Individual and Organisational Capacities for Assessing "Trustworthiness" of AI Systems in Healthcare Settings

The Crucial Role of Structural Empowerment

Oliver Behn, Marc Jungtäubl, Michael Leyer, Mascha Will-Zocholl

Abstract *The integration of AI systems in healthcare settings fundamentally transforms professional work practices and introduces new requirements for professionals' competence in assessing AI system trustworthiness. While regulatory frameworks such as the EU AI Act mandate sufficient AI knowledge from employees, current studies reveal a significant skills gap in managing AI systems. This article examines what governance structures organizations must establish to enable employees to assess AI system trustworthiness and ensure continuous capability development. Drawing on Structural Empowerment Theory (SET), it analyses how human-AI interaction evolves from a tool-based to a collaborative-like relationship and identifies the organizational conditions required to empower professionals. Through scenario-based case analyses from the medical field, the study demonstrates that assessing AI trustworthiness is not solely a matter of individual competence but depends significantly on organizational structures that ensure access to information, resources, and participatory governance processes. The analysis reveals that with – at least seemingly – ever-advancing AI systems, the complexity of trustworthiness assessment grows, making individual evaluations alone insufficient. The findings indicate that successful AI integration goes beyond technical training and requires a comprehensive approach that embeds structural empowerment across multiple organizational levels. AI governance should therefore not only focus on technology regulation but shape human-AI interaction in ways that strengthen professional autonomy, competence, and trust. This necessitates establishing new learning and design processes, communication channels, and reflection spaces. The article develops human-centric design principles and presents a capability-oriented framework for effective organizational AI governance. While structural empowerment represents a promising approach to fostering AI competencies, its limitations in complex and dynamic environments are also highlighted. Embedding structural empowerment within an adaptive governance framework proves crucial for equipping healthcare professionals to navigate the challenges of (potentially increasing) autonomous AI systems while maintaining human-centered decision-making and ethical standards.*

1. Introduction

As artificial intelligence (AI) continues to be integrated into business operations, expectations from the public and regulatory bodies regarding the proficiency and accountability of professionals utilizing AI systems have increased (Dlugatch et al. 2023; Shneiderman 2020). Professionals must cultivate a profound understanding of AI functionalities and wide-ranging implications. For instance, the European Union's (EU) AI Act (EU, 2024) mandates that companies developing and deploying AI systems must ensure that their employees as well as any individuals operating or using these systems have sufficient knowledge about AI (ibid.). Although the term "sufficient" remains undefined, the EU AI Act stipulates that employees must at least grasp fundamental AI principles, accurately interpret AI-generated outputs, and understand the ethical and legal ramifications of AI system usage. However, it remains uncertain whether such a level of understanding and competence is widespread, as AI systems differ significantly from traditional information systems (IS) and are still, for many employees at least perceived as, relatively new in (the possibility of) everyday use, leaving many professionals with limited experience in their use and oversight (Zhang et al. 2023). Additionally, the human-AI relationship is transforming from a simple tool-based interaction to a more intensive collaboration (Baird and Maruping 2021). A recent survey of AI practitioners, actively involved in building, deploying and/or maintaining AI solutions, found that many lack confidence in managing AI systems, with only one in four believing they possess adequate skills and knowledge (DataRobot 2024). Moreover, studies reveal a widespread sentiment among employees of feeling unprepared to effectively handle AI systems, thereby underscoring a critical skills gap that permeates AI-driven workplaces (Cetindamar et al. 2022; Cheong 2024).

Furthermore, those new technologies change existing power relations in the field where they are implemented and used. New players are being added, mechanisms and modes of cooperation and coordination are changing. These changes can be subsumed under governance as the principle of coordination and cooperation between actors from different sectors, beyond purely state or market-based regulation, taking into account hybrid structures and diverse patterns of organization (Matys 2025). In the context of AI, these are generally speaking state institutions, universities and companies that develop or/and use these technologies, society as an applying actor, etc.

Against this backdrop, this article explores how the adoption of AI at an organizational level, understood as the implementation and use of AI systems within operational workflows as well as the appropriation by employees, affects work practices of professionals and how those work practices affect the AI systems. In doing so, it emphasizes the need for organizations not only to integrate AI systems into their processes, but also to establish governance structures that enable employees

to use these systems responsibly and effectively. A central challenge in this context is ensuring that employees are equipped to assess the trustworthiness of AI systems and continuously develop the capacities needed to engage with them in a meaningful way. Specifically, this contribution addresses the following research questions:

1) What governance structures should organizations establish to enable employees to assess AI trustworthiness?
2) How can these governance structures ensure the continuous development and maintenance of AI-related capacities to deal with AI-systems?

To illustrate these challenges, we later present two scenarios from the medical field (see 5.2). This field is particularly relevant, as it is on the verge of a significant transformation, shifting from using AI systems in a supportive manner to integrating them as actively involved in decision-making processes (Zou and Topol 2025). Also, this field is paradigmatic, as the physicians working in it have characteristics such as high qualifications, a strong understanding of the profession with a traditionally influential profession, and representation of interests as a whole, as well as far-reaching, important and necessary autonomy of action on the one hand and responsibility for decisions and actions on the other. Following the capacity-oriented approach, introduced in the general introduction of this book, we demonstrate why assessing AI trustworthiness is especially complex for healthcare professionals. However, although our analysis centers on experts in the medical field such as physicians, its insights are broadly applicable to other industries, particularly those involving high-reliability decision-making (e.g., finance, law enforcement). We emphasize the critical role of structures and resources on an organizational level as well as autonomy and subjective strategies (work practices) on an individual level to make employees capable of dealing with AI systems. This means being able to use it but also being involved in its organizational embedding and design. Along the Structural Empowerment Theory (SET), we will show how organizations can (and need to) support the development of capacities of employees. Using a scenario-based case design, we illustrate implications for practical realization. Additionally, this article seeks to highlight the importance of developing previously missing organizational AI governance guidelines that reflect the needs and perspectives of employees. Specifically, it aims to present a capacity-oriented framework for effective AI governance in organizations and – finally – 'good governance for good work'.

The article is structured as follows: Section 2 examines the differences between AI-enabled software and traditional software, the distinction between narrow AI and in some disciplines so called 'agentic' AI,[1] as well as capability-based perspec-

1 Not only – but especially – from a sociological perspective, descriptions of technical artefacts such as AI that imply, they themselves possess agency, must be treated with caution. Even if

tives of AI competencies. Section 3 introduces the Structural Empowerment Theory and its relevance to AI governance in organizations, followed by Section 4 presenting a capacity-based approach towards shaping "good" governance including two scenarios of AI use in a medical context and the exploration of elements with impact on individual and organizational capacities required to effectively interact with and, in general, use AI systems. Section 5 provides a conclusion and practical implications.

2. Background

2.1. How Does AI Differ from Traditional Software?

AI-driven or at least AI-supported information systems represent a novel and increasingly prevalent type of technical systems (e.g., especially software systems) that stand apart from traditional counterparts (Murray et al. 2021). While both aim to achieve defined goals efficiently and logically through a set of rules (Barocas and Selbst, 2016), the key distinction lies in the possibility and ability of some AI systems to learn and adapt. Unlike, e.g., standard software, which relies on pre-programmed instructions, AI systems process vast amounts of diverse data over varying time spans and derive conclusions independently (Zuboff 2023). This capability stems mainly from machine learning (ML), where algorithms identify patterns and refine their operations based on the data they receive, without requiring explicit programming for every scenario (Kellogg et al. 2020). By leveraging this learning mechanism, some AI systems are said to be able to build up their own understanding of decision problems and automatically generate optimized solutions to achieve specified goals (Dietvorst et al. 2015).

