded. We consider these organizational practices and their social interactions as one dimension of socio-technical systems. Most scholars consider socio-technical systems as the combination of the social and the technical dimension. We suggest that the intertwinement between both and its evolution are the most crucial challenges of socio-technical design (Fischer and Herrmann 2015; Herrmann et al. 2021). Human–computer interaction (HCI) plays a major role for this intertwinement and therefore also in our study. In this socio-technical context we suggest the concept of “keeping the organization in the loop” that complements “keeping the human in the loop.” We define this concept as set of managerial measures and interactively sustained organizational conventions that:

- (a) ensure regulation, control, and supervision of AI-based processes and effects by allowing and making workers competent to be in control, and
- (b) continuously answer the question of the extent to which organizational coordination is a subject of AI-based decision-making.

We characterize AI as a simulation of decision-making (Jarrahi 2018). In cases where decision-making is not based on a computational, explicable algorithm but rather on the exploitation of continuously changing data and where the quality of the decision can be evaluated only at some later point in time if at all. While all kinds of complex technologies (such as ERP-systems, telecommunication infrastructure etc.) might benefit from the concept of keeping the organization in the loop, AI in its current and prospective

stages of development represents one of the clearest challenges to be met by this concept.

With respect to keeping the organization in the loop, the paper investigates the following research questions:

- RQ1) How can organizations ensure that humans are kept in the loop within AI-based socio-technical systems and help workers to continuously develop their competences and autonomy?
- RQ2) How do organizational practices enable AI-based socio-technical systems to ensure reliable functioning—not only during implementation, but over time?
- RQ3) What does a systematic framework look like that integrates the organizational practices that can be empirically identified and takes them into account continuously in AI (re-)design and use?

To answer these questions, we identify existing approaches to organizational issues in the literature on the design of human-centered AI (Sect. 2). In an extension of existing studies, we present the methods (Sect. 3) and findings (Sect. 4) of an empirical exploration of two cases of AI implementation in the area of predictive maintenance. The case analyses show the many ways in which AI/ML-based predictive maintenance outcomes depend on organizational influences. In Sect. 5, we build on this analysis by developing two general models of how AI continuously interacts with human labor in production systems. First, we identify four different loops of how human tasks, interacting with AI and organizational practices, are interwoven. Increasing then the granularity of the analysis, we present a schema of organizational tasks, actors, and practices that were evident in the implementation and maintenance of AI systems in our cases and are likely to be characteristic generally of AI-using organizations.

2 Literature review: AI and organizations

In socio-technical systems, organizational practices and human–computer interaction are closely intertwined, shaping the roles humans play in their daily work—when interacting with technology as well as within organizational practices and processes. Therefore, we start the review with literature on human-AI interaction (2.1). First, we look at approaches that employ humans for the compensation of weaknesses in AI (2.1.1). Second, we look at aspects emphasizing the necessity of or possibility for humans to improve their capabilities by interacting with AI (2.1.2). The contrast between these two perspectives demonstrates the type of interaction modes that have to be employed in organizational measures (RQ1 and RQ2) that aim to keep workers in the loop. In the second subsection (2.2) we analyze

organizational issues related to using AI as addressed in the literature on human-AI interaction, and in the social sciences and managerial studies. We do this to get beyond a pure interaction perspective and to identify aspects to be integrated in a systematic framework (RQ3).

2.1 The design of human-AI interaction

2.1.1 The human as a gap-filler

Originally, AI design includes aspects of replacing humans with machines, and there is a still-growing literature that identifies tasks that can be or will likely be automated (Brynjolfsson and Mitchell 2017). Within these approaches, not only the relevance of AI experts but also the role of practitioners and domain experts as well as their knowledge and experience are acknowledged as necessary for improving AI. For instance, for ensuring continuous learning of anomaly detectors, Olivotti et al. (2018) propose feedback loops to integrate the knowledge and experience of machine builders, component suppliers, and machine operators. Similarly, in reference to a “hybrid intelligence system,” Kamar proposes a kind of workflow whereby an AI system compensates for its own shortcomings by asking human workers to verify or correct its results (2016). Cai et al. (2019a) also support creating possibilities for humans to interactively refine AI results. Lindvall et al. (2018) note that clinicians’ corrections could be used to improve ML. Finally, and along the same lines, Amershi et al. would require practitioners to be involved not only in the rollout and training phases of an ML system but also after it has begun normal operation (Amershi et al. 2014).

Some approaches seek to harness the strengths of humans as a means of reducing the need for their involvement later. However, the more highly developed concepts of “collaborative AI”, “responsible AI”, or “hybrid intelligence systems” see a long-term role for humans in the application of AI systems (Dellermann et al. 2019; Koch and Oulasvirta 2018; Schiff et al. 2020).

Within the context of automation, the concept of distributed strengths highlights why keeping humans in the loop might be necessary: “operational systems often employ intermediate levels of automation in which the operator sets the high-level goals and monitors the automation as it performs the low-level tasks” (Johnson et al. 2017, p. 229). Within the HCI community, this has been described as a mixed-initiate interaction strategy, where “each agent can contribute to the task what it does best,” and where “the agents who currently know best how to proceed coordinate the other agents” (Allen 1999, p. 14). These approaches focus on the cognitive load on human agents, their capabilities, and on support techniques such as “[c]onsidering the status of a user’s attention” or “employing dialog to resolve

key uncertainties” (Horvitz 1999, p. 159). Others propose allowing users to intervene systematically so that they can control AI outcomes (Schmidt and Herrmann 2017). Schmidt (2020) proposes various ways human-centered AI can keep the human in control. However, these authors do not consider the fact that humans are involved not only in human–machine loops but are also simultaneously engaged in organizational “loops” like managerial decisions, coordination, and negotiation of agreements. This organizational context surrounds human actors and affects the way they handle their tasks. Further clarification is needed to understand how human–machine and human–human (organizational) loops are socio-technically integrated.

Dellerman et al. discuss—under the assumption that the strengths of humans and AI are complementary—how humans and machines working together achieve better results than each working separately (Dellermann et al. 2019). Jarrahi describes their different strengths in decision-making, arguing that humans are better at dealing with uncertainty and equivocality (Jarrahi 2018). While AI can extend humans’ ability to handle complexity, people’s strengths lie in enabling communication and negotiation. Consequently, we assume that for every domain of AI application—such as predictive maintenance—activities and the strengths needed to accomplish them have to be analyzed in the context of organizational practices.

2.1.2 Supporting the development of human strengths and competences

Understandability and predictability of machine behavior is a key prerequisite for “keeping humans-in-the-loop” approaches and for collaborative AI. This insight has fostered an emerging field of research on explainable AI (XAI) that aims to help users understand the background of results and decisions that are proposed by ML systems (Adadi and Berrada 2018). With XAI, the user can justify, control, or improve the outcome of AI and discover new insights about AI’s possibilities and limits. We suggest that explanations of the background of decisions proposed by AI do not only help improve a single instance of task handling but contributes to an overall understanding of the functioning of AI and its potentials and limits, and thus promote the development of users’ competences when dealing with AI and the related tasks.

Collaborative explanation between humans is explicitly mentioned as a model for XAI (Mueller et al. 2019). However, our extensive review of the XAI literature yielded no indication that XAI researchers have reflected on the relevance of explanations that are directly provided by humans. How human-to-human explanation and AI system-based explanation can be integrated as a socio-technical basis for human development is still a wide-open field of research.

In this context and in contrast to the mainstream, some approaches pursue the goal of developing human strengths within collaborative AI. This includes continuous improvement by collaborative learning (Dellermann et al. 2019; Herrmann 2020). In a broader sense, collaboration is characterized by symmetry and negotiability, e.g., about the division of labor (Dillenbourg 1999). Furthermore, it includes a recognition of human work within the machine-learning workflow that allows for reciprocal adaptation between humans and ML systems (Amershi et al. 2014). Job design has to combine AI and non-AI training approaches (Wilkins 2020). Others focus on how to support learning and identify various dimensions where humans and AI can augment each other (Holstein et al. 2020). These include areas such as sensing, attention, interpretation, and decision-making. The question of how learning support is achieved is mostly answered with respect to the features of the AI system. Examples include:

- allowing individuals the option to respond to incorrect predictions of an AI system (Nguyen et al. 2018),
- giving prompts that help counter the human tendency to place too much trust in automated systems (Okamura and Yamada 2020),
- and integrating knowledge management with learning (Li and Herd 2017).

However, the organizational practices required to realize benefits from these solutions are not extensively discussed within these approaches.

2.2 AI and organization

2.2.1 Human collaboration and organization within HCI

Some research into collaborative AI or hybrid intelligence systems occasionally makes reference to organizational factors and the socio-technical approach (Dellermann et al. 2019). “A key challenge to the implementation and adoption of intelligent machines in the workplace is their integration with situated work practices and organizational processes” (Wolf and Blomberg 2019, p. 546). For example, integrating AI systems and employees is seen as an organizational task—an onboarding process that has to be planned (Cai et al. 2019b). The implementation of AI is also seen as a process of socialization (Makarius et al. 2020). While some scholars refer to team processes and collaboration without taking collaboration between humans into account, others take a broader perspective. Seeber et al. (2020) address research topics about the integration of AI into teams of humans. For instance, they suggest that humans likely need to learn to adapt to these new teammates and “organizations could facilitate this adaptation by training people in the required collaboration competences” (Seeber et al. 2020, p.

6). Others suggest that XAI has to be complemented with network-based trust: “we trust an application because somebody we know (and trust) trusts the application as well” (Andras et al. 2018, p. 79).

Schiff et al. (2020) point out that companies that intend to comply with high-level principles for responsible AI have to implement these principles as organizational practices. Rakova et al. give concrete examples such as “the importance of being able to veto an AI system” and “[b]uilding channels for communication between people” (Rakova et al. 2020, p. 17). Shneiderman (2020) emphasizes including management to achieve reliability and security (Shneiderman 2020). Others point out that specific challenges arise from employing AI in the context of inter-organizational collaboration (Campion et al. 2020).

