Keeping the Organization in the Loop as a General Concept for Human-Centered AI: The Example of Medical Imaging

Thomas Herrmann
Information and Technology Management,
Ruhr-Universität Bochum, 44780 Bochum, Germany
thomas.herrmann@rub.de

Sabine Pfeiffer
Chair of Sociology, Friedrich-Alexander-Universität
Erlangen-Nürnberg, 90429 Nürnberg, Germany
sabine.pfeiffer@fau.de

Abstract

This study emanates from work on human-centered AI and the claim of "keeping the organization in the loop". A previous study suggests a systematic framework of organizational practices in the context of predictive maintenance, and identified four cycles: using AI, customizing AI, original task handling with support of AI, and dealing with contextual changes. Since we assume that these findings can be generalized for other kinds of applications of Machine Learning (ML), we contrast the management activities that support the four cycles and their interplay with a widely different domain: the usage of AI for radiology. Our literature analysis reveals a series of overlaps with the existing framework, but also results in the need for extensions, such as holistic consideration of workflows or supervision and quality assurance.

1. Introduction

Various approaches attempt to capture how the roles between humans and artificial intelligence (AI) – particularly machine learning (ML) – are changing with respect to autonomy and distribution of decision making. In this paper, we focus on the contribution of human-centered AI (HCAI) approaches [1]–[3] to this discussion. HCAI argues generally for employing a socio-technical perspective (for instance [1]), and for keeping the human in the loop [4]. We understand HCAI as an approach that seeks to recognize and to employ human capabilities, and to further develop them when accomplishing goals by using AI. Although organizational alignment is considered an indispensable component of socio-technical systems [5], HCAI approaches do not systematically consider the organization.

We conceive organizations as a multidimensional and paradoxical phenomenon, as an entity that is integrated into complex external environments and performs there as an actor; as structures and processes – such as hierarchies – of responsibilities, and competencies, all of which are subject to their own logic and rules [6]. The actual characteristics of these levels depend on the development of everyday organizational practices. Decision-making by individual members of the organization – even when supported by ML – depend as much on diverse and conflicting interests as on collective action. All this is not only strategically orchestrated by management in organizations, it is also negotiated and processed within organizational practices on a daily basis. Therefore, in the following we use the term "organizational practices" because such practices are crucial for how an organization is translated into reality.

To take organizational practices into account systematically, we follow Herrmann and Pfeiffer's [7] concept of "keeping the organization in the loop" (KOITL). We have chosen this recently published approach because it fills the meso-level gap between keeping individual humans and keeping the society in the loop [4]. However, their framework relies empirically on case studies in the single field of predictive maintenance (PM) in industrial manufacturing, begging the question of its generalizability.

Therefore, we investigate the question (RQ1) of what changes to the KOITL framework must be made if applied to domains that are fundamentally different from PM. To answer RQ1, we chose ML applications for medical-image detection in the field of radiology. This field is comparable with PM in that it also involves well-established and complex workflows. It differs from PM in that the purpose of radiology is to prepare medical treatment for people, where the efficient use of resources is of secondary consideration rather than the primary goal as in PM. PM is a strategy for detecting the condition of an equipment to identify maintenance needs to avoid faults and malfunctions, but "not to do it too often from the viewpoints of reliability and cost" [8, p. 135]. For PM, large sensor data generated in industrial plants can be used to train ML to detect undiscovered correlations or critical anomalies [9]. By contrast, ethical issues and privacy concerns are of higher relevance in the field of radiology. Radiology seeks data sharing between institutions to

URI: https://hdl.handle.net/10125/103278
978-0-9981331-6-4

improve its quality and usage for machine learning [10], while PM prefers to keep most of the data within the company. Due to the relevance of applying huge datasets, we refer to ML as the most relevant version of AI in this paper.

Although radiology is only one exemplary field, we suggest that it provides enough contrast to industrial PM so that we can go one step further when discussing the generalizability of the keeping the organization in the loop framework. In what follows we analyze literature on the interplay between HCAI and the organizational perspective (Section 2). Since many case studies about implementing ML in the field of radiology are already documented in the literature, this paper focuses on a systematic literature review to identify the relevant facets of organizational practices of ML usage in radiology (Section 3), followed by the description of our findings (Section 4). The discussion compares the findings in both areas (Section 5), and summarizes our recommended modifications to the KOITL framework.

2. Organizational Challenges: Neglected in Human-Centered AI Research

In the engineering and social science research on human-AI interaction, organizational practices are either entirely overlooked or discussed only tangentially to topics like productivity, biases in personnel management, or profitability. Some human-centered AI research occasionally points to organizational factors in the socio-technical context [1] or to organizational factors such as the initial integration of AI with situated work practices and organizational processes [11]. Others suggest that explainable AI (XAI) has to be complemented with network-based trust [12]. Organizational practices are supposed to be needed to implement principles for responsible AI [13] or to achieve reliability and security [14].

The sociology of technology has gotten past a dichotomization of human action versus technical operation, replacing it with the more differentiated framework of "distributed agency" [15][16]. However, these approaches focus on sociological analysis rather than organizational design. The organization is addressed more as a static environmental condition than as a dynamic set of actors and practices. The interaction between humans and AI is often reduced to a few aspects or even just one, e.g. how deeply the outcome of AI affects humans [17], how individuals are connected to a technical system, or how organizational measures might achieve certain productivity gains [18]. Some authors discuss only the distribution of moral decision-making as in the case of autonomous driving [19].