One of the key distinctions among AI systems is the difference between narrow AI and 'agentic' or, to put it more nuanced, advanced AI. While narrow AI operates within predefined constraints and requires direct human oversight, advanced AI exhibits a higher degree of automatic processing, enabling it to act automatically in complex environments. According to Mitchell et al. (2025), an AI is considered "agentic" if it:

- Pursues goals independently without direct human control.

---

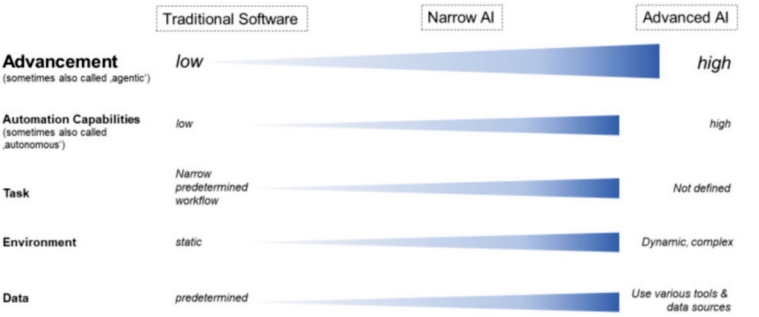
some Science-and-Technology-Studies schools speak of certain actants, sometimes referring to capacities inscribed in technical systems, yet here the point is usually only that artefacts exert effects. Genuine agency – with meaning, autonomy, and accountable responsibility – is usually reserved for human beings (ibid.). Labelling AI indiscriminately as agentic therefore risks obscuring human (individual and organizational) responsibility and forgetting that technology is always socially embedded in its origins and consequences.

- Uses tools and resources independently to complete tasks without requiring explicit human commands.
- Makes 'autonomous' (or rather: automated) decisions by calculated selecting, processing, and responding to information on its own.
- Operates effectively in open environments, meaning it can adapt flexibly to new situations.

Thus, unlike traditional narrow AI systems that mainly follow rule-based operations, novel advanced AIs are more proactive and can (help) solve unforeseen problems independently, choose based on calculations between different strategies and actively utilize external tools or APIs to achieve objectives (Kapoor et al. 2024). Historically, AI technologies were developed to enhance human decision-making rather than functioning independently. Berente et al. (2021) note that earlier AI systems, such as decision support and expert systems, primarily relied on structured data provided by human analysts. In contrast, modern AI systems are increasingly capable of processing information that has not been explicitly delegated by humans.

However, the distinction between narrow and 'agentic', advanced AI systems is not clear-cut. Instead, there is a continuum of AI systems varying in terms of the degree of (possible) automation (Mitchell et al. 2025; see Figure 1). As AI technologies advance from narrow, task-specific systems to advanced entities, their capabilities and applications expand significantly. On the other side, the rapid advancement of AI offers not only chances but also introduces new risks (Chan et al. 2023). For example, due to their automated decision-making capabilities, advanced AI systems may generate unintended consequences that are difficult to predict or control. While narrow AI systems typically operate under human oversight, advanced AI systems can operate in ways that go beyond initial programming, potentially leading to ethical dilemmas, security risks, and legal concerns. Chan et al. (2023) also highlight the potential for higher systemic and delayed harms caused by such AI systems. These harms may gradually accumulate, disproportionately affecting marginalized communities and reinforcing existing biases in ways that are difficult to detect and mitigate. Consequently, the increasing capabilities of AI-driven automated processing present new challenges in assessing their trustworthiness, particularly in high-reliability fields such as healthcare. For example, Zou and Topol (2025) note that the U.S. Food and Drug Administration traditionally evaluates AI-enabled medical devices as tools for addressing specific tasks. However, due to new developments, such regulatory approaches have to be at least evaluated and probably adjusted.

Figure 1: Levels of advancement in software and AI tools. Own illustration based on Mitchell et al. 2025.



2.2. What Is the Impact of AI Systems in Organizations?

This technological shift, distinguishing AI systems from earlier IS, brings significant implications across social, ethical, and organizational dimensions. For example, since then, AI systems have differed fundamentally from traditional IS by automated (self-)learning from data rather than relying on predefined, human-provided instructions (Samuel 1959). That capability allows AI systems to find solutions based on large amounts of data that are difficult or impossible for humans to manage, and offers the opportunity for unique and complementary outputs (Fügener et al. 2022). So, AI technology can outperform human decision-makers especially in processing vast datasets at exceptional speeds while maintaining a primarily (at least mathematical-formal) rational approach (Smith and McKeen 2011), leading to new forms of human-technology collaboration. With these capabilities, AI systems are getting increasingly more attractive also for high-reliability environments like healthcare (Lebovitz et al. 2022), becoming a key factor by assisting with complex tasks traditionally managed by human experts. This opportunity of collaboration can leverage the strengths of both humans and AI, enhancing precision and efficiency in critical fields like diagnostics, treatment planning, and operational safety. However, the characteristics of AI systems present challenges for real-world deployment and human-AI-collaboration. Since AI systems rely on statistical methods, they are inherently probabilistic and prone to errors (Berente et al. 2021). These statistical approaches can result in inconsistent behavior, further compounded by the opacity of the systems (Amershi et al. 2019; Schuetz and Venkatesh 2020).

Overall, the characteristics of AI systems also necessitate the development of new human skills. Individuals must learn to make critical decisions about whether and when to delegate tasks to AI, assess the reliability of AI-generated outcomes, and effectively communicate these results to stakeholders, such as customers or patients. This shift emphasizes the importance of fostering AI competencies, trust calibration, and communication expertise to navigate the complexities of human-AI collaboration.

2.3. Capacity-Based Perspective of AI Competencies

The integration of AI systems into organizational contexts – particularly in high-reliability environments such as healthcare – is not merely a technical development. It marks a deeper transformation of professional roles, decision-making structures, and (thus) governance arrangements. The shift from traditional software toward increasingly advanced AI expands both the opportunities and the uncertainties professionals face. The EU AI Act, among other regulatory initiatives, reflects this shift by requiring sufficient 'competence' to use, oversee, and evaluate AI systems responsibly. But what exactly constitutes this 'sufficient competence'? What does it mean, in practice, to have the capacity to evaluate, e.g., the trustworthiness of AI systems, particularly when their logic, limitations, and impacts often are opaque?

This article proposes that such evaluation requires more than isolated technical knowledge. It entails developing labor capacity which leads to the ability of evaluating the trustworthiness of AI usage. This includes, but is not limited to (see Ramezani et al. 2025):

- technical understanding of how AI systems function, including awareness of their limitations and potential biases.
- confidence and competence to question or interpret AI-generated recommendations.
- understanding of how AI decisions affect different stakeholders (e.g., patients, colleagues).
- AI systems broader implications for labor processes (see Pfeiffer 2014), work practices, clinical outcomes, and long-term societal consequences.

This leads us to a capacity-based perspective: Employees and especially professionals must not only possess individual skills, competences, and experience, but also work under organizational conditions that enable them to use and develop these. In short, they are individual, institutional and relational. They must be supported by organizational governance structures that ensure transparency, feedback, discretion, and participation. In this regard, several key concepts have to be clarified:

Ability refers to a person's actual or perceived skill to perform a specific task – for instance, interpreting AI systems' outputs. *Capacity*, by contrast, refers to the real freedom to act in a way that aligns with one's values and goals. It includes the presence of enabling conditions such as time, support, and access to knowledge. In our context, capability is the broader and more relevant term, since it incorporates both individual skills and institutional support (also see Sen 1999).