2.2.2 AI and organization in the social sciences and in management studies

It may not be surprising that engineering disciplines tend to leave organizational factors out of their equations, but also social sciences and management studies focus strongly on the “distributed agency” between human and machine, addressing the organization often just as a static environmental condition than as a dynamic set of actors and rules (Latour 1988; Rammert 2008). Current considerations often reduce the interaction between humans and AI to a few or even single aspects, e.g., how AI decisions affect humans or the distribution of moral decision-making in the case of autonomous driving (Awad et al. 2018). In engineering and computer science, the issue of machine behavior is commonly understood as an interdisciplinary task that raises ethical questions—appeals to scholars of law, humanities, and the social sciences to weigh in are common (Rahwan et al. 2019). This big leap from looking at interactions of individual human beings with AI to philosophical questions concerning the whole society or even “humanity” is often made. Generally speaking, the sociology of media and technology, which is influenced by cultural studies, is focused on applications of AI outside of the world of work and on related social discourses. It takes a more metaphorical and philosophical view of the algorithm, and theoretical concepts of transparency and explainability are at the center of interest (c.f. Reichert et al. 2018). Overlooked are the multiple layers between these micro and macro views: the meso level of organizations, institutions, networks, communities, processes, and agreements in which (any) use of technology is (and must be) decided, realized, concretized, micro-managed, and processed as well as changed and developed on a daily basis:

- *Organization as Yet Another Context* is addressed when individuals—here again the micro-level—are connected

to a technology application or how organizations benefit from the productivity gains to be realized through new technology (Wilson and Daugherty 2018, p. 117).

- *Decisions about distribution of agency as organizational challenge:* Using the empirical case of the Volkswagen emissions scandal, Johnson and Verdicchio show that in the case of AI, the simple comparison of technology and humans must be expanded to include the question of the combination of “causal agency of artifacts and the intentional agency of humans to better describe what happens in AI as it functions in real-world contexts” (Johnson and Verdicchio 2019, p. 639). Athey et al. argue along the same lines, noting that the allocation of decision authority between humans and AI is a problem that organizations have to solve; in a taxonomy, they distinguish between four conceivable functionalities of AI within the organization: replacement AI, augmentation AI, unreliable AI, and antagonistic AI (Athey et al. 2020, pp. 7–9). However, they mention the organization as such only once in their introduction. Jarrahi emphasizes the importance of a complementary “augmented” use of AI in organizations (Jarrahi 2018). In this way, the human capacity for intuition can be supported by AI rather than replaced, as this human ability is of particular importance in organizational decision-making processes. Although these three studies start with the stated intention of analyzing organizational practices, they mostly end up as nearly organization-free considerations of single decisions by individuals in interaction with AI.
- *Only few studies discuss organization as AI enabler:* Mateescu and Elish show that the implementation of AI in organizations not only impacts work but is itself work, stressing “technology’s rough edges“, and serves only to reduce the organizational challenge to *an once-only process of integration* (Mateescu and Elish 2019, p. 51).

These over-simplified perspectives were corrected in part recently by demands that management recognizes the characteristics of AI systems that distinguish them from other technologies and thus require “continuous oversight” by the organization (Silverman 2020, p. 2).

2.2.3 AI and organization: missing links, blind spots, and some hope

There as yet hardly any reliable data on the use of AI from an organizational perspective; even quantitative data with regard to AI at the “firm level” is still a clear research lacuna (Seamans and Raj 2018). There is also a dearth of qualitative case studies, which are necessary to reveal how organizations integrate AI in practice. Thus, the well-known interrelations of individuals, tasks, organizations, and institutions continue to be mostly ignored, as are questions of power

within organizations (Cohen 2016). In fact, this is a blind spot not only in the AI debate, but one that organizational sociology also criticizes for its own discipline (Vallas and Hill 2012).

Performance legitimacy, task legitimacy, and value legitimacy are of particular importance for the successful adaptation of new technologies, they have to be established and negotiated in organizational practices (Ren 2019). Also, the relationship between management and employees is particularly relevant but rarely addressed in AI research. Kellogg et al., a noteworthy exception, describe the organization as “contested terrain,” where management implements AI to “reshape organizational control”; the authors identify six mechanisms (“6 Rs”) of algorithmic employee control: restricting and recommending, recording and rating, replacing and rewarding. They also find signs of worker resistance which authors call “algoactivism” (Kellogg et al. 2020).

Crawford und Joler show in impressive depth the diverse human and natural resources that make Amazon Echo and Alexa possible, including also the various companies linked into their complex value chain (e.g., distributors, assemblers, component manufacturers, smelters, refiners, and mine operators), but they do not address the coordination and process-design efforts that are necessary within these organizations to guarantee value creation (Crawford and Joler 2018). Ross et al. (2017) even proclaim the end of the organization in the face of AI.

Only recently have more studies been found that more strongly address the organization itself and are also increasingly focusing on the use of AI in companies. Given that even “at highly tech-oriented firms like Amazon, ambitious AI projects often fail or dramatically exceed their budgets and timelines” (Davenport 2021, p. 169). Davenport and Miller (2021) are among the few who have looked more systematically at deploying AI systems in organizations. On the one hand, they distinguish between the technology-based approach. This includes, for example, the decision on the appropriate platform (exploration support, transaction support, or automated decision platforms), but also flanking software systems for intelligent case management (CMS) or robotic process automation (RPA). On the other hand, they emphasize a people- and job-role-based approach that not only addresses employees, but also highlights the importance of, for example, customers in certain situations. Some compare, for example, the assessments of management and employees in terms of fairness, trust, and emotion (e.g., Lee 2018). But they do not address the organization itself. Others address how organizational processes are changed by AI, for example, tracing the effects of automated decisions in HR management on employment relations (Duggan et al. 2020). Jarrahi et al. (2021) analyze the impact of algorithmic management on power and social structures within organizations. Both studies impressively demonstrate the

consequences for the organization and its members, but they leave open to a large extent which organizational embedding would be the prerequisite for satisfactorily functioning and error-robust applications in this context and how they should be organized. This topic is partly addressed in a study which presents new organizational roles and practices, and adopting alternative technology design principles—although here, too, the focus is on the application for peoples’ analytics (Gal et al. 2020). Our study, on the other hand, specifically addresses the conditions of organizational embedding of AI systems for a topic that is not itself social or organizational. Precisely because PM can have consequences not only for the organization and its members, but also for production technology, product quality, delivery reliability, etc., we ask about the organizational conditions for this entire complex—and, also technical—system, not only for the impacts and consequences for the organization.

The literature review yields many clues relevant to our research questions about the significance of organizational practices, but the current literature addresses mainly the organization of human contributions to advancing AI rather than addressing the development of human competences by organizing the distribution of agency (RQ1). Also, even when the organization is seen as more than just another context, the focus lies on the initial integration of AI and not on organizing AI and human development over long periods of time (RQ2). Thus, the current literature does not provide a systematic framework of how to integrate various organizational measures and practices for maintaining control (RQ3).

3 Predictive maintenance: case description and methods

Predictive maintenance (PM) is considered to be one of the most important AI applications for Industry 4.0 (Garcia et al. 2020). It is used for process technology such as energy supply as well as for infrastructure facilities (Doleski 2020; Timofeev and Denisov 2020). A McKinsey study pointed to predictive maintenance as one of the most economically lucrative use cases for AI and ML, valuing it at 0.5–0.7 trillion USD (Chui et al. 2018, p. 21).

Maintenance and repair have always been the subject of anticipation and planning. However, in large-scale plants, for example in chemical production, in energy and waste-water infrastructure, the aim is to minimize the intervals and duration of downtimes, to avoid unexpected malfunctions as far as possible, and to procure or replace spare parts as often as necessary but as late and infrequently as possible. A distinction has long been made between “corrective maintenance” following a malfunction and “preventive maintenance.” In the latter case, the aim has always been to avoid faults and malfunctions, but at the same time “not to do it too often

from the viewpoints of reliability and cost” (Nakagawa 2006, p. 135). Simply put, the goal of maintenance is “to keep...physical assets in an existing state,” while predictive maintenance (PM) is a way “to view data and does not necessarily require a lot of equipment” (Levitt 2003, p. 91).

It has become technically possible not only to evaluate the large amount of data generated in the embedded systems of the plants in real time, but also to examine data sets generated over long periods of time for previously undiscovered correlations and to combine them with other data, such as data on shift intervals from the Production Planning System (PPS). Predictive maintenance is largely understood as an application of AI and machine learning, with organizational approaches slipping into the background in favor of data-driven, technical measures. Predictive maintenance (PM) is defined as a “technology to recognize the condition of an equipment to identify maintenance requirements to maximize its performance” (Zonta et al. 2020, p. 49).

The use of machine learning in PM is a hot topic with a number of unresolved issues, such as scalability, latency, and data security (Dalzochio et al. 2020). Recent patents see maintenance workers and their activities as part of a neural maintenance network (Hishinuma and Osaki 2020). Other strategies focus more on models of collaboration between humans and AI in a maintenance system and emphasize the importance of both sets of knowledge (Illankoon and Tretten 2020). Some approaches focus more on the reduction of working hours and thus the proportion of human labor using more ML (Khalid et al. 2020; Silva 2020). Other research address the new skill requirements for maintenance staff whenever ML is used (Windelband 2017). Furthermore, maintenance workers’ knowledge and experience is of widely underappreciated value for the successful introduction of ML-based maintenance systems (Pfeiffer 2020).

Previous studies on the application of machine learning in the field of maintenance are predominantly focused on concepts for a future application or on the prediction of possible effects on the level of work. In contrast, the integration of this new, dynamically changing technology into the complex organizational structures around all processes of maintenance play a marginal role. Thus, for advancing research from this perspective, we re-investigated two cases of PM in the automotive sector. These case studies allowed an in-depth investigation into organizational structures. The next section provides a description of the cases and the methods used for investigating them.