Overlooked are the multiple layers between the micro level of human-computer interaction (HCI) and the macro level of societal regulation [4]. Between these lies the meso level of organizational practices, in which the way of using technology is concretized and processed as well as changed and developed on a daily basis. However, the assumption is that it is a once-off-process of integration, and the organizational practices remain mere context of an ongoing process of maintaining the integration [20]. Rarely does current research address that AI can be the subject of disputes [21] or of questions of power [22]. What is currently needed is a systematic overview of how AI and organizational practices have to be intertwined as a prerequisite and continuous context of human-centered AI. This gap may be practically justified by pursuing a clear focus of research. However, it has a systematic background, as conducting empirical research that addresses the organizational context might be more challenging than conducting HCI laboratory studies [23].

We argue that a technical concept for HCAI must be accompanied by a specification of the organizational practices that precondition its success. For example, interventions in automated processes [24], [25] or vetoing an AI outcome [26] require an organizational context that explicitly promotes or at least allows these actions. Similarly, the technical approaches that seek to provide XAI [27] cannot succeed if the organizational workflows do not provide time and opportunities to use XAI features. Concepts of how to activate and develop the complementary strengths of humans and AI [1] or to avoid oppressive ML systems [28] require appropriate organizational practices, for example sequences that have humans make their decisions [29] before considering proposals provided by ML. Within the Information Systems discourse, the need for managerial activities – such as communicating, leading, coordinating, and controlling [30] – or for organizational practices that overcome the perceived limitations of AI in radiology [31] is acknowledged, and the question of how roles will change is considered [32].

These approaches are complemented by Herrmann and Pfeiffer [7], who provide an approach that proposes a systematic and detailed framework of considering organizational practices that support HCAI. They conducted case studies in the context of predictive maintenance (PM) – considered to be one of the most important fields of ML applications for industry 4.0 [9]. Their KOITL framework identifies four types of managerial activities (Table 1) and, within these, 12 sub-activities. Managerial coordination (Table 1, field 1) is necessary for determining how ML-based PM notifications are handled and – if applicable – used for maintenance. This is closely related to the leadership and HR task of making people competent to use AI and of determining the roles of those interacting with AI (Table 1, field 2). The quality of managerial decisions also depends on coordination with an external environment (Table 1, field 3) that comprises all stakeholders and related organizational units that are not part of the plant that is subject to the PM application. Similarly, ongoing changes have to be considered (Table 1, field 4) with respect to the contextual factors that influence the performance of a PM-system. Despite the analytical separation shown in Table 1, the four types of activities are highly interrelated. They all interact with and shape the use of AI outcomes for task-handling and for evolving the AI system. We focus on the managerial activities shown in Table 1 because they allow a direct comparison with organizational measures proposed for using ML within radiology.

Table 1. Types of Managerial Activities Supporting Organizational Practices (following [7], Fig. 2).

1. Coordination of AI-related tasks a. Defining workflows and new tasks of using AI for original tasks b. Defining the ecology of relevant roles c. Reflection of interdependencies between contextual factors and AI	**2. Leadership and HR as enablers** a. Aligning roles, people, and tasks b. Employing and developing human experience and competence
3. Coordinating with the external world a. Regulation of access from the outside b. Identifying and observing various and varying external influences to be addressed c. Regulating the sharing of risks and benefits	**4. Dealing with contextual factors and changes** a. Assigning measures and resources to various types of prediction b. Tracing causes of problems c. Deciding about risk-benefit trade-offs d. Organizing multi-level long-term changes

Herrmann and Pfeiffer [7] characterize the metaphor of keeping the organization in the loop by four cycles and their interplay, which all have to be intertwined with the managerial activities described in Table 1. The first two cycles refer to dealing with AI (using and assessing AI output; customization of the AI system). The third cycle refers to the original tasks that are intended to be supported by AI. Taking changes and contextual factors into account for evolving an AI system is emphasized by the fourth cycle. All four circles are intertwined. The KOITL framework [7] has an important limitation. Although it aspires to provide a theoretical view of relevant facets of organizational practices, its original development depended on observations from the field of industrial predictive maintenance alone.

3. Systematic Literature Review

While Section 2 was based on a narrative review [33], for the literature on organizational aspects of using ML-based medical imaging in the field of radiology we conducted a systematic literature review (SLR) that allows for reproducibility and borrows from Brocke et al. [33]. To develop an overview on applying ML for radiology, we referred to Wang and Summers [10] and Liang et al. [34].

The systematic literature review focuses on the question, what statements or topics regarding organizational or managerial decisions on measures in the field of radiology can be found? We expect that grouping similar statements and comparing the set of these groups with the framework provided by Herrmann and Pfeiffer [7] will point out new requirements for reflecting and adapting the original framework.

To delimit the area of interest we started with the search term >“medical imaging” AND “radiology” AND “machine learning” AND “human-centered” AND “organizational”< (August 31, 2022) and included publications from the last 10 years. We used Google Scholar to include a wide scope of interdisciplinary research (144 hits), Web of Science (0), Scopus (15), IEEE Explore (0) and ACM digital library (7). The inclusion of the term "human-centered" was the reason that this search yielded so few hits. We used Scopus to see if broadening the search would yield more relevant results. To this end, we searched literature that also included AI or ML as keywords and also resulted in hits when either "medical imaging" or "radiology" or both were mentioned. This test did not yield any additional relevant literature after applying the exclusion criteria noted below. When we used terms such as "management" or "organization" in addition to "organizational," the search results were predominantly unsuitable for our purposes because these terms were semantically ambiguous (e.g., "the organization of this paper" or "information management system"). When we replaced “organizational” by “socio-technical”, the search results proved too narrow (27 hits with Google Scholar). However, we found one additional relevant source [35] that was not identified with the first search term.