While *trust* is a relational or emotional stance – one's willingness to rely on others – *trustworthiness* refers to whether a system or actor actually deserves trust. AI governance must focus not only on encouraging professionals to trust, but also on ensuring that AI systems and their governance structures are trustworthy, i.e., transparent, accountable, and ethically sound (also see Hurley et al. 2025; Rubeis 2024).

Competence is the demonstrated ability to perform a task effectively; *confidence* is one's belief in that ability. AI governance must support both – professionals need to know how to interact with AI systems, but also feel secure enough to question, validate, or reject their outputs (also see Donia et al. 2024).

Responsibilisation describes a governance logic in which responsibility is shifted to individuals – such as clinicians – without giving them the resources or control necessary to act effectively. In the context of AI systems, professionals may be held accountable for decisions influenced by systems they did not design, understand, or fully control (also see Bleher and Braun 2022).

This conceptual foundation allows us to better understand why it is essential for trustworthy and effective AI governance to not only locate the responsible use of AI systems at the individual level, but to structurally empower users and (at best all) stakeholders to do so – especially in settings where decisions impact not only individual employees, but entire professional and organizational systems. The following sections build on this basis to outline how empowerment can be structurally embedded in governance and how such embedding shapes the evolving human–AI relationship. As a paradigmatic field of application for various reasons already mentioned and still to be shown, we continue to illustrate the present topic at appropriate points with work and organizations in the healthcare sector.

3. Structural Empowerment Theory and AI Governance

3.1. Structural Empowerment Theory

From social science research into technology and digitalization over the past 40 years we know that the introduction and use of new technologies is not a purely technical process, but always socially embedded (Pinch and Bijker 1984). Technical artefacts are charged with expectations, interests and meanings by various stakeholders, for example management, technology companies, employees, service

providers, politicians, etc. The concrete design and use of technology are therefore the result of a social negotiation process.

However, this process is not free of hierarchies and power asymmetries. This applies at a global and social level as well as at an organizational level. Organizational power relations are changing as a result of the implementation and use of these technologies, existing inequalities, e.g., within the workforce, are reinforced or new ones created (Huws 2014; Schrape 2021).

Many studies show that the actual use of new technologies often deviates from originally intended goals (originally Suchman 2007). Employees develop so-called 'workarounds' or use tools creatively to solve everyday work problems. Digital technologies are thus not merely 'implemented', but 'integrated' into everyday work practices. This means that the implementation in this practice should sensibly leave room for user-side appropriation and influence on implementation and further development. A participatory and context-sensitive implementation of digital technologies assumes that all those involved are able to participate in this process. This requires empowerment and an extensive recourse to the employees' labor capacity (Boes et al. 2020; Pfeiffer 2014).

This also applies to the current discussion on AI and work, where most debates on digitalization research are being continued, e.g., on questions of inequality, the automatability and replaceability of workers, qualification requirements, co-determination and design issues, new forms of work and tasks or the evolvement of new occupational fields. More recent aspects are the increased interest in ethically orientated questions, the penetration of the functional mechanisms and the question of a (re-)distribution of responsibility.

As we focus on the organizational level and issues relating to the design of AI-related governance, this does not mean leaving out the external institutions and actors that play a role in this process, such as legislation, professional and trade associations, or technology companies, but rather looking at their role in governance in organizations from the internal perspective of the organization. The concern to point out which aspects strengthen the role of employees leads us to the question of how structures of empowerment and the development of important capacities can be analyzed and assessed.

This is where structural empowerment theory comes into play. At the time, Kanter raised the question – relatively independently of technical developments – of how the structural inequality of employee groups (in this case men and women) could be offset by systematic support from the organisation and how this could ensure greater and more equal participation in developments within the organization (Kanter 1993).

Structural empowerment refers to the organizational conditions that enable employees to access essential resources, information, support, and opportunities necessary for professional autonomy and decision-making (Kanter 1993). In the

context of AI integration, structural empowerment is crucial for ensuring that professionals can effectively engage with AI systems, maintain trust in their outputs, and make informed decisions. Thus, empowerment can be understood as the process of equipping employees with the necessary knowledge, tools, and institutional support to confidently interact with AI systems, critically assess their recommendations, and make informed decisions that align with ethical, professional, and organizational standards (Thomas and Velthouse 1990). Kanter's framework identifies four key dimensions of structural empowerment:

1) Access to information: information and knowledge necessary to perform tasks
2) Access to resources: assets in terms of money, material and working time
3) Access to support: reflecting own work practices by receiving guidance and feedback from colleagues and supervisors
4) Access to opportunities: learning opportunities to allow for knowledge and skills growth.

Our conceptual basis adopts these four dimensions as its guiding theoretical framework. We argue that structural empowerment is essential for employees to develop both the confidence needed to assess the trustworthiness of AI system usage and the capacity to deal with AI systems in a useful way. For example, access to information helps employees understanding AI systems' capabilities and limitations, enabling critical evaluation of outputs. Access to resources provides the necessary tools and time to engage with AI systems responsibly. Access to support through feedback from colleagues and supervisors fosters collaborative problem-solving and continuous improvement. Finally, access to opportunities ensures employees stay updated on technological advancements, adapt to emerging ethical and legal challenges and rethink the way of using AI systems.

The integration of AI systems into decision-making introduces the need for new capacities among professional knowledge workers, including physicians and managers. For instance, physicians must develop the ability to understand and explain AI-driven decisions to patients, discern when to trust AI systems' recommendations, and recognize situations where human judgment should override AI systems' output (Zhang et al. 2023).

In the medical field, the use of AI-driven diagnostic tools illustrates how the demand for professionals has shifted. For example, consider AI systems that analyze medical imaging, such as X-rays or MRIs. In the past, radiologists relied heavily on their ability to manually assess each image, leveraging their pattern recognition skills honed through years of study and experience. The process required careful attention to detail, deep expertise in imaging interpretation, and an understanding of subtle variations that indicate diseases. Today, AI systems can process and analyze large volumes of medical images rapidly, often highlighting potential issues with

high accuracy (Smith and McKeen 2011). While this can reduce the time required for diagnosis, it also can shift the necessary skills for physicians. Instead of focusing solely on manual interpretation, healthcare professionals now need to: (1) be able to interpret AI-generated insights critically and validate these recommendations against their clinical judgment; (2) collaborate effectively with AI systems, which requires competencies in integrating machine-generated insights into treatment planning, ensuring seamless communication across medical teams, and maintaining adherence to clinical protocols; (3) assess potential biases in AI systems, ensure compliance with ethical standards, and prioritize patient safety.

As AI systems continue to evolve, the emphasis in healthcare shifts from manual diagnostic expertise to critical data interpretation, collaborative workflows, and ethical decision-making. But with this shift comes the need to ensure that AI systems can complement human expertise while maintaining high clinical standards and ensuring patient safety. These challenges underscore the need to empower professionals effectively to navigate this new environment – and to take part in its shaping.