Our two case studies in the automotive industry were selected for the following analysis. The reason for this was that PM implementations have long been widespread in this industry due to series production with extremely large batch sizes and complex products. In addition, the industry is highly automated and has been very active in the field of Industry 4.0 since the beginning, so there is already a

lot of digital data with relevance for maintenance here. A database for AI-based PM is therefore available here and is currently already being used in the industry. The automotive industry is therefore particularly suitable for an analysis of the challenges for AI-based PM, both for production-related and economic reasons, to analyze the potential for and the limits of ML using the example of maintenance. Both case studies took place in a similar data collection period and in the same industry, but deliberately contrast in two respects: (a) in the size of the plants studied and (b) in the disciplinary access as originally, case study A was researched from the perspective and with the methods of business informatics and case study B from the perspective and with the methods of sociology. Therefore, the original research questions and approaches varied.

- *Case A: Predictive Maintenance for a Car Manufacturer.* This case, from the perspective of informatics, dealt with a PM implementation project at a German car manufacturer in a chassis manufacturing division with 1500 employees. The PM project team generated hypotheses about what symptoms should initiate warnings for a car manufacturer's chassis assembly line. Because such hypotheses are based on fluctuations and outliers in a continuous stream of sensor data, it was clear that a machine-learning approach could replace a human-controlled procedure of hypothesis-generation as long as enough data and sufficiently granular sensor data are available. Study materials included the company's own thorough documentation of the planning and implementation of the PM process. Conducted were four interviews in 2017 with individuals with implementation responsibilities in different roles: plant operator, master craftsman, systems specialist, and planner. The interviews were transcribed and content-analyzed with respect to a set of problem categories.

In this case, the plant operator was the main addressee of PM notifications, discussions and evaluations of warnings, malfunctions, and false alarms by the PM system and according reactions were object of team meetings on the shopfloor. The PM-implementation process in this case thus could be interpreted as a *bottom-up* approach. From the interviews 77 specific challenges for PM integration were derived (Herrmann 2020).

- *Case B: Predictive Maintenance within Industry 4.0.* The second case is an automotive factory that mainly produces smaller commercial vehicles. At the time of the sociological study (2016), it employed around 15,000 people. A qualitative case study was combined with quantitative surveys on the views of employees on Industry 4.0. Predictive maintenance was just

one aspect of broader issues of technology that were addressed in the case study that included 28 qualitative interviews, several workplace observations, four group discussions, and six feedback workshops (Pfeiffer et al. 2019). The qualitative interviews, each lasting about one hour, were conducted with engineers, shop floor employees, and maintenance workers. The interview guidelines were loosely structured; the main aim was to encourage the development of self-determined narratives. The survey method and the selection of the interviewees were based on approaches from the German sociology of work tradition (Pfeiffer et al. 2019). These are similar to the company-case-study method more often used in international research (Yin 2008). The interviews were transcribed in full and content-analyzed (Mayring 2007). The introduction of PM was planned primarily for the assembly and body shop areas. The aim was to standardize the control software of the industrial robots in the PM train. Thus, the PM-implementation process in this case could be interpreted as a *top-down* approach. This one decision alone—to control all robots centrally from the headquarters simultaneously—triggered the need for a high level of organizational coordination and led to numerous technical problems with sometimes considerable economic consequences.

The material of both case studies was re-examined for this analysis and thereby jointly subjected to interdisciplinary re-interpretation. Inter-disciplinarity is a challenge and is often understood as a “one-way cross-disciplinarity [...] with researchers in one field using knowledge or output from a different field without feedback in the other direction that contributes to more integrative knowledge” (Morss et al. 2021, p. 1154). In contrast, this analysis followed a “cross-disciplinary” approach. In our joint re-interpretation of the empirical material of both case studies we combined “knowledge, approaches, tools, or other core elements” from both contributing fields. This joint, multilevel, and iterative process of analysis not only confronted the research perspectives of two scientific disciplines, but also confronted the empirical material with the challenges of an AI-based approach in the field of production and maintenance.

Although the different cases varied in terms of the organizational embedding of PM integration (Case A bottom-up, Case B top-down), the joint interdisciplinary re-analysis showed that in both cases, a multitude of organizational preconditions emerge as necessary for successful and sustainably functioning implementations—organizational preconditions, however, that were also strongly underestimated by the actors in the field themselves. The following analysis systematically traces these organizational categories based on both case studies and then

discusses the generic organizational requirements for the design of human–machine systems of AI-based systems.

4 Case study findings: organizational embeddedness of AI in predictive maintenance

When analyzing the data of both case studies from the perspective of embedding ML/AI within complex business organizations, we grouped the challenges and problems found into six categories where AI/ML use within PM requires or affects specific organizational practices. Below, we describe these categories in general terms that could also be valid for any organization. In the following, we illustrate these core organizational categories for coping with AI using our empirical example of Predictive Maintenance (see also the discussion in Sect. 5 and in detail Fig. 2).

4.1 Leadership and human resources as enablers of AI/ML deployment

Leadership and human resource management (HR) are decisive factors in shaping organizational processes, supporting change, ensuring employee skill development, and enabling a learning organization—this is all the more true for embedding AI technology within organizations. While leadership and competence development tend to take place on site in the operational departments (in the context of PM this usually means production, assembly, maintenance, and IT), change management tends to be orchestrated by central HR. In implementing AI, the more technically oriented operational departments and HR administrators (who usually have little technical expertise) need to cooperate more than usual. Thus, the organizational processing and embedding of PM is anything but a self-starter.

Two organizational approaches to the challenge of PM are observable. One is to apply standard procedures of central HR and local management to the implementation and development of PM. Another is to augment this business-as-usual approach with new and PM-specific requirements, which themselves require new organizational learning processes, especially within the usually non-technically driven HR. Beyond systematic coupling of the PM implementation and adaption processes with HR tools such as change management and skills and organizational development, additional novel requirements can be seen in at least three central dimensions. Each is discussed separately in the following three sub-sections.

4.1.1 Aligning roles and people to PM-related tasks

Designing and improving PM systems, creating new learning queues, and deciding on new training data necessitates having qualified (or qualifiable) employees. This necessity increases with the inclusion of ML-based techniques for PM. Even already-implemented PM systems are in continuous need of intervening action of varying degrees of urgency and with varying serious consequences (planned production downtimes, ordering expensive spare parts, etc.). These interventions draw on qualified staff from different disciplines (mechanics, hydraulics, embedded systems, etc.) and different backgrounds of experience (with specific systems, machines, product lines, process technology, etc.). The complex question of who is allowed to decide, initiate, order, or remedy what AI-provided decision, whether alone or with others, and at what level has to be resolved in the context of a PM system time and again. It is resolved in the ongoing evolution of the PM system (see Sect. 4.6) and has to be prepared and represented organizationally.

4.1.2 Dealing with the shortcomings of AI

PM systems can make bad recommendations. Employees may also have good reasons for not following the PM system's recommendation in certain situations, e.g., for an experimental handling of assumed false positive PM notifications. These human decisions can later prove to be bad or can lead to quality losses, high costs, or even legal liability problems. How to proceed in such cases and what consequences employees will face are questions the organization must answer (see Sect. 4.2). The answers they develop must comply with law, be robust, and must not only protect employees, customers, and the company when mistakes are made, but must also initiate and accompany all subsequent learning processes. These too are also tasks HR and leadership have to address, generate, and carry out.

4.2 Decision-making about AI outcome

At the moment when the PM system issues a notification, an employee has to decide what to do about it and how to integrate the response into routine maintenance tasks. At a first glance, this problem seems mainly to be about an individual's activity in interaction with the PM system. A closer look reveals the underlying organizational issues.

4.2.1 Responsibilities, duties, and rights

First, the organization has to decide who is authorized to deal with the PM messages (see Sect. 4.3). This is usually the person responsible for the system to which the notification refers. However, this person can decide to

forward the PM notification to someone else. It is a matter of organizational decision rules to whom and under what circumstances PM notifications can or must be forwarded. Since AI plays the role of a decision-proposer, the persons addressed must have the standing and rights to oppose such a proposal; whether this is the case must be checked and determined at least once by superiors, process managers, and/or HR (see Sect. 4.1).

4.2.2 Collaborative decisions about AI-initiated tasks

In case A, a plant operator who is authorized to accept a notification is responsible for the coordination of its complete handling. This too is an example of an organizational rule. The typical procedure is that the operator will inspect the part of the plant that is potentially affected. In complex systems, the initial notification and its root cause may differ and may require a longer search and a testing process and also feedback loops with others (see 4.2.3). The next decision is whether fixes are necessary and when. Depending on the identified problem, the plant operator might decide that their knowledge or experience is insufficient for handling the notification. If so, and if there is sufficient time, the operator will discuss the PM notification at a meeting for coordinating maintenance work and identifying potential improvements of the plant's technical processes. These types of meetings are another instance when organization is involved. When a PM system is introduced, time management, the management of topics, their prioritization, and the subsequent documentation work have to be reorganized for running these meetings, and it has to be determined how far aspects such as prioritization become a matter of AI when its data-based capacities will increase over time.

4.2.3 Involving others in maintenance work

Even if the plant operator solves the problem indicated by the PM system directly, additional specialists may have to get involved. Doing the repair work is usually not a job for one individual, and the plant operator has to check the availability of the others involved, their capabilities, the availability of replacement parts, the availability of tools, etc. All in all, this coordination can possibly include complex negotiations, for example if a works council has to be involved because the repair work needs to be scheduled for a weekend or requires unplanned overtime. In such cases, it may be necessary to involve other actors from the organization who are not ordinarily involved in maintenance (see also Sect. 4.4). Again, the question is how far AI will be used to coordinate the involvement of others.

4.2.4 Continuous improvement of AI-based PM-outcome

After having inspected the plant, the operator might also realize that the PM notification is not appropriate (false positive). Or the operator might detect problems that should have but were not flagged by the PM system (false negative). In the case of AI—where its quality cannot sufficiently be tested before runtime—documentation and the initiation of quality improvement is continuously necessary. In case A, this cannot be done by the operator alone but needs the inclusion of other organizational units such as the master craftsman, specialized workshops, and the key user assigned to the novel implemented PM system. Frequent refinement of PM-results points to the need for evolving the PM-system (see Sect. 4.6).