In all, we identified 21 relevant literature sources using the following selection criteria:

- long papers written in English published in journals or conference proceedings (except one highly relevant thesis),

- papers that dealt explicitly with organizational measures and practices in context of radiology work – but not those that merely mention that the organizational level is relevant,
- papers that consider the implementation of AI or ML, not just its development.

We identified 19 further papers in the course of the narrative review, by applying backward and forward research, and by consulting an expert in the field of ML for medical imaging who pointed to the activities of the European Commission [36] and the US Food & Drug Administration [37]. Note that a literature search for publications dealing generally with the relation between ML and organizational practices would have revealed much more literature than the search focused on radiology alone. However, this limitation is justified for the purposes of this paper, which seeks to contrast two application domains on the concrete level of their typical workflows.

In order to match the content of the identified contributions with the KOITL-framework, we searched the 40 papers for statements on organizational or management-related challenges or measures. We consider it an advantage that the contexts of this papers are quite diverse, as we aim for an exploratory search that reveals a variety of different aspects. The contributions were systematically subjected to a content analysis according to Mayring [38]. Identified statements were first grouped by searching for semantically similar statements or subclasses that fit the four management activities summarized in Table 1. In a second step, we proceeded without predetermined coding and exploratively found new categories. This approach resulted in a total of seven categories, three of which are not included in Table 1 and include 15 new subcategories. In the next section, we present these categories and indicate in parentheses after each subheading how many papers contained information contributing to the category (one paper can be relevant for several categories).

4. Findings

4.1. Coordination of a Variety of Original Tasks and Workflows (13)

Applying ML in radiology not only covers selected tasks such as the analysis of CT or MR images but a whole workflow [39] of activities such as medical image registration, brain function or activity analysis, content-based image retrieval systems, medical image segmentation, text analysis of radiology reports [10], or detection of similarities between images to clarify a diagnosis [40]. Thus, a broad range of technical support is relevant, such as natural language processing for reporting [41] or making scanning tasks more efficient [42]. These tasks are related to different interests of many participants and stakeholders. A wide variety of tasks, departments, professions, and managers and employees at various hierarchical levels are involved, and their roles with respect to medical imaging have to be clarified [43].

Appropriate, seamless workflow integration [39] and reconfiguration of processes [44]–[46] is a key challenge for the success of ML applications [47] because of the highly situated nature of activities in clinical environments. ML has to contribute to a ‘‘best’’ procedure for a given clinical circumstance, including factors such as diagnostic power, radiation exposure, etc. [48]. ML applications should be well integrated into the underlying technical infrastructure, covering image archiving and communication systems [49] and procedures of providing and curating the data needed to train or retrain and evaluate ML models [48]. Thus, the entire life cycle of ML implementation needs support [50].

4.2. Leadership: Promoting Complementary Strength and Customization of AI(16)

Employees make countless experience-based decisions in their daily work [51] that are affected by ML. Some statements from studies emphasize the strength of ML [52], e.g. how algorithms can detect details that would be neglected by humans [49] [53]. However, most papers argue that the relative strengths of human radiologists and ML systems are complementary [10], [54]. ML systems “… are mainly viewed or treated as an aid or helper in a secondary position“ [10, p. 947]. Consequently, radiologists are expected to remain involved and engaged [55] so as to improve quality and efficiency [53], e.g. to guarantee forward-looking responsibility [35]. Human strengths lie in adapting to practice patterns, maintaining relationships, communicating findings appropriately, and dealing with incomplete information, e.g. of scans [42] [55] [47].

To employ and to support human strengths, certain interaction modes are offered that help draw attention to highly complex constellations or refining the search of images related to the case at hand [41] [40]. As a further example, organizational measures can specify that the human should first clarify his or her own expectations before being presented with an AI output [29] [48] [56].

Organizational structurers should allow human end-users to innovate new technologies [57], e.g. by providing feedback to the ML system to enhance its possibilities by processing new data [48]. According to Cai et al. [40], when using an interaction mode for refining ML output, users experienced less effort and more benefits, and methods of experimenting with the system's outcomes led to higher trust.

4.3. Coordination with the External World: Networking (9)

For the continuous improvement of ML-based medical image detection, studies reported several institutions cooperating in a networked effort [50], building strategic partnerships [57], and developing trust, e.g. between hospitals [58]. This concerns image sharing networks, mutual access to verified-case datasets, and exchange of experience and criteria for optimization and standardization [53], as it takes place within the emerging field of radiomics [59]. It pursues extracting useful imaging features from radiological data [48], [60], detecting relations between image-based findings and genetics, optimizing treatment and decisions about biopsies [59], and supporting a broader scope of clinical decision making [39]. Thus, radiomics helps to integrate the knowledge of many radiologists. However, data exchange is basically affected by increasing privacy regulations and ethical considerations [43].

4.4. Handling Contextual Factors and Changes (3)

Establishing AI in the practice of radiology – and its impact once established – are matters of continuous change in relation to advances in technology. Thus, a constant re-evaluation of relevant criteria and values is required [48]. These dynamics make continuous training necessary and therefore are dependent also on leadership tasks (4.2). The success of ML in regard to diagnostic performance, nursing workflows, and patient experiences depends on several contextual factors [47]. For example, monitoring the evolving legal context and adapting to regulations is a continuous organizational task [43].