Using SET as a guiding framework, the following requirements per SET dimension emerge based on the current state of our argumentation in order to enable empowerment in the context of AI systems in healthcare:

1) Access to information means that healthcare professionals and managers need clear, comprehensive, and transparent insights into how AI systems function. This includes understanding underlying algorithms, data inputs, and decision-making processes of AI to explain them effectively and make informed choices.
2) Access to resources: To work effectively with AI, healthcare professionals require resources such as state-of-the-art AI systems, adequate time to engage in AI-related training and decision-making, and institutional/organizational support for implementing AI systems.
3) For access to support continuous feedback and collaborative learning opportunities with colleagues, AI developers, and supervisors are critical. Support systems must be in place to help professionals evaluate AI-generated recommendations, troubleshoot issues, and improve their ability to work alongside AI systems.
4) Access to opportunities: Empowerment requires investment in ongoing education and training programs to help professionals and managers enhance their labor capacity. For example, workshops on interpreting AI systems' outputs, ethical usage of AI systems, and decision-making in complex scenarios can foster confidence and competence in leveraging AI systems effectively.

By structuring empowerment initiatives around these four dimensions, organizations may enable professionals to remain or become capable of acting, designing, and adapting to challenges and opportunities presented by AI systems in high-re-

liability decision-making environments. This approach might ensure that they can make the most of AI systems while safeguarding the integrity of their decisions, actions, and work in general.

3.2. The role of Structural Empowerment in AI Governance

As the integration of AI into healthcare goes far beyond technological change, broader questions need to be asked and implications considered and addressed. Ultimately, there are even fundamental questions of trust not only in the technology itself, but also in the organizations using it and the actors involved with their interests, knowledge, motives, scope for decision-making and action, skills and much more. And ultimately, it is also about trustworthiness at various levels and the ability to assess the trustworthiness of technical and organizational systems (at all). This concerns both patients in the healthcare sector as well as professionals from different disciplines who are (directly or indirectly) affected by the use of AI systems.

While AI systems promise enhanced efficiency, their implementation simultaneously introduces new forms of standardization and formalization, shifts in professional identity, autonomy, and (the need for) changing governance structures (Bartsch et al. 2025; Jones et al. 2023; Jungtäubl 2024). AI governance actively shapes how AI systems are embedded in clinical environments, determining who has access to knowledge, who retains decision-making authority, and – depending on the concrete governance design, directly or indirectly or intentionally or unintentionally – how professional roles change or evolve. Moreover, the tension between AI-driven neo-taylorisation of work[2] on the one hand, and increasing work intensification and rising job complexity on the other, must also be considered – particularly in healthcare settings (Altenried 2020; Mantello et al. 2023; Söllner et al. 2025).

Drawing on SET outlined before, this section explores how AI governance frameworks influence professionals' access to information, resources, support and opportunities – key dimensions relevant to their autonomy and engagement, and ultimately their ability to assess the trustworthiness of AI systems.[3] AI governance, however, is neither neutral nor purely technical. It can shape professionals' experiences with AI systems and their trust in organizational structures that (should) oversee the implementation of AI systems and not outsource it to employees choices or

2 Neo-Taylorisation of work refers to the contemporary revival and intensification of Taylorist practices – e.g., strict task decomposition, AI-driven surveillance and algorithmic control – often implemented at the expense of employee discretion and autonomy to maximize efficiency.

3 Ultimately, the employees, who in turn are in direct contact with the organizational environment – the patients – also represent trustworthiness to the outside world. They are therefore under particular pressure.

primarily to them with insufficient support.[4,5] If we now additionally extend SET through the lens of labor processes and subjectification, we can understand how AI systems influence the self-perception of professionals, their identity and the risks associated with increasing algorithmic accountability, and what impact this ultimately has on the (perceived) capacities of employees.

3.2.1. AI Governance as Structuring Force for Work, Trust(worthiness), and Empowerment

In the healthcare sector, issues relating to trust are not only about trust in technical systems such as AI reliability, but also about the implementation of embedding in and thus overarching trust in "the healthcare system", the organization of a hospital, etc. This is closely linked to governance structures that enable professionals to practically use AI in a meaningful way. "Meaningful" here implies improved productivity, enhanced work quality, healthy working conditions, self-efficacy, and professional identity. Such factors are also basic conditions for establishing and maintaining trustworthiness of AI (Kiseleva et al. 2022; Petersson et al. 2022; WHO 2021). Following SET, empowerment for 'good' AI governance[6] occurs when professionals are actively involved and have capacities to do so. Such active involvement and participation are strongly linked to SET dimensions. Professionals (and other employees as well) need transparency, explainability and contextualization in order to be able to interpret AI systems' results in a competent manner. Otherwise, there is a risk that AI governance will create a "black box bureaucracy" (Ananny and Crawford 2018; Pasquale 2015) that undermines trust and professional commitment. The active *information* of professionals is therefore essential in the context of governance and must be regulated, whereby employees should at best already be involved in the development of AI governance in order to embed their own aspiration in form of needs, knowledge and experience. It is further important to maintain and improve the *resources* of professionals by providing them with adequate training, sufficient time for professional learning and solid technical and organizational support as necessary prerequisites for enhancing their labor capacity that enables maintaining and developing AI-related competencies. AI governance structures must explicitly support these resources to reach the goal of strengthening trust, trustworthiness, professionalism, and identity. This is necessary to enable other positive effects of and for mutual learning between healthcare professionals and AI developers as well as between professionals and AI systems themselves.

4 According to initial, explorative empirical impressions from the DigiGov project interviews, the latter appears to be the case.

5 Ergo, questions for and perspectives on digital/AI ethics arise as discussed, e.g., Jungtäubl, Zirnig and Ruiner (2024).

6 In analogy to "good work" as a normative framing of working conditions.

Healthcare professionals need the *support* of their organization and other institutions (such as professional societies in the medical field, etc.) in the context of professional empowerment, for example in form of continuous collegial support and exchange as well as feedback mechanisms to the organization and to AI-developing companies, collaborative interactions and clearly regulated and jointly defined responsibility structures. AI governance must also prevent algorithmic fatigue (Budhwar et al. 2023) by maintaining supportive human interactions rather than replacing them with purely algorithmic controls. Ultimately, such structures and mechanisms also encourage developments around AI to promote, rather than restrict, career and professional development (*opportunities*), allowing professionals to develop new capacities through continuous, supported skill-building and co-creation processes (Jiang et al. 2021). Thus, AI governance actively shapes both the professional and organizational trust necessary for successful AI integration, enabling professionals to – finally – confidently and autonomously assess AI trustworthiness.

3.2.2. Balancing Trust, Autonomy, and Algorithmic Control

Subsequently to those considerations about effective AI governance in the medical field respectively in healthcare, three critical tensions must be carefully balanced in order to provide professional empowerment.

- *Autonomy vs. Algorithmic Standardization:* AI systems may enhance autonomy by reducing repetitive tasks; however, excessive reliance on AI systems' recommendations can undermine professional judgment and autonomy. And, autonomy can be reduced by the programming of the systems as they allow some cases and restrict others. AI governance frameworks must ensure that AI systems remain advisory rather than prescriptive or predictive, maintaining professional human agency in clinical decisions. It is also important to note that experienced staff seem less likely to use AI systems anyway and, even when they do, are good judges of how far they want to rely on AI systems' outputs, whereas less experienced staff are more likely to do so and, precisely because of their lack of experience, are less able to judge how good the quality of AI systems' outputs are. These and other points therefore have an impact on autonomy and the scope for action.
- *Competency Enhancement vs. De-skilling:* AI systems shift professional roles towards interpretative oversight. Without structured empowerment and ongoing training, professionals risk experiencing de-skilling, diminishing confidence in their own expertise and AI systems. Governance must thus emphasize capacity-based approaches, ensuring AI systems complement rather than replace professional judgment, and must address the question of what degree of dependency on those systems is acceptable.