4.2.5 Complying with organizational rules

For all successive decision-making after a PM warning and in its context, organizational rules have to be specified and conventions have to be established. The rules also have to consider the larger context, for example whether equipment is still under warranty or whether a contract for maintenance by external partners is applicable. It is a task of the organization that has implemented the PM system to provide the rules for these instances of decision-making, to align them with its strategies, and to continuously adapt the rules in accordance with changes in conditions and strategies.

Furthermore, the organization has to make sure that these rules are known and accepted by the employees dealing with PM notifications and by all employees in the organizational loops within which this decision-making is embedded. Potentially, a machine learning-based PM system could support this long-term alignment of personnel and rules.

4.3 Coordination of original tasks supported with AI

4.3.1 Workflows and tasks

If an operator decides autonomously or in communication with others that the PM messages should trigger further action, the subsequent activities have to be coordinated. One of the first things that happens is that the PM message is forwarded to others who should know about it. This is especially important if other specialists have to contribute to the maintenance work. Beyond the involvement of people, the acquisition of spare parts or special tools also has to be organized. This could involve complex organizational procedures, for example if parts have to be ordered or if the order has to be approved, the time needed until the replacement part is available also has to be factored into the coordination of the maintenance task. A long chain of coordination

processes with internal or external actors can result (cf. also Sect. 4.4). A continuously evolving PM system (see Sect. 4.6) could also offer AI-based suggestions for how the indicated problems could be dealt with and who has to be included.

4.3.2 Involving an ecology of the roles involved

Every decision regarding how an anticipated problem has to be dealt with and how the upcoming tasks are handled and assigned has to be coordinated with upstream and downstream organizational functions and/or departments because of possible consequences concerning economic questions, technical issues, questions about the stream of material, productivity or quality, etc. Further organizational units with which coordination should or can take place are, for example, the works council, quality assurance, stock keeping, procurement, and production management. Under specific conditions, if a certain machine part is affected by an irregular series of breakdowns, it might be reasonable to include the supplier of the machine concerned. The inclusion of external partners also usually involves an organizational procedure (again see also 4.4). Under the condition of AI usage, all of these organizational procedures have to set under consideration that they are triggered by more or less reliable forecasts that can become obsolete after a procedure has been started.

4.3.3 Reflection of context and interdependencies with other maintenance activities

An elaborated handling of PM notifications and the problem that generated them will include a reflection and understanding of their underlying causes (see Sect. 4.5.2) and a consideration of possible dependencies with other material and technical facts as well as relations to other ongoing or pre-scheduled maintenance work. That means that not every PM notification initiates a separate workflow, because coupling with other maintenance steps also takes place. The action that follows a PM message is split into different steps that are separately coordinated. For example, first an improvised fix of a leaking pipe might take place, and then the whole pipe is replaced a week later together with a larger unit of an assembly machine. This type of consideration and coordination requires deliberate reflection and an understanding of the reasons behind a PM notification and of the related interdependencies, and it generates expectations regarding what advanced AI might also have to consider.

4.4 Coordinating with the external world

Maintenance processes are integrated into a network of actors from outside the company, with whom constant coordination must occur. This necessity will not disappear

after PM implementation. Rather, the way the interplay with the external world is managed will change and will face new challenges. A distinction must be made between situational versus general coordination requirements. Moreover, new coordination requirements will likely only emerge after PM has been in use for a longer period of time.

4.4.1 Regulation of access from outside: when, who, and what

A basic kind of coordination, which has to be clarified both generally and in particular situations, takes place between internal maintenance and the external providers of plant and equipment. The general clarification, beyond other intentions, also includes who may view, use, and process which data for which purpose. The same data can be used for the AI-supported teleservice by the external manufacturer and/or the internal PM system. Who has access to which data must first be clarified in general but also checked and adjusted during software updates (of the embedded system as well as of the AI/PM systems).

In the event of a malfunction, which kinds of problems can be fixed internally and which ones require calling in an external service must be clarified. For this purpose, maintenance contracts must be negotiated that also deal with the question of the extent to which external actors are allowed to influence AI-based PM.

4.4.2 Various and varying external influences to be addressed

In addition to coordination with the manufacturers of the machines and systems, coordination is also necessary with other external players. For example, with the suppliers of peripheral and handling technology or with the suppliers of production-relevant consumables (lubricants, coolants, compressed air, small sealing parts, etc.). Their products have a direct or indirect influence on the question whether ML-based PM system have to be adapted, and they are sometimes also relevant in the event of a malfunction. Often the contractual basis of the respective cooperation is regulated very differently. New contractual frameworks have to be designed that consider the peculiarities of ML-based PM and its permanent necessities and possibilities for change.

The coordination with external partners is an ongoing task. It is not only necessary when implementing the PM system, but also in the case where external relations change (change of supplier, changed supplier evaluation, etc.) and, if necessary, also in case of the frequent changes in ML-based PM itself.

4.4.3 Sharing risks and benefits

In addition, organizational processes with suppliers must be set up if materials or vendor parts change—even within the specification—because this can also have an effect on the PM training-data. Another PM-specific, organizational issue to be processed is feedback within the organization as well as between the organization and its stakeholders in the case of recurring problems. The service and development departments of equipment manufacturers will need to be called in if the causes of recurring faults are known to originate in their equipment. In light of possible negative outcomes of PM-based decisions, such as production downtimes, cost risks must be anticipated and protection against legal liability (e.g., contractual penalties due to late delivery or non-compliance with just-in-time specifications) must be in place. Finally, it is conceivable that external partners will exploit PM vulnerabilities in their pricing. For example, if a supplier becomes aware of just how tightly planned and urgent the procurement of a spare part is, it may raise its price accordingly.

4.5 Dealing with changes

4.5.1 From prediction to efficient measures and sharing of resources

Many decisions within the socio-technical framework of a PM system require general organizational clarification. For example, a PM system may generate recommendations, based on technical parameters, about which spare parts should be ordered only when needed and which ones should be stored on site. Using storage space and scheduling what may then be longer production downtimes may make sense technically, but they can increase costs. General procedures and organizational decision-making structures are thus necessary for such coordination, as they require negotiating among different operational logics and paradigms. These general procedures must also be integrated into the existing organizational processes and integrated with other IT tools, for example with various processes and modules of the ERP system (e.g., supplier evaluation, quality management, or just-in-time procedures).

4.5.2 Tracing the causes of problems

PM systems based on machine learning and constantly changing learning data sets are dynamic systems, which means that completely new legal and liability challenges arise. If, for example, defective parts produced on the equipment are returned at a later date, it must be legally clarified whether the cause of the defect lay with the parts manufacturer, the equipment manufacturer, or the supplier of the PM

system. The parameters of the PM system at the time the defective part was produced may have to be reconstructed weeks later. Procedures must be developed for this purpose that regulate how system states can be continuously documented and what to do in the event that liability has to be clarified.

4.5.3 Risk–benefit trade-offs

In the long run, PM systems should not only be able to take costs into account, but the systems themselves create new cost variables that usually were not part of previous operational or strategic controlling procedures. For example, and especially in the PM implementation phase, the cost of dealing with “false positives” needs to be assessed and compared to PM-caused savings. The cost side must also be considered strategically. Depending on the business model, process technology, and products, general strategies (cf. Section 4.2) for dealing with reports must be defined. The cost and liability risks have to be integrated into a trade-off equation, and such strategic decisions must be made at higher level in the organization. They involve more than just the PM system and its shop-floor operator alone.

4.5.4 Organizing multi-level long-term changes

Decisions are necessary not only in response to PM notifications (see Sect. 4.2.2), but also for the long-term handling of the system. A decision is needed again and again as to whether the PM has to be re-trained based on new data (cf. Section 4.6). Criteria for such decisions must be negotiated and defined. Further, a procedure must be generated and mapped in IT and organizational processes for the following question: When should new learning (not) take place? For example, for technical, economic, and/or liability reasons, a kind of “freeze” for a certain period of time can be more sensible than a permanent learning process using constantly changing data. The changes may be of contradictory nature, their consequences and the related cascades of causes may be unknown, and yet they must be considered in the PM system and the organizational requirements in which it is embedded.

4.6 Evolving the PM system

Every PM notification (or failure to appropriately notify) requires two human responses: whether the plant has to be maintained and whether the PM system itself needs to be adjusted (see Sect. 4.2). In this sense, PM necessitates two types of maintenance, and also the second type of maintenance is surrounded by organizational roles and procedures. It is useful to differentiate between small and large maintenance cycles. A short-cycle example is sorting out

outlier data from PM input or modifying trigger thresholds for notifications. A long-cycle example would be reactivating the developers (see 4.4.2) to employ new training data for overcoming a larger set of documented failures.

4.6.1 Organizational procedures of improvements

The organization has to decide who is in charge of continuous quality assurance for PM and who is responsible for and has the permission to modify the PM system. A typical workflow might be (as in Case A) that the most experienced plant operators can propose and test modifications, and eventually submit them to other individuals for approval or rejection, for example a key user higher up in the hierarchy. Similarly, long maintenance cycles that involve developers must be coordinated within the organization. To expand on the example noted above, the preparation and handover of new training data, perhaps due to changes in the context of the PM system such as the addition of new sensors, and the decision about when or how often the underlying models of the PM-system need to be re-trained are decisions that affect many parts of the organization. Another type of decision that needs approval is to ignore a PM notification to test whether it is indeed appropriate. Such an experiment can in the worst case cause a large-scale shutdown. Even if the operators are standing ready to manage this emergency, additional costs can be expected, and the organization has to develop a culture that accepts these potential losses and develop procedures to estimate them.

4.6.2 Organizational culture of reflexive PM use

It seems to be a generally accepted requirement that the output of a PM system has to be assessed critically, especially with regard to benefits versus costs. However, such critical reflection must first become part of the organization's value system and culture. The responsibility for risks associated with experimenting with and modifying the PM system has to be distributed among multiple stakeholders. The organization must guarantee that no single employee bears the responsibility for failure as long as their actions remained inside a clearly defined corridor determined by the organization. Moreover, promoting critical reflection requires that employees know that their input has a chance of resulting in concrete decisions and actions relevant to the functioning of the PM. Thus, a culture of critical reflection must emerge as an integral part of organizational procedures in the context of quality assurance, continuous improvement, and organizational learning.