4.5. Investing Additional Resources (5)

In addition to selecting appropriate ML algorithms, AI initiatives require a justification for the investment of significant resources such as hardware, software, and the human expertise (both computational and medical) [48] into which ML has to be efficiently integrated [58]. High-quality ground-truth data [39] comprising a substantial number of verified cases are a basic resource for ML training [53]. Providing these data is costly and requires regulated exchanges between institutions, expert involvement, and substantial time [49].

4.6. Ensuring Supervision and Quality (17)

The quality of ML-based results and the supervision of these systems is an important issue throughout most of the identified literature. Algorithms, quality methods, performance measures, monitoring, feedback, and accountability have to be combined [48]. Quality requirements can be grounded in ethical criteria [61] [36] [62] covering aspects such as informed consent, privacy and data protection, ownership and data quality, objectivity, human agency and oversight, technical robustness and safety, transparency, diversity, fairness, accountability, etc.

A particular quality challenge is the deployment of ML out of highly controlled environments into widespread use [55], or from one context to another [49] [42]. The question arises whether the modification of validated models by individual institutions or by empowered end-users should be permitted or whether the ML code should be frozen [55]. Modification needs testing and organizational regulations for quality assurance and possibly the involvement of the software developers' expertise [48]. There are concerns that the data used to train ML systems might be a source of uncertainty [56] [39].

Not all types of ML image-analysis applications are suitable to be subject to human monitoring [48]. A review process for ML results must be organized involving experts or institutions such as the American College of Radiology Data Science Institute [39] or the US Food & Drug Administration [37]. Radiologists must receive training for supervisory tasks [41]. Further organizational challenges are the compensation for patients affected by low-quality outputs [52] in compliance with evolving laws [43]. Obviously, dealing with quality challenges involves non-technical, organizational measures [36] such as policies for accountability [63], governance frameworks [64], guidelines [58], or checklists [65].

4.7. Considering the Subjective Factor and Communication between Stakeholders (11)

Even if quality problems with the technical systems can be solved, this does not necessarily mean that the people concerned have confidence in technically derived analyses. If people trust human decisions more than machine decisions par tout, the benefit of machine learning for radiology will be limited [10]. If people are provided with explanations about an ML result, one has to consider that their different professional and educational backgrounds will affect their perception of such explanations. Therefore, different people who are part of the social and organizational environment in which an ML application is embedded might differ in their degree of trust in the application's outputs [66]. Currently, the general population does not support a fully independent use of such systems without involving a radiologist [67], and physicians are concerned about their own reputation and relevance [68]. Radiologists are still unsure about their future role and their responsibility for AI outcomes in relation to unsolved ethical and legal issues [69, p. 10]. They have low confidence in the results such systems generate [70]. Medical students are also skeptical [71]. Thus, the subjective attitudes – as well as the goals [62] – of different stakeholders need to be addressed at the beginning of AI implementation [11].

Exchanging knowledge in clinical practice requires much communication and mutual understanding, as well as the integration of multiple stakeholder perspectives [50] [62]. A common "language" has to be established between multiple stakeholders including physicians, scientists, technologists, industry leaders, and patients [48].

Table 2: The extension of management activities

1. Coordination of AI-related tasks a. Defining workflows and new tasks of using AI **within workflows of** original tasks (4.1) b. Defining the ecology of relevant roles (4.1) c. Reflection of interdependencies between contextual factors and AI (4.4.) **d. Checking for integration of all supports (4.1)** **e. Adding resources (4.5)** **f. Preparing the interplay between quality assessment and adaptation of AI (4.6)** **g. Supporting exchange between stakeholders' perspectives (4.7)**	2. Leadership and HR as enablers a. Aligning roles, people, and tasks (4.1) b. Employing and developing human experience and competence (4.2) **c. Preparing users to contribute to adaptations (4.2)** **d. Preparing staff for the shortcomings of AI and for supervisory tasks (4.6).** **e. Preparing staff for dealing with the subjective experience and attitudes of stakeholders (4.7)**
3. Coordinating with the external world a. Regulation of access from the outside (4.3) b. Identifying and observing various and varying external influences to be addressed (4.3) **c.** Regulating the sharing of risks and benefits, **e.g. of exchanging data and technology (4.4)** **d. Building strategic partnerships (4.3)** **e. Support and regulation of cross-institutional exchange of underlying data (4.3)**	4. Dealing with contextual factors and changes a. Assigning measures and resources to various types of prediction (4.4) b. Tracing causes of problems (4.4) **c.** Deciding about risk-benefit trade-offs, **e.g. of exchanging data and technology (4.4)** d. Organizing multi-level long-term changes **e. Observing and predicting societal changes (4.4)**
5. Supervision and quality assurance (4.6) **a. Organizing supervision routines** **b. Organizing tests when new algorithms are deployed** **c. Regulating the modification of ML algorithms** **d. Quality assurance for data to be used and shared** **e. Organizing the involvement of external expertise**	

5. Discussion

In what follows we discuss each of the categories of the findings and decide where an extension of management activities (Table 1) is reasonable. These extensions are indicated in Table 2 in boldface. In parentheses, we additionally refer to the subsection that mainly supports the subcategories of Table 2.