- *Trust in AI vs. Trust in Governance:* Professionals' trust in AI systems depends significantly on transparent, accountable and participatory AI governance. Even high-performing AI systems may be rejected if AI governance lacks transparency and participatory mechanisms. AI governance structures must therefore include clear accountability processes, participatory design approaches, and explicit dispute-resolution mechanisms. Ultimately, this not only leads to a potentially improved identification with the organization and its AI governance, but increases the chance of acceptance and appropriation as well as the quality of reliance on new technology such as AI.

These tensions emerge differently depending on the complexity and apparent 'autonomy' of AI systems, which highlights the need for flexible governance approaches that are tailored to the different levels of AI integration and are continuously reviewed and realigned where necessary – from narrow to (highly) advanced AI systems.

3.2.3. Subjectification of Work, Control, and AI Governance

AI governance also directly influences professionals' subjective experiences of work, affecting self-perception, professional identity, and accountability structures. On the one hand, the introduction of AI can paradoxically lead to increased demands for autonomy (professionals interpret and integrate AI systems into work processes). On the other hand, it also harbors the risk of new forms of control and heteronomy (e.g., through algorithmic performance monitoring, prescriptive standardization). This risk sometimes arises precisely from the desired transparency, which can extend beyond the technical systems to work processes in which AI systems are embedded.

Without structural empowerment, professionals risk "algorithmic responsibilization", becoming accountable for AI systems' outcomes without decision-making authority (Adensamer et al. 2021). This leads to increased cognitive burdens, identity erosion, and challenges in overriding AI systems' recommendations – especially under economic pressures prevalent in healthcare. To counteract these negative subjectification effects, AI governance must prioritize:

- *Human-in-the-loop models:* Maintaining genuine human oversight, reducing performative validations of AI systems' outputs.
- *Co-design and participatory governance processes:* Ensuring professionals actively shape AI systems development and integration, aligning AI systems with real-world work practices.
- *Explicit accountability structures:* Clarifying responsibility for AI systems' decisions, preventing unintended accountability shifts onto frontline professionals.

4. Towards "Good" AI Governance: A Capacity-Based Approach

A genuinely empowerment-oriented AI governance model ensures that AI systems support professional autonomy, capacity-building, and trust rather than imposing rigid control or labor intensification. Key principles include:

- *Transparency and explainability:* Ensuring AI interpretability so professionals trust AI systems because they understand them, not merely due to external expectations.
- *Autonomy-preserving AI design:* AI systems as assistive technology, reinforcing clinical discretion rather than enforcing algorithmic compliance.
- *Continuous skill development and organizational learning:* Investing in continuous competence-building to develop genuine trust based on professional skill, not blind acceptance.
- *Participatory AI governance:* Involving professionals in governing AI systems actively ensures context-sensitive governance aligned with professional needs, not abstract managerial or purely technological priorities.

Embedding these principles into AI governance can create a trust-based work environment where AI reinforces rather than undermines professional expertise and autonomy.

Table 1: Structural empowerment from the perspective of AI governance.

SET Dimension: Access to…	**AI Governance Considerations**	**Impact on work, trust(worthiness), and professional identity**
Information	· Transparency regarding technical functionalities and organisational decision-making processes (*governance transparency*) · AI system explainability and interpretability of results · Clear disclosure of data provenance and decision criteria underlying AI systems' outputs	· Trust increases when professionals can autonomously assess and understand AI systems' outputs. · Enhances self-efficacy and reduces risks of "black-box bureaucracy." · Prevents risks of algorithmic responsibilization (*accountability without meaningful authority*).

SET Dimension: Access to...	**AI Governance Considerations**	**Impact on work, trust(worthiness), and professional identity**
Resources	· Investments in continuous professional development and labor capacity · Provision of sufficient time, technical infrastructure, and organisational resources · Supporting individuals in developing necessary competencies for effective AI system use and oversight	· Enables meaningful integration of AI systems into daily practices, preventing overload and de-skilling. · Builds trust as professionals possess adequate capacities and resources for safe, responsible, and efficient AI system use. · Supports healthy, sustainable, and professionally enriching working conditions.
Support	· Establishing clear accountability structures and responsibilities for AI-based decisions · Fostering collaborative and participatory AI governance models (*co-design*, *human-in-the-loop*) · Institutionalized support mechanisms (*peer networks*, *feedback systems*) to avoid algorithmic isolation and fatigue	· Reduces algorithmic isolation and cognitive burdens through collective reflection, discussion, and validation of AI recommendations. · Increases organisational trust and prevents negative subjectification effects (e.g., over-responsibilization, identity loss). · Ensures sustainable, human-centric interaction with AI systems.

SET Dimension: Access to…	AI Governance Considerations	Impact on work, trust(worthiness), and professional identity
Opportunities	· Promoting continuous development of AI competencies as integral to professional career paths and work environments · Leveraging AI systems as an enhancement (not a replacement) of professional expertise and decision-making autonomy · Creating new professional roles and opportunities through co-design processes and active professional involvement in AI system implementation	· Supports positive perception of AI systems as extending and strengthening professional competencies rather than restricting them. · Avoids negative subjectification effects (role degradation, loss of autonomy and professional discretion). · Reinforces professional identity and autonomy by enabling active shaping and meaningful integration of AI systems into professional roles.

4.1. Conceptual Framework: AI Governance, Structural Empowerment, and Professional Outcomes

The conceptual framework depicted in Figure 2 provides a structured overview of the core theoretical relationships addressed in this article. Building on previous discussions, it illustrates how organizational AI governance directly influences both organizational capacities and individual employee capacities. Central to this relationship is the SET, which serves as a critical mediator between governance structures and individual outcomes, particularly the (perceived) ability to evaluate trustworthiness of AI system usage.

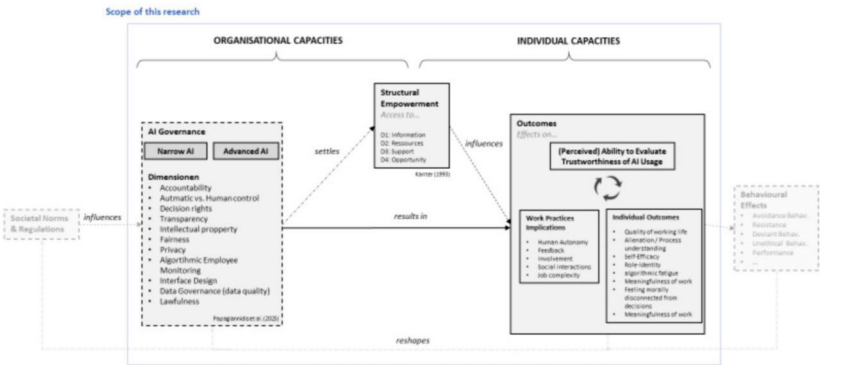
At the organizational level (left side of Figure 2), AI governance structures encompass distinct dimensions such as accountability, decision rights, transparency, fairness, privacy, and data governance. These governance dimensions vary according to concrete regulations or structures within organizations as well as according to whether the AI systems are narrowly defined or more advanced and (can be) used more extensively. The interplay of these AI governance dimensions shapes the organizational conditions that professionals encounter, defining their access to crucial resources for effectively integrating AI systems into their work. The four SET dimensions provide a direct pathway through which organizational AI governance

conditions influence individual capacities, shaping both practical and psychological outcomes at the workplace.