4.6.3 Explainability and accountability as organizational challenge

Explainability, as a standard for human-AI interaction, is often thought of as a technical feature inherent to the machine or as a product of machine learning. The discussion of explainability puts great emphasis also on eliminating discriminating biases, e.g., in ML-powered HR tools (Köchling and Wehner 2020). However, explainability is just another technical feature, not a quality of ML alone. The outputs that enable explainability can create more explanations of PM notifications than the organization requires or can even process.

The possibility of analyzing and experimenting is a core issue for enabling explanations. The ability to ask a “what if?” question is a promising approach for helping users to understand what effect slight variations of a situation would have on system output. A basis for exploring the effects of variations is a collection of examples. Collecting examples—as a kind of a knowledge-management environment—is an organizational task involving a variety of people, procedures, responsibilities, and decisions.

5 Discussion: keeping the organization in the loop

What do the empirical observations of PM in our case studies, as described in Sect. 4, tell us generally about human–machine interaction in dynamic organizations trying to implement any kind of AI/ML? The very broadest lessons learned from the preceding analysis can be summarized in the following four points.

- a) The study of human–computer interaction has to include the organizational context.
- b) Interactions in the organization should be investigated continuously, not just during AI rollout.
- c) Keeping humans in the loop depends not so much on individual workers but mainly on organizational practices in which task handling is embedded.
- d) To realize benefits from the complementary strengths of humans and AI and from fostering their coevolutionary improvement, it is not sufficient to just plug in AI technology. By contrast, it is necessary to also develop new organizational practices including organizational learning and improvement of human collaboration.

Against this backdrop, it seems clear that “keeping the human in the loop” in any organization that uses AI involves a variety of interactions on many levels, both between humans and machines and among humans in different organizational roles. In the following discussion, we offer

two ways—at different levels of abstraction—to understand these interactions. First of all, we argue in the following section that the “loop” of human–machine interaction is actually four different loops (subsection 5.1). In the subsequent Sect. (5.2), we bore down to an increased level of granularity in a schematic representation of how AI-relevant organizational tasks, actors, and practices are interrelated.

5.1 Not one but four organizational loops

To understand the complex organizational implications of “keeping the human in the loop” we have to go beyond the quite limited addendum of just “keeping management in the loop.” Rules of coordination and the assignment of roles and tasks can only be developed and continuously evolved if all types of relevant actors participate interactively and reiteratively. Our empirical findings confirm that organizational rules for dealing with AI and its outputs, as an outcome of management decisions, are not automatically internalized as standard practice. Rules are practiced as a result of people’s constantly changing behavior and their reflection; as rules are adopted, they are thus constantly undergoing a process of adaptation (Mark 2002). Consequently, managerial coordination and organizational procedures are dynamic and “loopy.” These dynamics also apply to the socio-technical relationship between organizational practices and human–computer interaction with the AI system (Eason 1989).

Against this backdrop and in accordance with our definition of “keeping the organization in the loop” (see Sect. 1) we argue that ensuring regulation, control, and supervision of AI-based processes needs four feedback cycles. These four loops, derived from our empirical findings, refer to task handling that:

- 1) directly deals with AI (UA-tasks, loop 1) using and assessing the outcome of AI and includes steps of refining the outcome or experimenting with it,
- 2) directly deals with AI (CA-tasks, loop 2) by contributing to the evolution of the AI system based on its customization and improvement,
- 3) exist before AI usage as those original tasks (O-tasks, loop 3) that are now carried out using the AI outcome as an information basis, such as replacing a part that was the subject of a PM notification
- 4) deal with those changes and contextual factors (C-tasks, loop 4) that influence the usage and evolution of AI.

If we want to understand—whether for analytical reasons or to achieve good design—why technologies such as AI are more successful in one situation than another, then we need to understand not only the impact of user behavior, but also

the impact of organizational practices and their interaction within the four loops mentioned above.

Table 1 describes these four loops both generically and concretely in our empirical example of PM. In the last column of the table, we also add the use of AI in medical image analysis (Currie et al. 2019). This is deliberately an initial, tentative approach and merely illustrates that the abstract representation of the four loops is suitable not only for the application field of maintenance—even if it was developed from it—but also for the use of AI in organizations in general.

With respect to Fig. 1, we can spell out some central aspects of taking typical AI characteristics into consideration:

- (1) Every single instance of using AI is also a source for deciding whether and how an AI improvement should take place. This is particularly true for ML-based AI that draws on underlying training data and where every usage of AI produces new data that is potentially relevant for (re-)training. Thus, loop 1 and 2 are closely integrated.
- (2) Since AI is not based on conventional algorithms that could be tested with respect to complete knowledge about its internal structure, the quality of AI outcome has to be checked against its value for repeatedly handling the series of original tasks (maintaining of machinery in the case of PM). Thus, original task handling immediately feeds back to the question of potential AI adaptation as shown by the overlap of loop 3 with the intertwined loops 1 and 2.
- (3) Not only AI usage produces potential training data for ML but also its context. The ways in which contextual factors can trigger technical change of AI is very different compared to conventional IT solutions where re-engineering work by human experts is crucial. Contextually produced data can permanently be used for new training cycles and it is more about organizational than technical questions about whether, when, and with which data the re-training will take place.

Figure 1 gives an overview of the structure of a systematic framework that integrates organizational practices that help keep humans in the loop (RQ3). The next section outlines the interplay between these practices within the framework.

5.2 Schematic representation of organizational aspects of AI

In the following, we derive detailed crossover influences from the empirical observations (Sect. 4), describe them as interacting organizational practices, and thus provide a deeper understanding to the scheme of the four loops. We

Table 1 Four loops of organizational practices enabling AI with examples in maintenance and medicine

Loop #	Description	Typical activities	PM example	Medical image analysis
1) UA-Tasks	Using and assessing AI output	Inspecting the results of ML output; seeking explanations	An operator becomes aware of a notification that a robot's pipe will break in the near future; inspects similar notifications with comparable pipes	Radiologists inspect images of a lung where potential tumors are indicated and apply another ML model to the image and mark the differences
2) CA-Tasks	Customizing the AI system	Entering corrected data; extending a data set for re-training the ML model	Enters data that not the pipe but the engine that generates the air pressure was the problem	Enters a rule under which condition which the ML model should be applied
3) O-Tasks	Original tasks to be supported with AI	Planning measures based on the ML results; conducting the measures	The engine will be replaced on a Saturday	The images with the marked, suspicious spots are used at a tumor conference; decisions about surgery are made
4) C-Tasks	Contextual changes and factors in relation to AI	Checking whether a retraining of the ML model is needed; modifying contracts with external partners	New sensors to monitor the robots' movements are added; the supplier offers to provide aggregated data of the sensors' output	The hospital provides a new MRT machine that generates images with a much higher resolution

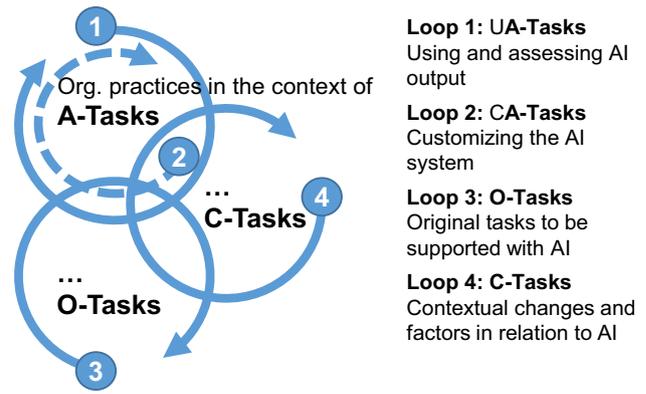


Fig. 1 Four Loops of Organizational Practices Enabling AI Implementation

describe a framework that can provide orientation both for the scientific discussion in relation to and beyond the literature analyzed above and for the practical design and implementation of AI applications. The schema presented in Fig. 2 shows how each of the six categories identified and discussed in Sect. 4 mutually interact. These categories are represented as boxes in Fig. 2, numbered 4.1 to 4.6. The various organizational roles involved are represented as ovals in Fig. 2, with a representation (arrows) of their responsibilities and the influences they are exposed to. Four specific types of interacting organizational practices (in accordance with the four loops in Fig. 1) are shown in Fig. 2 as bi-directional arrows numbered 1 through 4 that link the activities of the six categories.

5.2.1 Interplay 1: between AI-supported tasks and the preparation and coordination of AI usage.

How the workforce deals with the outcome of an AI system is not only a matter of HCI but of complex, collaborative decisions (Fig. 2, 4.2) that need ongoing coordination (Fig. 2, 4.3). In interaction (Fig. 2, 1b) with decision-making, this coordination covers the ongoing specification of duties and rights of roles and workflows that have to be aligned with accompanying processes. The coordination of AI-supported original tasks (4.3) interacts (1a) with leadership and HR (4.1) to ensure—either by centralized or on-site activities—that the relevant competences and people are assigned to the handling of the AI application. This interaction (1a) includes clarifying who is authorized to decide what (Athey et al. 2020). It therefore necessarily encounters power relationships in organizations is mostly neglected in AI-related discourses (Ren 2019; Vallas and Hill 2012). To develop an interaction between 4.1 and 4.2, a complementary combination of human and AI strengths is to be pursued (Dellermann et al. 2019; Jarrahi 2018). Organizational practices (1a) have