The distinction proposed by Herrmann und Pfeiffer [7] between AI usage and the original tasks AI is designed to support may also reasonably be applied to radiology. However, ML-based image detection influences a much broader scope of tasks than PM. The entire image-detection workflow context (see 4.1) is affected (Table 2, 1a). Thus, the suggested [7] interplay between coordinative management activities with AI use for original tasks should include checking for integration of all kind of possible support with AI/ML (Table 2, 1d) and seamlessly integrating AI into workflows (see 4.1). This integration specifically includes also organizational practices such that generated data can be exchanged, compared, and processed to serve as a basis for mutual knowledge and for advancing ML within the entire ML lifecycle. Furthermore, the practice of radiology, similarly to PM, is structured by the various roles to which tasks, rights and duties have been assigned (Table 2, 1b, 2a). For example, not only image detection but also the task of image-scanning is an integral part of the relevant workflow, and management has to decide whether nurses are authorized to judge the quality of a scan [47].

The relevance of complementary strengths (see 4.2) in the field of clinical ML support is similar to what is found in the PM context (Table 2, 2b). Relevant as well is the promotion of human strengths and its interplay in refining AI results and in contributing to AI evolution through continuous customization, e.g. by retraining the ML models. The current proposals of how human strengths can be involved to help customize AI (see 4.2) are mostly technical. However, the HCAI discourse (see Section 2) points to the involvement of organizational practices in making customization possible in the course of quality assessment (Table 2, 1f) and in preparing the users accordingly (Table 2, 2d). We assume that in the case of the PM-related framework, preparing the user for contributing to ML adaptation is also of certain relevance, but it is not explicitly mentioned as a management activity in Herrmann and Pfeiffer [7]. It was, however, implicitly addressed in that leadership is considered relevant. We suggest listing these preparation tasks (Table 2, 2c to 2e), since they are explicitly discussed as being relevant for medical practitioners [11] .

During the development and use of ML-applications, multiple communication exchanges take place between stakeholders with widely different backgrounds and experience. Thus, building strategic partnerships is relevant (Table 2, 3d). Additionally, medical ML is characterized by a higher relevance of cross-institutional exchange of ML-relatable data (Table 2, 3e) as demonstrated by the example of radiomics as a developing field (see 4.3). While in the field of radiology the exchange of data has to be regulated primarily under ethical and privacy aspects, for PM it is of interest to avoid access to competition-relevant data from the outside.

The deliberate consideration of external factors and their continuous change (4.4) appears to be relevant in both areas of medical imaging and PM. Both, for example, have to be aware of and prepared for continuous technical change and challenges (Table 2, 4a, 4b, 4d). While PM is more focused on the economic context, e.g., availability or price of spare parts, radiology must additionally deal with societal changes, such as ethical and political discourses (Table 2, 4e). Of particular interest to radiology is how risks and benefits are shared when data or ML algorithms are exchanged (Table 2, extensions of 3c and 4c).

The implementation of ML-based image recognition can be cost-intensive and requires many additional resources (see 4.5). This aspect has to be added to the management activities of coordinating AI related tasks (Table 2, 1e).

As ML-based predictions can have a direct impact on patient well-being, the reliability and validation of the quality of ML results (see 4.6) is of high relevance. This is reflected in the discussion about different methods of quality assurance and about organizational measures that should take place. Because of the high relevance of quality assurance and supervision, this aspect is newly inserted in Table 2 as an extension of the PM-related findings (Table 2, 5). It comprises managerial activities such as organizing supervision, the testing and regulation of modifications, and involving external expertise. These measures are intertwined with leadership (preparing staff for the shortcomings of AI and for supervisory tasks; Table 2, 2d) and coordination (preparing the interplay between quality assessment and adaptation; Table 2, 1f).

To be considered also is that different stakeholders (see 4.7) – all of whom differently affected – influence how ML systems and their outcomes are subjectively experienced. Staff have to be prepared to deal with this subjective factor (Table 1, 2e). Within radiology (see 4.7), it is highly relevant for the coordination between dealing with AI and the original tasks (such as diagnosing and treating) that communication with patients and relatives and involving them into decision making is included (Table 2, 1g, 2e). This subjective factor seems to be of higher relevance in the case of radiology than with PM.

When expanding Table 1 to Table 2, it becomes clear that the five types of management activities must be closely intertwined to induce sustainable organizational practices in support of HCAI. These practices also determine the extent to which management actions are effectively implemented in dealing with AI and the use of AI results for the original tasks (e.g., repair work or medical treatment). In both domains, the notion of ML systems as autonomous agents [72] does not play a dominant role, yet it is recognized that ML technology is dynamically evolving, as described as a basic feature of AI [30]. Other, more general overviews of relevant organizational measures in the context of AI that address management challenges [30], principles of human-AI collaboration [73], or mixed human-AI initiative [74] do not provide the level of detail found in the field of radiology and PM in terms of considerations of management activities and their interplay with the use and adaptation of AI in the context of existing workflows.

6. Conclusion

The analysis of the two widely different domains points to the conclusion that keeping the organization in the loop (KOITL) is of general relevance for human-centered AI. The analyzed papers hint at various organizational measures being relevant. For radiology, there is no theoretical framework for analyzing the interplay between ML and organizational practices. To fill this gap and with respect to the research question (RQ1), we extended the framework of Herrmann and Pfeiffer [7] by the extensions boldfaced in Table 2. Those managerial activities that had already been identified (see Table 1) also proved just as relevant in the context of radiology.