On the individual level (right side of Figure 2), the outcomes are categorized into two interconnected areas: implications for labor processes and work practices as well as subjective, psychological effects. Work practices include dimensions such as autonomy, complexity, social interactions, and meaningfulness of tasks. Individual outcomes on employees, as subjects, cover aspects like psychological ownership, role identity, perceived legitimacy of AI systems, trust, and potential negative effects such as algorithmic fatigue or feelings of over-responsibilization. Collectively, these individual outcomes determine the ability of professionals to competently assess the trustworthiness of AI systems – an essential competency increasingly demanded by regulatory frameworks such as the EU AI Act (EU, 2024). Furthermore, the conceptual framework acknowledges external influences, such as societal norms and regulations, and behavioral effects like avoidance behaviors, resistance, or deviations from prescribed AI system usage, indicating the broader social and organizational implications of AI governance decisions.

In sum, the conceptual framework clearly illustrates that AI governance is not merely a technical consideration but an essential structural dimension shaping professional roles, identities, and trust relationships within organizations. By leveraging SET, it is, on the one hand, an analyzation-tool and, on the other hand, it offers concrete pathways to actively design governance frameworks that empower professionals and foster meaningful human-AI collaboration (Chowdhury et al. 2022).

Figure 2: Conceptual Framework.



4.2. Medical AI Scenarios

Now we present two hypothetical scenarios. The first scenario focuses on a medical setting where AI-powered software assists radiologists by analyzing medical imaging, helping to identify potential health concerns like tumors. The second scenario shifts to the perspective to advanced AI systems in triage decisions. By examining these two scenarios, we will demonstrate these relationships concretely, highlighting specific governance challenges and empowerment strategies in different clinical contexts. The interplay of trust, empowerment, and subjectification varies substantially depending on the scope of the respective AI system, so the following section illustrates (some) distinctions concretely through two contrasting scenarios, demonstrating why adaptive, empowerment-oriented AI governance – grounded explicitly in SET – is essential. Such governance does more than manage technological risks; it actively shapes professional roles, work experiences, and trust in AI system integration.

4.2.1. Scenario – Narrow AI: Cancer screening and professional autonomy

Dr. Anna Berger, an experienced radiologist at a leading hospital, starts her working day assisted by an AI-powered diagnostic tool designed to enhance cancer screening accuracy. The AI system is designed to analyze MRI and CT scans, identifying potential tumors to improve diagnostic accuracy. It utilizes deep learning algorithms trained on millions of real-world medical images.

As the AI system delivers its initial analysis results, Dr. Berger notices that the AI has flagged a suspicious lesion – something she might not have noticed at first glance. She manually reviews the findings and confirms the need for further investigation. While the AI automatically detects standardized patterns in routine cases, it is up to Dr. Berger to assess complex and unusual cases, consider individual patient factors, and ultimately determine the final diagnosis and next medical steps.

This division of working tasks (and responsibilities) between human and AI illustrates how AI systems can effectively support professionals by relieving them of routine analytical tasks, thus allowing clinicians to focus on more challenging cases. However, it also increases job complexity, as Dr. Berger must now manage the intersection of her clinical judgment with the AI generated recommendations. Furthermore, the phenomenon of algorithmic opacity emerges: Dr. Berger understands that the AI system flags certain images, but she may not fully comprehend why minor variations in patient data alter AI decisions, reflecting pragmatic opacity as discussed in recent research (Bhat 2025; Shahidi Hamedani et al. 2025).

From a SET perspective, the success of such AI system integration critically depends on governance structures that ensure transparency and continuous vocational training. Dr. Berger requires targeted training, sufficient resources, and organizational support to confidently interpret, validate, and sometimes override

AI recommendations. Without clear governance guidelines, she risks experiencing heightened cognitive burdens and algorithmic responsibilization – where accountability for diagnostic outcomes is placed upon her, even as her ability to influence the underlying logic of the AI system remains limited.

Previous studies support the realism of this scenario. Eisemann et al. (2025), for instance, found that AI-assisted mammography significantly improved cancer detection rates without increasing false positives, highlighting AIs' potential to augment professional expertise without replacing it. Yet, to achieve such positive outcomes sustainably, AI governance frameworks must prioritize transparent and explainable AI, structured feedback loops, and the active participation of medical professionals in AI system development.

This scenario underscores that the perceived ability to evaluate the trustworthiness of AI systems is not merely a function of individual technical expertise, but emerges at the intersection of personal competence, institutional support, and transparent AI system design. For professionals like Dr. Berger, structural empowerment – including targeted training, resource provision, and participatory (AI) governance mechanisms – constitutes a critical precondition for developing the evaluative confidence and contextual understanding required to responsibly interpret and, where necessary, contest AI-generated outputs.

4.2.2. Scenario – Advanced AI: Automated Emergency Triage

Dr. Jonas Meier, an experienced emergency physician, begins his shift at one of the busiest emergency departments in the city. Recently, an (advanced) AI system has been assisting physicians by accessing real-time digital patient records, current hospital data, societal health statistics, and news. Its main purpose is to support triage decisions: Unlike narrow AI that operate within predefined boundaries, this AI system dynamically pulls from multiple real-time data sources to assess case urgency, prioritize patients, and allocate resources more efficiently.
The AI is capable of independently and flexibly deciding which data is relevant and which is not. It can also automatically 'decide' which tools to use in problem-solving. The AI system can independently calculate a strategy and weigh the decisions that need to be made based on that. For Jonas the AI suggestions seem to be useful, but he never knows exactly which information and tools the AI considers and which it disregards.

What sets this AI system apart is its high degree of automation: based on calculations it independently determines relevant data, analytical tools, and appropriate weighting of competing priorities. While Dr. Meier often finds its suggestions helpful, the logic behind these outputs remains opaque – raising concerns about transparency, particularly in ethically complex decisions such as prioritizing between equally urgent patients with different prognoses or resource needs.

This scenario reveals the critical importance of AI governance, particularly when advanced AI systems influence not only clinical workflows but also ethical judgment and institutional authority. In the absence of structured oversight or in-

terdisciplinary governance bodies, clinicians like Dr. Meier are placed in vulnerable positions – legally and morally accountable, yet without real visibility into how AI outputs are calculated. This "algorithmic responsibilization" risks eroding both trust and professional autonomy. From the perspective of SET, technical reliability alone is insufficient. What is needed are institutionalized structures for explainability, clinician participation, and dynamic feedback. (AI) Governance frameworks must clearly delineate accountability lines for decision outcomes – especially when AI recommendations are accepted or overridden. Furthermore, professionals must be provided with systematic channels to shape AI system development, contribute to risk evaluation, and evaluate emerging patterns of bias or opacity.

The trustworthiness of advanced AI systems depends not only on the accuracy of these systems but also on the institutional scaffolding that ensures their contextual use. Trust arises from meaningful engagement – AI governance that ensures transparent AI integration, preserves professional discretion, and embeds human oversight as a non-negotiable standard. In this context, the perceived ability to evaluate the trustworthiness of AI (usage) becomes both more fragile and more essential. For professionals like Dr. Meier, this ability hinges not only on cognitive competencies or ethical awareness, but fundamentally on whether (AI) governance structures enable transparent insight, facilitate critical reflection, and foster a climate of shared responsibility in navigating opaque and high-reliability AI outputs.

4.3. Perceived Ability to Evaluate Trustworthiness of AI Usage

In line with the EU AI Act, professionals must not only follow AI recommendations – they must be capable of critically evaluating them. This calls for the development of user impact awareness – an emergent competence involving the capacity to foresee how AI system decisions impact patients, professional workflows, and wider organizational outcomes.