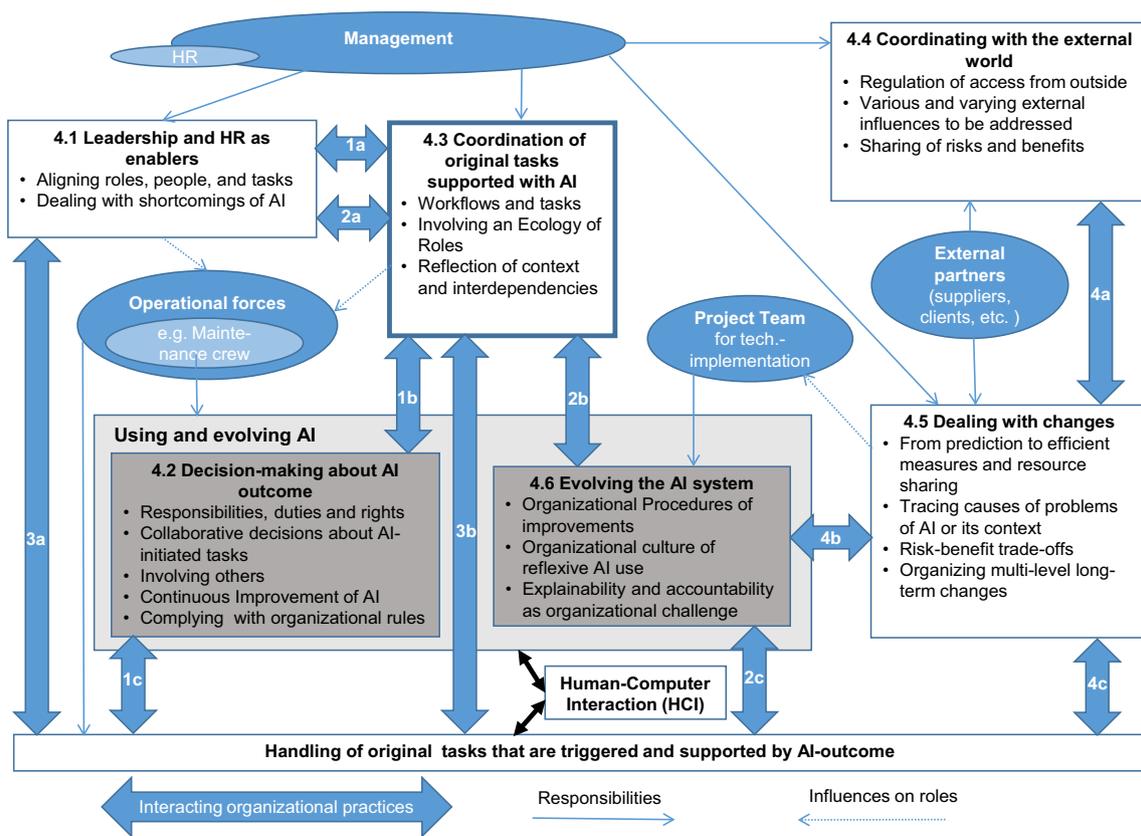


Fig. 2 Organizational embeddedness of AI

to provide the channels and procedures that allow negotiation and communication to take place as both can still be seen as superior human strengths (Jarrahi 2018).

Most important, the interplay (1c) between handling an AI outcome and the carrying out of original tasks to be supported or triggered by this outcome has to be coordinated. As shown in Fig. 1, we differentiate between A-Tasks (AI-related tasks) and O-Tasks (original tasks). AI is usually implemented to make O-Tasks more efficient and productive. But we go beyond these foci in the literature (Crawford and Joler 2018; Ren 2019). Our empirical findings (4.1, 4.2, and 4.3) reveal that these efficiency gains can only be achieved by appropriately aligned organizational practices. The shortcomings of AI become only completely apparent in relation to the O-Task, in our cases the maintaining of machinery. To overcome shortcomings, one has to go beyond approaches by which humans are asked for help only if the algorithm detects its own insufficiency as proposed by Kamar (2016). Rather, a large set of organizational measures is needed to deal with AI-related shortcomings and to engage in quality assurance and continuous improvement. These include collaboration among relevant decision makers and knowledge-holders. The training and socialization for this kind of AI-related collaboration cannot be reduced to interaction

with a “collaborative” AI (Makarius et al. 2020; Seeber et al. 2020).

5.2.2 Interplay 2: between the tasks to be supported and the preparation and coordination of AI customization

The continuous evolution of the AI System (4.6) is also a joint effort between management and operations. Also involved is the project team in charge of AI implementation or the technical infrastructure to which the AI system is connected. We consider the customization of the AI system (CA-tasks) as an ongoing challenge. This customization or configuration takes place not only during the initial design phase or AI implementation. It is an ongoing endeavor, continuously triggered by the need to deal with changes (4.5) and by the interplay (2c) with the fulfillment of O-Tasks as an indicator of the quality of AI outcome. However, it is not only a matter of an individual’s interaction with the AI system to veto an AI outcome or to intervene into an AI process (Rakova et al. 2020; Schmidt and Herrmann 2017). These individual interactions have to be supported (2b) by appropriate coordination with interrelated HR measures and leadership (4.1, 2a). Allowing for minimally intrusive ways

of refining AI results is not mainly a question of HCI design (Cai et al. 2019a). It is a challenge for the entire organization. Whenever the key actors of an organization negotiate who is authorized (4.1 & 4.3) to modify the AI system and to what extent modification is to be undertaken, dimensions of power and control are on the table (Kellogg et al. 2020; Ren 2019; Vallas and Hill 2012).

Explainability is a prerequisite of AI customization, but an explanation is not only the result of algorithmic support (Adadi and Berrada 2018; Mueller et al. 2019). Explainability is also the result of and embedded in communicative negotiation and reflection, or exploration and experiment in the interplay (2c) with O-Tasks. Experiments and interventions can be risky and have to be authorized by the organization (2b). Avoiding misplaced trust in AI results is not essentially a technical challenge (Okamura and Yamada 2020). Trust in AI is also rather a question of collaboratively negotiated assessment of the performance of an AI system as a consequence of organizational practices that eventually lead to networked trust (Andras et al. 2018). Continuously interwoven explanation of AI behavior and exploratory adaptation can improve both AI and human learning (Holstein et al. 2020). For this to occur, it must be promoted by the interplay (2a) between coordination and HR. This promotion has to acknowledge the value of developing human competence in dealing with AI and seek to protect the experiential property of the workers that is used and developed during decision-making with (4.2) and customization of AI (4.6).

5.2.3 Interplay 3: between AI-related management activities and the original tasks

This interplay supposes management and workers have to decide which subset of a larger set of original tasks will be supported by AI. Including AI into the realm of original tasks requires changes of the rights and duties of roles to be included (3a) and of the workflows to be coordinated (3b). These changes affect organizational practices and are a continuous outcome of other organizational practices that deal with decisions on the distribution of agency between human and AI (Athey et al. 2020). Specific human strengths have to be taken into account in these decisions—as described for instance by Jarrahi (2018). In the realm of the O-tasks, extending the level of AI support simultaneously implies shifting control from management to technology and a redistribution of power because control over repairs is no longer in the hands of management alone but is also relegated in part to PM-driven notifications. An AI result may not be considered when performing O-tasks, even if it is potentially relevant—this is a broader requirement than the approach of implementing the technical ability to veto as proposed by Rakova et al. (2020). The literature makes note of a requirement to evaluate the performance legitimacy of employing

AI (Ren 2019). But again, this evaluation has to be organized with respect to the effects on O-tasks. Furthermore, the desired increase in competence has to be measured in terms of O-tasks whenever AI is coupled with training for the human workforce (Holstein et al. 2020).

5.2.4 Interplay 4: between the drivers of dealing with changes and applying or evolving AI

As mentioned above, the evolution of the AI system is also influenced by contextual changes. Handling these influences requires a continuous monitoring of and coordination with the external world (4.4, 4a). Some changes result from external factors (4a) that include the societal context (Rahwan 2018), others usually include aspects of inter-organizational collaboration (Campion et al. 2020) with an ecology of external partners. We call all those tasks that deal with contextual factors and changes “C-Tasks.” The influence of C-tasks is often neglected by scholars that project an AI evolution that gradually reduces the need for human involvement (Brynjolfsson and Mitchell 2017). Other changes are triggered by the interplay of AI and O-Tasks (1c, 4c). Understanding and reflecting these dynamics and diagnosing the causes of problems in the performance of the O-Tasks (4.5, tracing causes) may lead to new tasks. For example, new strategies may be necessary to derive the most efficient and economical decisions from a series of PM warnings about the same machinery parts. Necessary changes can be supported by the internal project team, but external partners can help in understanding the limitations and opportunities of the AI application as well as the potential for further evolution (4b).

The continuous influence of external factors and their variation are the strongest support for our argument that specifying organizational measures and implementing organizational practices as described above are not a one-off necessity when an AI application is introduced but are rather ongoing throughout the whole life cycle of AI. While the loop between AI-related decision-making and O-Tasks might achieve stability with a minimum of human support, C-Tasks are continuous and will constantly trigger a need for decisions about revising the AI system and related organizational practices. Therefore, the processes of socialization into procedures of interacting with AI applications cannot be limited to onboarding (Cai et al. 2019b; Makarius et al. 2020). Instead these have to instill organizational readiness for ongoing change and learning.

Based on Fig. 2, the essence of our answers to the research questions can be framed as follows:

- RQ1: Crucial to the organizational support of keeping the human in the loop is the close integration of *decision-making about AI outcome* (4.5) and *evolving the AI sys-*

tem (4.6) through *coordination measures* (4.3) that are constantly triggered by reflecting the appropriateness of AI for the original tasks.

- RQ2: Continuous ensuring of the functioning of AI-based socio-technical systems requires dealing with the *external world* (4.4) and with *changes* (4.5) to understand how the evolution of AI is *coordinated* (4.3) with respect to the contextually triggered dynamics of original tasks.
- RQ3: A systematic framework of keeping the organization in the loop when applying AI has to acknowledge the interplay of organizational practices instead of regarding them as isolated measures. This interplay has to be adjusted in a way that keeps the four loops displayed in Fig. 1 synchronized.

6 Conclusion

This study addresses the human-centered AI-design research that emphasizes the relevance of constantly *keeping the human in the loop*. Based on sociological theory and two empirical cases in the field of predictive maintenance, the contribution of our study is to broaden human-centered AI-design research by *keeping the organization in the loop* in analyses and design of AI use. A typical phenomenon to be accounted for in socio-technical design is that organizational practices and technology continuously co-evolve and have to be continuously aligned. This applies particularly to the realm of AI and machine learning, too, where organizational practices have to accompany the continuing need and potential for improving AI based on new data that emerge by task-handling and contextual changes.