We suggest that Table 2 provides a theoretical basis for how the interplay between AI usage and customization, the support of original tasks by AI, and dealing with contextual factors and ongoing changes can be supported by managerial activities. Supervision and quality assurance are added as an additional type of management activity that might prove of general relevance for implementing HCAI since they point to human supervisory tasks. The KOITL framework, which includes these extended management activities, can help in the creation of checklists for practitioners.

However, such checklists would need to be adapted to each specific domain context [23].

A limitation of our results is that they are based on a literature review rather than additional case studies. However, there are already a number of case studies reported in the literature found. Nevertheless, certain aspects may be usefully explored in new case studies, e.g., whether adaptation and alignment can be promoted organizationally or whether this is prevented by regulatory contexts. Furthermore, it would be interesting to explore implementation in a third domain, e.g., AI applications for human resource management, in which bias avoidance and fairness are more central issues. Alternatively, the details found in this work could be compared to more general frameworks of management challenges as increasingly discussed in the Information Systems context [30].

7. Acknowledgement

This work was supported by the Humaine project (Human Centered AI Network), funded (02L19C200) by the German Federal Ministry of Education and Research (BMBF), and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 442171541. We thank Axel Mosig for hints to complete the literature review.

8. References

[1] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister, "Hybrid Intelligence," Bus. Inf. Syst. Eng., vol. 61, no. 5, pp. 637–643, Oct. 2019.

[2] A. Schmidt, "Interactive Human Centered Artificial Intelligence: A Definition and Research Challenges," in Proceedings of the International Conference on Advanced Visual Interfaces, Salerno Italy, 2020, pp. 1–4.

[3] B. Shneiderman, Human-Centered AI. Oxford University Press, 2022.

[4] I. Rahwan, "Society-in-the-loop: programming the algorithmic social contract," Ethics Inf. Technol., vol. 20, no. 1, pp. 5–14, 2018.

[5] C. Kirsch, P. Troxler, and E. Ulich, "Integration of people, technology and organization: the european approach," in Symbiosis of Human and Artifact, vol. 20, Y. Anzai, K. Ogawa, and H. Mori, Eds. Elsevier, 1995, pp. 957–961.

[6] N. Phillips and T. B. Lawrence, "The turn to work in organization and management theory: Some implications for strategic organization," Strateg. Organ., vol. 10, no. 3, pp. 223–230, Aug. 2012.

[7] T. Herrmann and S. Pfeiffer, "Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence," AI&Soc, 2022.

[8] T. Nakagawa, Maintenance Theory of Reliability. Springer Science & Business Media, 2006.

[9] J. Dalzochio et al., "Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges," Comput. Ind., vol. 123, p. 103298, 2020, doi: 10.1016/j.compind.2020.103298.

[10] S. Wang and R. M. Summers, "Machine learning and radiology," Med. Image Anal., vol. 16, no. 5, pp. 933–951, Jul. 2012.

[11] C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry, "'Hello AI': Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making," Proc. ACM Hum.-Comput. Interact., vol. 3, no. CSCW, pp. 1–24, Nov. 2019,.

[12] P. Andras et al., "Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems," IEEE Technol. Soc. Mag., vol. 37, no. 4, pp. 76–83, 2018.

[13] D. Schiff, B. Rakova, A. Ayesh, A. Fanti, and M. Lennon, "Principles to Practices for Responsible AI: Closing the Gap," ArXiv200604707 Cs, 2020, Accessed: Jan. 10, 2021. [Online]. Available: http://arxiv.org/abs/2006.04707

[14] B. Shneiderman, "Human-Centered Artificial Intelligence: Three Fresh Ideas," AIS Trans. Hum.-Comput. Interact., pp. 109–124, 2020.

[15] B. Latour, "Mixing Humans with Non-Humans: Sociology of a Door-Closer," Soc. Probl., vol. 35, no. 3, pp. 298–310, 1988.

[16] W. Rammert, "Where the action is: distributed agency between humans, machines, and programs," TU Berlin, Berlin, 4–2008, 2008. [Online]. Available: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-12331

[17] A. S. Jago, "Algorithms and Authenticity," Acad. Manag. Discov., vol. 5, no. 1, pp. 38–56, 2019.

[18] J. H. Wilson and P. R. Daugherty, "Collaborative Intelligence: Humans and AI Are Joining Forces," Harv. Bus. Rev., no. July-August, pp. 114–123, 2018.

[19] E. Awad et al., "The Moral Machine experiment," Nature, 2018.

[20] A. Mateescu and M. C. Elish, "AI in Context: The Labor of Integrating New Technologies," Data & Society pp. 1–54, 2019., Accessed: Aug. 2022 Available: https://apo.org.au/sites/default/files/resource-files/2019-01/apo-nid217456.pdf.

[21] K. C. Kellogg, M. A. Valentine, and A. Christin, "Algorithms at Work: The New Contested Terrain of Control," Acad. Manag. Ann., vol. 14, no. 1, pp. 366–410, 2020.

[22] S. P. Vallas and A. Hill, "Conceptualizing Power in Organizations," in Rethinking Power in Organizations, Institutions, and Markets, D. Courpasson, D. Golsorkhi, and J. J. Sallaz, Eds. Bingley: Emerald, 2012, pp. 165–197.

[23] R. M. Davison and M. G. Martinsons, "Context is king! Considering particularism in research design and reporting," J. Inf. Technol., vol. 31, no. 3, pp. 241–249, Sep. 2016.