For Dr. Meier, this is particularly relevant. The advanced AI system automatically filters and interprets vast data sets but provides little insight into its 'decision rationale'. This opacity becomes problematic when prioritization outcomes are potentially biased – for instance, if patients with more complete digital histories or from wealthier neighborhoods are systematically favored. Such patterns, if undetected, can solidify through feedback loops, reinforcing disparities. Assessing trustworthiness under these conditions requires time, support, and structured empowerment:

- Explainability tools must be embedded in the AI system to reveal key decision pathways.
- Institutional mechanisms for interdisciplinary validation and ethical review are essential.

- Professionals must be granted both the time and training to engage critically with AI system outputs.

Advanced SET provides a powerful lens here: access to information (here, e.g., how the AI system works), resources (training, time), support (interdisciplinary exchange), and opportunities (participation in governance) are essential for cultivating capacity to assess AI system trustworthiness. Only through such holistic empowerment can healthcare systems ensure just, effective, and ethically grounded AI integration.

5. Conclusion

This article has examined the essential role of structural empowerment in fostering the ability to evaluate the trustworthiness of AI systems, particularly within healthcare settings. Through the lens of SET, we explored how the integration of AI systems transforms work practices, professional roles, and governance structures. Our analysis shows that while AI systems can greatly enhance decision-making efficiency, they also pose new challenges.

Successfully interacting with AI systems requires labor capacity and competencies that differ from those needed for traditional information systems. A key ability is to evaluate AI systems' trustworthiness. However, we demonstrated that this ability is not solely a matter of individual competence but also depends on organizational conditions that support vocational training, access to information, collaborative practices, and participatory governance. Using scenario-based analysis, we illustrated how narrow AI and advanced AI systems differently impact healthcare professionals' ability to make informed and responsible decisions.

Our analysis also shows that the degree of 'agency' – understood as automated and independent functioning, calculating, and 'deciding' – of the AI system significantly influences the individual capacities required to assess the trustworthiness of AI. As AI systems become more advanced and workflow automation increases, the complexity of evaluating their outputs grows, making it increasingly difficult – even impossible in some cases – for individual healthcare professionals to assess trustworthiness on their own. Therefore, AI governance should establish structures that distribute the responsibility for evaluating trustworthiness rather than placing it solely on individuals. This approach not only mitigates risks associated with increasing system autonomy but also fosters a more collaborative and supportive evaluation process.

The findings suggest that successful AI integration requires more than just technical training; it necessitates a comprehensive approach that embeds empowerment at multiple levels of the organization. Empowering healthcare professionals through

transparent governance, continuous learning opportunities, and active involvement in AI governance fosters both individual and collective capacities. Thus, AI governance should not only focus on regulating technology but also on shaping the human-AI interaction in ways that reinforce professional autonomy, competence, and trust.

To achieve this, healthcare professionals need support from their organizations in the form of access to empowered structures. Practically, this means establishing new learning and design processes that enable ongoing skill development and adaptation. Digital governance should also focus on creating communication channels and reflection spaces that connect individuals and organizations, such as intranets or professional networks for continuous exchange and knowledge sharing. Designing these structures appropriately will be a core task of digital governance in the future.

Ultimately, the adoption of AI systems in healthcare must prioritize structural empowerment to ensure that professionals remain confident, capable, and accountable in their interactions with increasingly autonomous systems. This approach will enable organizations to harness the potential of AI while safeguarding human-centered decision-making and maintaining ethical standards in medical practice.

Nevertheless, our analysis shows that while structural empowerment is a promising approach to fostering skills and competencies in working with AI systems, it is not a universal solution. Creating structural conditions that ensure access to information, resources, support, and continuous development is crucial, but these measures alone cannot fully address all challenges related to AI use. In complex and dynamic environments, such as certain healthcare settings, even well-designed empowerment structures have their limitations. For example, the growing functions and functioning of advanced AI systems may continue to create uncertainties that structural measures alone cannot resolve. Moreover, continuously aligning individual competencies with rapid technological advancements remains a challenge. By embedding structural empowerment within a dynamic and adaptable governance framework, organizations can better equip healthcare professionals to navigate the evolving challenges posed by AI systems while ensuring professional autonomy, competence, and trust.

References

Adensamer, Angelika, Gsenger, Rita and Klausner, Lukas D. (2021): "'Computer says no'": Algorithmic Decision Support and Organisational Responsibility", in: Journal of Responsible Technology 7, 100014.

Altenried, Moritz (2020): "The Platform as Factory. Crowdwork and the Hidden Labour Behind Artificial Intelligence", in: Capital & Class 44(2), pp. 145–158.

Amershi, Saleema et al. (2019): "Guidelines for human-AI interaction", Proceedings of the 2019 chi conference on human factors in computing systems, pp. 1–13.

Ananny, Mike, Crawford, Kate (2018): "Seeing Without Knowing. Limitations of the Transparency Ideal and its Application to Algorithmic Accountability", in: new media & society 20(3), pp. 973–989.

Baird, Aaron and Maruping, Likoebe M. (2021): "The Next Generation of Research on IS Use. A Theoretical Framework of Delegation To and From Agentic IS Artifacts", in: MIS quarterly 45(1).

Barocas, Solon and Selbst, Andrew D. (2016): "Big data's disparate impact", in: Calif. L. Rev. 104, p. 671.

Bartsch, Sebastian et al. (2025): "Governance of High-Risk AI Systems in Healthcare and Credit Scoring", in: Business & Information Systems Engineering, pp. 1–19.

Berente, Nicholas et al. (2021): "Managing Artificial Intelligence", in: MIS quarterly 45(3).

Bhat, Maalvika (2025): "Designing AI Interfaces for Transparent Decision-Making and Ethical Reflection", Companion Proceedings of the 30th International Conference on Intelligent User Interfaces, pp. 211–214.

Bleher, Hannah and Braun, Matthias (2022): "Diffused Responsibility. Attributions of Responsibility in the Use of AI-driven Clinical Decision Support Systems", in: AI and Ethics 2(4), pp. 747–761.

Boes, Andreas et al. (2020): "Empowerment in der agilen Arbeitswelt", in: Bauer et al. (eds.), Arbeit in der digitalisierten Welt. Praxisbeispiele und Gestaltungslösungen aus dem BMBF-Förderschwerpunkt, Springer, p. 307.

Budhwar, Pawan et al. (2023): "Human Resource Management in the Age of Generative Artificial Intelligence. Perspectives and Research Directions on ChatGPT", in: Human Resource Management Journal 33(3), pp. 606–659.

Cetindamar, Dilek et al. (2022): "Explicating AI literacy of employees at digital workplaces", in: IEEE transactions on engineering management 71, pp. 810–823.

Chan, Alan et al. (2023): "Harms From Increasingly Agentic Algorithmic Systems", Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pp. 651–666.

Cheong, Ben C. (2024): "Transparency and Accountability in AI systems. Safeguarding Wellbeing in the Age of Algorithmic Decision-making", in: Frontiers in Human Dynamics 6, 1421273.

Chowdhury, Soumyadeb et al. (2022): "AI-employee Collaboration and Business Performance. Integrating Knowledge-based View, Socio-technical Systems and Organisational Socialisation Framework", in: Journal of Business Research 144, pp. 31–49.

DataRobot (2024): "The Unmet AI Needs Survey", https://www.datarobot.com/resources/the-unmet-ai-needs-survey/?utm_campaign=unmetaineedssrvyDBPR&utm_source=database&utm_medium=direct, last access: July 17, 2025.

Dietvorst, Berkeley J., Simmons, Joseph P. and Massey, Cade (2015): "Algorithm Aversion. People Erroneously Avoid Algorithms After Seeing Them Err", in: Journal of experimental psychology: General, 144(1), p. 114.