In this study, we identified four intertwined loops of interacting organizational practices of collaborative task handling and compiled a detailed schematic of tasks, actors, and interacting organizational practices that we expect to find in most AI-using organizations. These models and ways of thinking about the influence of organizations on AI lead to several topics for a future research agenda.

Further research could focus on the co-evolution not only of human individuals and AI algorithms but also of organizational practices. This includes design-oriented implications like these:

- AI-based decision-making could be designed to include the support for organizational practices, e.g., by integrating features for knowledge management (Li and Herd 2017). This seems to be the case for communicative coordination of workflows with assigned roles, tasks, and rights. For example, explaining the appropriateness of an AI-based proposal for a decision might involve those human experts who are well informed about how an ML model has been trained.

- Continuous customization and improvement of ML models could be aligned with such types of technical support and organizational practices that contribute to the development of human competencies and experience and aim to value and protect experiential property.

Organizational and management theory is currently debating whether and how AI/ML could be used to test and prune medium-scale or even large-scale organizational theories (c.f. Leavitt et al. 2020; critically Lindebaum and Ashraf 2021). Algorithms are seen as “supercarriers of formal rationality for organizational decision-making” (Lindebaum et al. 2020, p. 255). Still, practical implications of how to process the various loops of AI/ML in real organizations seem surprisingly less relevant for organizational sociology and management theory. With respect to advancing theory within these strands of organizational theory, the following consequences can be derived from our analysis.

- For the academic perspective in social sciences two aspects are of special relevance. First, in order to adequately explore the complex relationships between work process level, AI system, and organizational practices, it is necessary to work across the often-rigid divisions of labor between sociology of work, sociology of organization, and science and technology studies (STS) in the social sciences. Secondly, social science should open itself more to participative methodology, bringing in AI experts involved at all organizational levels into the conceptualization of research projects and the transfer of knowledge.
- Even practitioners in leadership positions within organizations tend to involve the organization and its experts from the various organizational contexts too little or too late when implementing AI systems. Our empirical and conceptual elaborations may also make evident change processes in practice: AI implementations in organizations are not a case for purely technical AI experts, but a complex, ongoing organizational process—from the very beginning and beyond.

Funding Open Access funding enabled and organized by Projekt DEAL. Some of the conceptual work for this article was made possible by the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the priority programm 2267 “Digitalisation of Working Worlds: Conceptualizing and capturing a Systemic Transformation” (project number 442171541). Case study A was funded by Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (MKW NRW), Germany.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6:52138–52160
- Allen JE (1999) Mixed-initiative interaction. *IEEE Intell Syst* 14(5):14–23. <https://doi.org/10.1109/5254.796083>
- Amershi S, Cakmak M, Knox WB, Kulesza T (2014) Power to the people: the role of humans in interactive machine learning. *AI Mag* 35(4):105–120
- Andras P, Esterle L, Guckert M, Han TA, Lewis PR, Milanovic K, Payne T, Perret C, Pitt J, Powers ST, Urquhart N, Wells S (2018) Trusting intelligent machines: deepening trust within socio-technical systems. *IEEE Technol Soc Mag* 37(4):76–83. <https://doi.org/10.1109/MTS.2018.2876107>
- Athey SC, Bryan KA, Gans JS (2020) The Allocation of Decision Authority to Human and Artificial Intelligence (NBER Working Paper No. 26673). National Bureau of Economic Research. <https://doi.org/10.3386/w26673>
- Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon J-F, Rahwan L (2018) The Moral Machine experiment. *Nature*. <https://doi.org/10.1038/s41586-018-0637-6>
- Baxter G, Sommerville I (2011) Socio-technical systems: From design methods to systems engineering. *Interact Comput* 23(1):4–17
- Brödner P (2018) “Super-intelligent” machine: Technological exuberance or the road to subjection. *AI & Soc* 33(3):335–346
- Brynjolfsson E, Mitchell T (2017) What can machine learning do? Workforce implications. *Science* 358(6370):1530–1534
- Cai CJ, Reif E, Hegde N, Hipp J, Kim B, Smilkov D, Wattenberg M, Viegas F, Corrado GS, Stumpe MC et al (2019a) Human-centered tools for coping with imperfect algorithms during medical decision-making. In: *Proceedings of the 2019a CHI Conference on Human Factors in Computing Systems*, 1–14
- Cai CJ, Winter S, Steiner D, Wilcox L, Terry M (2019b) “Hello AI”: uncovering the onboarding needs of medical practitioners for Human-AI Collaborative Decision-Making. In: *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–24. <https://doi.org/10.1145/3359206>
- Campion A, Gasco-Hernandez M, Jankin Mikhaylov S, Esteve M (2020) Overcoming the challenges of collaboratively adopting artificial intelligence in the public sector. *Soc Sci Comput Rev*. <https://doi.org/10.1177/0894439320979953>
- Chui M, Manyika J, Miremadi M, Henke N, Chung R, Nel P, Malhorta S (2018) Notes from the AI Frontier. Insights from Hundreds of Use Cases. Discussionpaper. McKinsey Global Institute. https://www.mckinsey.de/~media/McKinsey/Locations/Europe%20and%20Middle%20East/Deutschland/News/Presse/2018/2018-04-20/mgi_notes-from-ai-frontier_discussion-paper2.ashx
- Cohen LE (2016) Jobs as Gordian knots: a new perspective linking individuals, tasks, organizations, and institutions. In: Cohen LE, Burton MD, Lounsbury M (eds) *The structuring of work in organizations*. Emerald, pp 25–59
- Crawford K, Joler V (2018) Anatomy of an AI system. The Amazon Echo as an anatomical map of human labor, data and planetary resources. AI Now Institute. <https://anatomyof.ai>
- Currie G, Hawk KE, Rohren E, Vial A, Klein R (2019) Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci* 50(4):477–487. <https://doi.org/10.1016/j.jmir.2019.09.005>
- Dalzochio J, Kunst R, Pignaton E, Binotto A, Sanyal S, Favilla J, Barbosa J (2020) Machine learning and reasoning for predictive maintenance in Industry 4.0: current status and challenges. *Comput Industry* 123:103298. <https://doi.org/10.1016/j.compi.2020.103298>
- Davenport TH (2021) Enterprise adoption and management of artificial intelligence (SSRN Scholarly Paper ID 3916451). Social Science Research Network. <https://papers.ssrn.com/abstract=3916451>
- Davenport T, Miller S (2021) Working with smart machines: insights on the future of work. *Asian Manage Insights* 8(1):18–25
- Dellermann D, Calma A, Lipusch N, Weber T, Weigel S, Ebel PA (2019) The future of human-ai collaboration: a taxonomy of design knowledge for hybrid intelligence systems. *HICSS*.
- Dillenbourg P (1999) What do you mean by collaborative learning? In: Dillenbourg P (ed) *Collaborative learning*. Elsevier, pp 1–19
- Doleski OD (2020) Energieversorgungsunternehmen neu denken: Utility 4.0. In O. D. Doleski (Ed), *Realisierung Utility 4.0 Band 1: Praxis der digitalen Energiewirtschaft von den Grundlagen bis zur Verteilung im Smart Grid* (pp. 3–62). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25332-5_1
- Duggan J, Sherman U, Carbery R, McDonnell A (2020) Algorithmic management and app-work in the gig economy: a research agenda for employment relations and HRM. *Hum Resour Manag J* 30(1):114–132. <https://doi.org/10.1111/1748-8583.12258>
- Eason KD (1989) *Information technology and organisational change*. Taylor & Francis
- Fischer G, Herrmann T (2015) *Meta-design: transforming and enriching the design and use of socio-technical systems. Designing socially embedded technologies in the real-world*. Springer, London, pp 79–109
- Gal U, Jensen TB, Stein M-K (2020) Breaking the vicious cycle of algorithmic management: a virtue ethics approach to people analytics. *Inf Organ* 30(2):100301. <https://doi.org/10.1016/j.infoandorg.2020.100301>
- García E, Costa A, Palanca J, Giret A, Julian V, Botti V (2020) Requirements for an intelligent maintenance system for industry 4.0. In: Borangiu T, Trentesaux D, Leitão P, Giret Boggino A, Botti V (eds) *Service oriented, holonic and multi-agent manufacturing systems for industry of the future*. Springer International Publishing, pp 340–351. https://doi.org/10.1007/978-3-030-27477-1_26
- Herrmann T (2020) Socio-technical design of hybrid intelligence systems—the case of predictive maintenance. In: *International Conference on Human-Computer Interaction*, 298–309
- Herrmann T, Jahnke I, Nolte A (2021) A problem-based approach to the advancement of heuristics for socio-technical evaluation. *Behav Inf Technol*. <https://doi.org/10.1080/0144929X.2021.1972157>
- Hishinuma T, Osaki K (2020) Maintenance work support system and maintenance work support method (United States Patent No. US20200151450A1). <https://patents.google.com/patent/US20200151450A1/en>
- Holstein K, Aleven V, Rummel N (2020) A conceptual framework for human–AI hybrid adaptivity in education. In: Bittencourt II, Cukurova M, Muldner K, Luckin R, Millán E (eds) *Artificial*