[24] T. Herrmann, "Socio-Technical Design of Hybrid Intelligence Systems–The Case of Predictive Maintenance," in International Conference on Human-Computer Interaction, 2020, pp. 298–309.

[25] A. Schmidt and T. Herrmann, “Intervention user interfaces: a new interaction paradigm for automated systems,” interactions, vol. 24, no. 5, pp. 40–45, 2017.
[26] B. Rakova, J. Yang, H. Cramer, and R. Chowdhury, “Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for shifting Organizational Practices,” ArXiv200612358 Cs, Nov. 2020, Accessed: Jan. 10, 2021. [Online]. Available: http://arxiv.org/abs/2006.12358
[27] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI),” IEEE Access, vol. 6, pp. 52138–52160, 2018.
[28] G. C. Kane, A. G. Young, A. Majchrzak, and S. Ransbotham, “Avoiding an Oppressive Future of Machine Learning: A Design Theory for Emancipatory Assistants,” MIS Q., vol. 45, no. 1, pp. 371–396, 2021.
[29] R. Fogliato et al., “Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging.” arXiv, May 19, 2022. Accessed: Jun. 03, 2022. [Online]. Available: http://arxiv.org/abs/2205.09696
[30] N. Berente, B. Gu, J. Recker, and R. Santhanam, “Managing artificial intelligence,” MIS Q., vol. 45, no. 3, pp. 1433–1450, 2021.
[31] B. Kim, I. Koopmanschap, M. H. R. Mehrizi, M. Huysman, and E. Ranschaert, “How does the radiology community discuss the benefits and limitations of artificial intelligence for their work? A systematic discourse analysis,” Eur. J. Radiol., vol. 136, p. 109566, 2021.
[32] S. Seidel, N. Berente, A. Lindberg, K. Lyytinen, B. Martinez, and J. V. Nickerson, “Artificial Intelligence and Video Game Creation: A Framework for the New Logic of Autonomous Design,” J. Digit. Soc. Res., vol. 2, no. 3, pp. 126–157, 2020.
[33] J. vom Brocke, A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, “Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research,” Commun. Assoc. Inf. Syst., vol. 37, 2015.
[34] Y. Liang, L. He, and X. ‘Anthony’ Chen, “Human-Centered AI for Medical Imaging,” in Artificial Intelligence for Human Computer Interaction: A Modern Approach, Y. Li and O. Hilliges, Eds. Cham: Springer International Publishing, 2021, pp. 539–570.
[35] M. Sand, J. M. Durán, and K. R. Jongsma, “Responsibility beyond design: Physicians’ requirements for ethical medical AI,” Bioethics, vol. 36, no. 2, pp. 162–169, Feb. 2022.
[36] European Commission, C. and T. Directorate General for Communications Networks, and High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI. 2019. Accessed: May 23, 2021. [Online]. Available: https://data.europa.eu/doi/10.2759/346720
[37] Food & Drug Administration and others, “Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD),” 2019.
[38] P. Mayring, Qualitative content analysis: a step-by-step guide. Sage, 2021.
[39] G. Choy et al., “Current Applications and Future Impact of Machine Learning in Radiology,” Radiology, vol. 288, no. 2, pp. 318–328, Aug. 2018.
[40] C. J. Cai et al., “Human-centered tools for coping with imperfect algorithms during medical decision-making,” in Proc. of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–14.
[41] S. H. Tajmir and T. K. Alkasab, “Toward Augmented Radiologists,” Acad. Radiol., vol. 25, no. 6, pp. 747–750, Jun. 2018.
[42] P. Lakhani et al., “Machine Learning in Radiology: Applications Beyond Image Interpretation,” J. Am. Coll. Radiol., vol. 15, no. 2, pp. 350–359, Feb. 2018.
[43] R. Crootof, M. E. Kaminski, and W. N. Price II, “Humans in the Loop,” SSRN Electron. J., 2022.
[44] S.-L. Wamba-Taguimdje, S. Fosso Wamba, J. R. Kala Kamdjoug, and C. E. Tchatchouang Wanko, “Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects,” Bus. Process Manag. J., vol. 26, no. 7, pp. 1893–1924, May 2020.
[45] M. Kargl, P. Regitnig, H. Müller, and A. Holzinger, “Towards a Better Understanding of the Workflows: Modeling Pathology Processes in View of Future AI Integration,” in Artificial Intelligence and Machine Learning for Digital Pathology, vol. 12090, A. Holzinger, R. Goebel, M. Mengel, and H. Müller, Eds. Cham: Springer International Publishing, 2020, pp. 102–117.
[46] K. Santosh and L. Gaur, Artificial Intelligence and Machine Learning in Public Healthcare: Opportunities and Societal Impact. Singapore: Springer, 2021.
[47] E. Beede et al., “A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy,” in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA, 2020, pp. 1–12.
[48] G. Currie, K. E. Hawk, E. Rohren, A. Vial, and R. Klein, “Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging,” J. Med. Imaging Radiat. Sci., vol. 50, no. 4, pp. 477–487, Dec. 2019.
[49] T. Martín Noguerol, F. Paulano-Godino, M. T. Martín-Valdivia, C. O. Menias, and A. Luna, “Strengths, Weaknesses, Opportunities, and Threats Analysis of Artificial Intelligence and Machine Learning Applications in Radiology,” J. Am. Coll. Radiol., vol. 16, no. 9, pp. 1239–1247, Sep. 2019.
[50] J. Roski, E. J. Maier, K. Vigilante, E. A. Kane, and M. E. Matheny, “Enhancing trust in AI through industry self-governance,” J. Am. Med. Inform. Assoc., vol. 28, no. 7, pp. 1582–1590, Jul. 202.
[51] S. Pfeiffer, “Kontext und KI: Zum Potenzial der Beschäftigten für Künstliche Intelligenz und Machine-Learning,” HMD - Prax. Wirtsch., vol. 57, pp. 465–479, 2020.
[52] K. Chockley and E. Emanuel, “The End of Radiology? Three Threats to the Future Practice of Radiology,” J. Am. Coll. Radiol., vol. 13, no. 12, pp. 1415–1420, Dec. 2016.