Dlugatch, Rachel, Georgieva, Antoniya and Kerasidou, Angeliki (2023): "Trustworthy Artificial Intelligence and Ethical Design. Public Perceptions of Trustworthiness of an AI-based Decision-support Tool in the Context of Intrapartum Care", in: BMC Medical Ethics 24(1), p. 42.

Donia, Joseph et al. (2024): "Lifecycles, Pipelines, and Value Chains. Toward a Focus on Events in Responsible Artificial Intelligence For Health", in: AI and Ethics, pp. 1–14.

Eisemann, Nora et al. (2025): "Nationwide Real-world Implementation of AI For Cancer Detection in Population-based Mammography Screening", in: Nature medicine, pp. 1–8.

EU (2024): "The EU Artificial Intelligence Act: Up-to-date developments and analyses of the EU AI Act", https://artificialintelligenceact.eu, last access: July 17, 2025.

Fügener, Andreas et al. (2022): "Cognitive Challenges in Human–artificial Intelligence Collaboration. Investigating the Path Toward Productive Delegation", in: Information Systems Research 33(2), pp. 678–696.

Hurley, Meghan E. et al. (2025): "Patient Consent and the Right to Notice and Explanation of AI systems Used in Health Care", in: The American Journal of Bioethics 25(3), pp. 102–114.

Huws, Ursula (2014): Labor in the Global Digital Economy. The Cybertariat Comes of Age, NYU Press.

Jiang, Lushun et al. (2021): "Opportunities and Challenges of Artificial Intelligence in the Medical Field. Current Application, Emerging Problems, and Problem-solving Strategies", in: Journal of International Medical Research 49(3), pp. 1–11.

Jones, Caroline, Thornton, James and Wyatt, Jeremy C. (2023): "Artificial Intelligence and Clinical Decision Support. Clinicians' Perspectives on Trust, Trustworthiness, and Liability", in: Medical law review 31(4), pp. 501–520.

Jungtäubl, Marc (2024): Steuerung von Arbeit durch Formalisierung [Control of Work through Formalization], Nomos.

Jungtäubl, Marc, Zirnig, Christopher and Ruiner, Caroline. (2024): "HCI Driving Alienation. Autonomy and Involvement as Blind Spots in Digital Ethics", in: AI and Ethics 4(2), pp. 617–634.

Kanter, Rosabeth M. (1993): Men and women of the corporation. New edition. Basic books.

Kapoor, Sayash et al. (2024): "Ai agents that matter", arXiv preprint arXiv:2407.01502.

Kellogg, Katherine C., Valentine, Melissa A. and Christin, Angèle (2020): "Algorithms at Work. The New Contested Terrain of Control", in: Academy of management annals 14(1), pp. 366–410.

Kiseleva, Anastasiya, Kotzinos, Dimitris and De Hert, Paul (2022): "Transparency of AI in Healthcare as a Multilayered System of Accountabilities. Between Legal Requirements and Technical Limitations", in: Frontiers in Artificial Intelligence 5, pp. 1–21.

Lebovitz, Sarah, Lifshitz-Assaf, Hila and Levina, Natalia (2022): "To Engage or Not to Engage with AI for Critical Judgments. How Professionals Deal with Opacity When Using AI for Medical Diagnosis", in: Organization science 33(1), pp. 126–148.

Mantello, Peter et al. (2023): "Bosses Without a Heart. Socio-demographic and Cross-cultural Determinants of Attitude Toward Emotional AI in the Workplace", in: AI & SOCIETY 38(1), pp. 97–119.

Matys, Thomas (2025): "Entgrenzungen und Globalisierung", in: Abels et al. (eds.), Macht, Kontrolle und Entscheidungen in Organisationen. Eine Einführung in organisationale Mikro-, Meso- und Makropolitik, Wiesbaden: Verlag für Sozialwissenschaften, pp. 165–184.Springer.

Mitchell, Margaret et al. (2025): "Fully Autonomous AI Agents Should Not be Developed", arXiv preprint arXiv:2502.02649.

Murray, Alex, Rhymer, Jen and Sirmon, David G. (2021): "Humans and Technology. Forms of Conjoined Agency in Organizations, in: Academy of management review 46(3), pp. 552–571.

Pasquale, Frank (2015): "The Black Box Society. The Secret Algorithms that Control Money and Information, Harvard University Press.

Petersson, Lena et al. (2022): "Challenges to Implementing Artificial Intelligence in Healthcare. A Qualitative Interview Study With Healthcare Leaders in Sweden", in: BMC health services research 22(1), p. 850.

Pfeiffer, Sabine (2014): "Digital Labour and the Use-value of Human Work. On the Importance of Labouring Capacity for Understanding Digital Capitalism", in: tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society 12(2), pp. 599–619.

Pinch, Trevor J. and Bijker, Wiebe E. (1984): "The Social Construction of Facts and Artefacts. Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other", in: Social studies of science 14(3), pp. 399–441.

Ramezani, Ramin et al. (2025): "Bench to Bedside. AI and Remote Patient Monitoring", in: Frontiers in Digital Health 7, pp. 1–4.

Rubeis, Giovanni (2024): "Relationships", in: Ethics of Medical AI, pp. 151–212.

Samuel, Arthur L. (1959): "Machine learning", The Technology Review 62(1), pp. 42–45.

Schrape, Jan-Felix (2021): Digitale Transformation, Bielefeld: transcript/utb.

Schuetz, Sebastian and Venkatesh, Viswanath (2020): "The Rise of Human Machines. How Cognitive Computing Systems Challenge Assumptions of User-sys-

tem Interaction", in: Journal of the Association for Information Systems 21(2), pp. 460–482.

Sen, Ayusman (1999): Development as Freedom, Oxford University Press, New York.

Shahidi Hamedani, Sharareh, Aslam, Sarfraz and Shahidi Hamedani, Shervin (2025): "AI in Business Operations. Driving Urban Growth and Societal Sustainability", in: Frontiers in Artificial Intelligence 8, pp- 1–5.

Shneiderman, Ben (2020): "Bridging the Gap Between Ethics and Practice. Guidelines for Reliable, Safe, and Trustworthy Human-centered AI systems", in: ACM Transactions on Interactive Intelligent Systems (TiiS) 10(4), pp. 1–31.

Smith, Heather A. and McKeen, James D. (2011): "Enabling Collaboration with IT", in: Communications of the Association for Information Systems 28(1), p. 16.

Söllner, Matthias et al. (2025): "ChatGPT and Beyond. Exploring the Responsible Use of Generative AI in the Workplace. An Interdisciplinary Perspective", in: Business & information systems engineering, pp. 1–15.

Suchman, Lucille A. (2007): Human-machine Reconfigurations. Plans and Situated Actions, Cambridge university press.

Thomas, Kenneth W. and Velthouse, Betty A. (1990): "Cognitive Elements of Empowerment.: "An 'Interpretive' Model of Intrinsic Task Motivation", in: Academy of management review 15(4), pp. 666–681.

WHO (2021): Ethics And Governance of Artificial Intelligence For Health, https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf, last access: July 17, 2025.

Zhang, Melody et al. (2023): "The Adoption of AI in Mental Health Care–perspectives From Mental Health Professionals. Qualitative Descriptive Study", in: JMIR Formative Research 7(1), e47847.

Zou, James and Topol, Eric J. (2025): "The Rise of Agentic AI Teammates in Medicine", in: The Lancet 405(10477), p. 457.

Zuboff, Shoshana. (2023): "The Age of Surveillance Capitalism", in: Longhofer, Wesley and Winchester, Daniel (eds.), Social Theory Re-wired, Routledge pp. 203–213.