- intelligence in education, vol 12163. Springer International Publishing, pp 240–254
- Horvitz E (1999) Principles of mixed-initiative user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 159–166
- Illankoon P, Tretten P (2020) Collaborating AI and human experts in the maintenance domain. *AI & Soc*. <https://doi.org/10.1007/s00146-020-01076-x>
- Jarrahi MH (2018) Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Bus Horiz* 61(4):577–586
- Jarrahi MH, Newlands G, Lee MK, Wolf CT, Kinder E, Sutherland W (2021) Algorithmic management in a work context. *Big Data Soc* 8(2):20539517211020332. <https://doi.org/10.1177/20539517211020332>
- Johnson DG, Verdicchio M (2019) AI, agency and responsibility: the VW fraud case and beyond. *AI & Soc* 34(3):639–647. <https://doi.org/10.1007/s00146-017-0781-9>
- Johnson AW, Duda KR, Sheridan TB, Oman CM (2017) A closed-loop model of operator visual attention, situation awareness, and performance across automation mode transitions. *Hum Factors* 59(2):229–241. <https://doi.org/10.1177/0018720816665759>
- Kamar E (2016) Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence. *IJCAI*, 4070–4073
- Kellogg KC, Valentine MA, Christin A (2020) Algorithms at work: the new contested terrain of control. *Acad Manag Ann* 14(1):366–410. <https://doi.org/10.5465/annals.2018.0174>
- Khalid W, Albrechtsen SH, Sigsgaard KV, Mortensen NH, Hansen KB, Soleymani I (2020) Predicting maintenance work hours in maintenance planning. *J Qual Maintenance Eng*. <https://doi.org/10.1108/JQME-06-2019-0058>
- Kirsch C, Troxler P, Ulich E (1995) Integration of people, technology and organization: The European approach. *Advances in human factors/ergonomics*, vol 20. Elsevier, pp 957–961
- Koch J, Oulasvirta A (2018) Group cognition and collaborative AI. *Human and machine learning*. Springer, pp 293–312
- Köchling A, Wehner MC (2020) Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Bus Res* 13(3):795–848. <https://doi.org/10.1007/s40685-020-00134-w>
- Latour B (1988) Mixing humans with non-humans: sociology of a door-closer. *Soc Probl* 35(3):298–310
- Leavitt K, Schabram K, Hariharan P, Barnes CM (2020) Ghost in the machine: on organizational theory in the age of machine learning. *Acad Manag Rev*. <https://doi.org/10.5465/amr.2019.0247>
- Lee MK (2018) Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management. *Big Data Soc* 5(1):2053951718756684. <https://doi.org/10.1177/2053951718756684>
- Levitt J (2003) Complete guide to preventive and predictive maintenance. Industrial Press Inc
- Li J, Herd AM (2017) Shifting practices in digital workplace learning: an integrated approach to learning, knowledge management, and knowledge sharing. *Hum Resour Dev Int* 20(3):185–193. <https://doi.org/10.1080/13678868.2017.1308460>
- Lindebaum D, Vesa M, den Hond F (2020) Insights from “the machine stops” to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Acad Manag Rev* 45(1):247–263. <https://doi.org/10.5465/amr.2018.0181>
- Lindebaum D, Ashraf M (2021) The ghost in the machine, or the ghost in organizational theory? A complementary view on the use of machine learning. *Acad Manage Rev*. <https://doi.org/10.5465/amr.2021.0036>
- Lindvall M, Molin J, Löwgren J (2018) From machine learning to machine teaching: the importance of UX. *Interactions* 25(6):52–57
- Makarius EE, Mukherjee D, Fox JD, Fox AK (2020) Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization. *J Bus Res* 120:262–273. <https://doi.org/10.1016/j.jbusres.2020.07.045>
- Mark G (2002) Conventions and commitments in distributed CSCW groups. *Computer Supported Cooperative Work (CSCW)* 11(3–4):349–387. <https://doi.org/10.1023/A:1021289427473>
- Mateescu A, Elish MC (2019). AI in Context. The Labor of Integrating New Technologies. *Data & Society*. https://datasociety.net/wp-content/uploads/2019/01/DataandSociety_AInContext.pdf
- Mayring P (2007) *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz
- Morss RE, Lazrus H, Demuth JL (2021) The “inter” within interdisciplinary research: strategies for building integration across fields. *Risk Anal* 41(7):1152–1161. <https://doi.org/10.1111/risa.13246>
- Mueller ST, Hoffman RR, Clancey W, Emrey A (2019) Explanation in Human-AI systems: a literature meta-review synopsis of key ideas and publications and bibliography for explainable AI, p 204
- Mumford E (2006) The story of socio-technical design: Reflections on its successes, failures and potential. *Inf Syst J* 16(4):317–342
- Nakagawa T (2006) *Maintenance theory of reliability*. Springer Science & Business Media
- Nguyen AT, Kharosekar A, Krishnan S, Krishnan S, Tate E, Wallace BC, Lease M (2018) Believe it or not: designing a human-AI partnership for mixed-initiative fact-checking. In: Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 189–199. <https://doi.org/10.1145/3242587.3242666>
- Okamura K, Yamada S (2020) Adaptive trust calibration for human-AI collaboration. *PLoS ONE* 15(2):e0229132. <https://doi.org/10.1371/journal.pone.0229132>
- Olivotti D, Passlick J, Axjonow A, Eilers D, Breitner MH (2018) Combining machine learning and domain experience: a hybrid-learning monitor approach for industrial machines. In: International Conference on Exploring Service Science, pp 261–273
- Pfeiffer S (2020) Kontext und KI: Zum Potenzial der Beschäftigten für Künstliche Intelligenz und Machine-Learning. *HMD - Praxis Der Wirtschaftsinformatik* 57:465–479. <https://doi.org/10.1365/s40702-020-00609-8>
- Pfeiffer S, Lee H, Held M (2019) Doing Industry 4.0—participatory design on the shop floor in the view of engineering employees. *Cuadernos De Relaciones Laborales* 37(2):293–311
- Rahwan I (2018) Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf Technol* 20(1):5–14
- Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon J-F, Breazeal C, Crandall JW, Christakis NA, Couzin ID, Jackson MO, Jennings NR, Kamar E, Kloumann IM, Larochelle H, Lazer D, McElreath R, Mislove A, Parkes DC, Pentland ‘Sandy’ A, Wellman M (2019) Machine behaviour. *Nature* 568(7753):477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- Rakova B, Yang J, Cramer H, Chowdhury R (2021) Where responsible AI meets reality: practitioner perspectives on enablers for shifting organizational practices. *Proc ACM Hum Comput Interact* 5(CSCW1):1–23. <https://doi.org/10.1145/3449081>
- Rammert W (2008) Where the action is: distributed agency between humans, machines, and programs (No. 4–2008; TUTS - WorkingPapers). TU Berlin. <https://nbn-resolving.org/urn:nbn:de:0168-ssaoar-12331>
- Reichert R, Fuchs M, Abend P, Richterich A, Wenz K (eds) (2018) *Rethinking AI: neural networks, biometrics and the new artificial intelligence*, vol 1. Transcript
- Ren M (2019) Why technology adoption succeeds or fails: An exploration from the perspective of intra-organizational legitimacy. *J Chin Sociol* 6(1):21. <https://doi.org/10.1186/s40711-019-0109-x>

- Ross PK, Ressia S, Sander EJ (2017) Data driven management, artificial intelligence, and automation. In: Ross PK, Ressia S, Sander EJ (eds) *Work in the 21st Century*. Emerald, pp 113–137
- Schiff D, Rakova B, Ayesh A, Fanti A, Lennon M (2020) Principles to practices for responsible AI: closing the gap
- Schmidt A (2020) Interactive human centered artificial intelligence: a definition and research challenges. In: *Proceedings of the International Conference on Advanced Visual Interfaces*, 1–4. <https://doi.org/10.1145/3399715.3400873>
- Schmidt A, Herrmann T (2017) Intervention user interfaces: a new interaction paradigm for automated systems. *Interactions* 24(5):40–45
- Seamans R, Raj M (2018) AI, Labor, Productivity and the Need for Firm-Level Data (NBER Working Paper No. 24239). National Bureau of Economic Research. <http://papers.nber.org/tmp/85128-w24239.pdf>
- Seeber I, Bittner E, Briggs RO, de Vreede T, de Vreede G-J, Elkins A, Maier R, Merz AB, Oeste-Reiß S, Randrup N, Schwabe G, Söllner M (2020) Machines as teammates: a research agenda on AI in team collaboration. *Info Manage* 57(2):103174. <https://doi.org/10.1016/j.im.2019.103174>
- Shneiderman B (2020) Human-centered artificial intelligence: three fresh ideas. *AIS Trans Human-Computer Interaction* <https://doi.org/10.17705/1thci.00131>
- Silva PCR (2020) Deep learning in the maintenance industry. <https://doi.org/10.21427/Q6YR-ET27>
- Silverman K (2020) Why your board needs a plan for AI oversight. *MIT Sloan Manag Rev* 62(1):1–6
- Timofeev AV, Denisov VM (2020) Machine learning based predictive maintenance of infrastructure facilities in the cryolithozone. In: Pricop E, Fattahi J, Dutta N, Ibrahim M (eds) *recent developments on industrial control systems resilience*. Springer International Publishing, pp 49–74
- Vallas SP, Hill A (2012) Conceptualizing power in organizations. In: Courpasson D, Golsorkhi D, Sallaz JJ (eds) *Rethinking power in organizations, institutions, and markets*. Emerald, pp 165–197
- Webb M (2020) The impact of artificial intelligence on the labor market (SSRN Scholarly Paper ID 3482150). Social Science Research Network. <https://papers.ssrn.com/abstract=3482150>
- Wilkins U (2020) Artificial intelligence in the workplace—a double-edged sword. *Int J Info Learning Technol*. <https://doi.org/10.1108/IJILT-02-2020-0022>
- Wilson JH, Daugherty PR (2018) Collaborative intelligence: humans and AI are joining forces. *Harvard Bus Rev*, July–August, 114–123
- Windelband L (2017) Work requirements and qualifications in maintenance 4.0. In: Schlick CM, Duckwitz S, Flemisch F, Frenz M, Kuz S, Mertens A, Mütze-Niewöhner S (eds) *Advances in ergonomic design of systems, products and processes*. Springer, Berlin, pp 89–102
- Wolf CT, Blomberg JL (2019) Intelligent systems in everyday work practices: integrations and sociotechnical calibrations. In: *International Conference on Intelligent Human Systems Integration*, 546–550
- Yin RK (2008) *Case study research: design and methods*. Sage
- Zonta T, da Costa CA, da Righi RR, de Lima MJ, da Trindade ES, Li GP (2020) Predictive maintenance in the Industry 4.0: a systematic literature review. *Comput Indus Eng* 150:106889. <https://doi.org/10.1016/j.cie.2020.106889>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.