[53] J. H. Thrall et al., “Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success,” J. Am. Coll. Radiol., vol. 15, no. 3, pp. 504–508, Mar. 2018.
[54] W. Bulten et al., “Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study,” Lancet Oncol., vol. 21, no. 2, pp. 233–241, Feb. 2020.
[55] M. Kohli, L. M. Prevedello, R. W. Filice, and J. R. Geis, “Implementing Machine Learning in Radiology Practice and Research,” Am. J. Roentgenol., vol. 208, no. 4, pp. 754–760, Apr. 2017.
[56] S. Lebovitz, N. Levina, New York University, H. Lifshitz-Assa, and New York University, “Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts’ Know-What,” MIS Q., vol. 45, no. 3, pp. 1501–1526, 2021.
[57] D. Wiljer and Z. Hakim, “Developing an Artificial Intelligence–Enabled Health Care Practice: Rewiring Health Care Professions for Better Care,” J. Med. Imaging Radiat. Sci., vol. 50, no. 4, pp. S8–S14, 2019.
[58] L. Kemppainen, M. Pikkarainen, P. Hurmelinna-Laukkanen, and J. Reponen, “Connected health innovation: data access challenges in the interface of AI companies and hospitals,” Technol. Innov. Manag. Rev., vol. 9, no. 12, pp. 43–54, 2019.
[59] M. L. Giger, “Machine Learning in Medical Imaging,” J. Am. Coll. Radiol., vol. 15, no. 3, pp. 512–520, 2018.
[60] P. Lambin et al., “Radiomics: Extracting more information from medical images using advanced feature analysis,” Eur. J. Cancer, vol. 48, no. 4, pp. 441–446, 2012.
[61] M. Kohli and R. Geis, “Ethics, Artificial Intelligence, and Radiology,” J. Am. Coll. Radiol., vol. 15, no. 9, pp. 1317–1319, Sep. 2018.
[62] S. Creemers, “Determinants for successful deployment of clinical prediction models,” PhD Thesis, Master’s thesis. University of Turku, Finland, 2019.
[63] B. R. Jackson et al., “The Ethics of Artificial Intelligence in Pathology and Laboratory Medicine: Principles and Practice,” Acad. Pathol., vol. 8, p. 2374289521990784, Jan. 2021.
[64] L. B. Weissinger, “AI, Complexity, and Regulation,” in The Oxford Handbook of AI Governance, J. Bullock, Y.-C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. Young, and B. Zhang, Eds. Oxford University Press, 2022.
[65] M. A. Madaio, L. Stark, J. Wortman Vaughan, and H. Wallach, “Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI,” in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA, Apr. 2020, pp. 1–14.
[66] U. Ehsan and M. O. Riedl, “Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach,” in HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence, vol. 12424, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds. Cham: Springer International Publishing, 2020, pp. 449–466.
[67] Y. P. Ongena, D. Yakar, M. Haan, and T. C. Kwee, “Artificial Intelligence in Screening Mammography: A Population Survey of Women’s Preferences,” J. Am. Coll. Radiol., vol. 18, no. 1, Part A, pp. 79–86, 2021, doi: 10.1016/j.jacr.2020.09.042.
[68] S. B. Maier, E. Jussupow, and A. Heinzl, “Good, bad, or both? Measurement of physician’s ambivalent attitudes towards AI,” ECIS 2019 Proc., 2019, [Online]. Available: https://aisel.aisnet.org/ecis2019_rp
[69] European Society of Radiology (ESR), “Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology,” Insights Imaging, vol. 10, no. 105, pp. 1–11, 2019, doi: https://doi.org/10.1186/s13244-019-0798-3.
[70] F. Jungmann et al., “Attitudes Toward Artificial Intelligence Among Radiologists, IT Specialists, and Industry,” Acad. Radiol., vol. 28, no. 6, pp. 834–840, 2021, doi: 10.1016/j.acra.2020.04.011.
[71] C. J. Park, P. H. Yi, and E. L. Siegel, “Medical Student Perspectives on the Impact of Artificial Intelligence on the Practice of Medicine,” Curr. Probl. Diagn. Radiol., 2020, doi: 10.1067/j.cpradiol.2020.06.011.
[72] A. Baird and L. M. Maruping, “The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts.,” MIS Q., vol. 45, no. 1, 2021.
[73] M. Sujan, R. Pool, and P. Salmon, “Eight human factors and ergonomics principles for healthcare artificial intelligence,” BMJ Health Care Inform., vol. 29, no. 1, p. e100516, Feb. 2022.
[74] M. Muller and J. Weisz, “Extending a Human-AI Collaboration Framework with Dynamism and Sociality,” in 2022 Symposium on Human-Computer Interaction for Work, Durham NH USA, Jun. 2022, pp. 1–12.