5-11-2023

Design, Development, and Implementation of Artificial Intelligence Technology: A Scoping Review

Imke Grashoff
University of Hamburg, imke.grashoff@uni-hamburg.de

Jan Recker
University of Hamburg, jan.christof.recker@uni-hamburg.de

DESIGN, DEVELOPMENT, AND IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGY: A SCOPING REVIEW

Research Paper

Grashoff, Imke, University of Hamburg, Hamburg, Germany,
imke.grashoff@uni-hamburg.de

Recker, Jan, University of Hamburg, Hamburg, Germany,
jan.christof.recker@uni-hamburg.de

Abstract

Research on artificial intelligence technology has not only increased rapidly over recent years, it is also no longer limited to the technical disciplines from which it originates. AI technologies are also at the center of social and sociotechnical studies including those conducted in information systems. Through a scoping review we explore how the research on different kinds of AI technologies has progressed over the past decade. Particularly, we explore whether the research on AI technologies has been informed by and informs our sociotechnical understanding of phenomena related to their design, development and implementation. For this purpose, we develop an analytical framework that differentiates sociotechnical perspectives, categorizes AI technologies into different kinds, and distinguishes research on AI technology design, development, and implementation. The findings from our review point to several directions for future research.

Keywords: Artificial Intelligence, Sociotechnical Perspective, Scoping Review.

1 Introduction

Artificial intelligence (AI) technologies, that is, digitally-embodied or -enabled agents that are able to perceive environmental data and perform actions based on these perceptions (Russell and Norvig, 2010), have permeated many areas of the human experience. In business alone, AI technologies have progressed from deterministic rule-based expert systems (Gill, 1995) to probabilistic decision-support systems (Pearl, 1988) and predictive models (Fu et al., 2021) that are now widely used in a variety of contexts, such as targeted advertising (Davenport et al., 2020), customer service (Schanke et al., 2021), robotic health care (Terry, 2019), algorithmic management (Möhlmann et al., 2021), operations and logistics (Tsolakis et al., 2022), or product development (Recker et al., 2023). By some estimates, more than half of businesses were implementing some form of AI technologies in 2020 (Balakrishnan et al., 2020).

With the increasing rate of development and diffusion, also research on AI technologies has grown rapidly, examining the advantages but also the critical sides of AI technology design and use. Because AI technologies are at least partially autonomous, learning, and inscrutable made-made objects (Berente et al. 2021), much research has focused on the technological challenges and advances in developing “machines who think” (McCorduck, 2004). But social consequences, especially unintended outcomes, such as discrimination, unfairness or otherwise unethical decision-making or behavior, have also been receiving increased attention because they denote potential threats that society and organizations must consider when incorporating AI technologies into regular practices (Taddeo and Floridi, 2018). For example, it is widely suggested that humans must stay “in the loop” when businesses implement or use AI technologies (Metcalf et al., 2019; Fügener et al., 2021; Murray et al., 2021; Raisch and Krakowski, 2021). Researchers also increasingly argue that human actors, such as developers and managers in charge of development, must be more aware and fully responsible already for the design, development and implementation of AI technologies to mitigate if not prevent already beforehand any unintended consequences of AI technologies during use (Taddeo and Floridi, 2018; Martin, 2019; Abbasi et al., 2018).

Maintaining an overview over the advancement of knowledge on the design, development and implementation of AI technologies across technical and social dimensions is difficult because research on AI has long ceased to be limited to computer science (McCorduck, 2004; Berente et al., 2021). Other fields such as neuroscience, psychology, sociology, philosophy, economics, and information systems have become interested in the phenomena that surround AI technologies, in particular their design, development and implementation, contributing different valuable knowledge with their different perspectives, methods and approaches (Russell and Norvig, 2010). But to date, the different contributions remain mostly separated, leading to a fragmented and scattered picture of relevant phenomena and issues (Sloane, 2022).

The time is ripe, therefore, to review the scope and nature of the current body of literature on AI technology design, development, and implementation to understand which knowledge contributions have been made, where, and how. Our goal is to provide a coherent summary of the different types of knowledge contributions that have been made about design, development and implementation of AI technologies as essential prerequisites for their deployment and use, and to understand which different types of AI technologies have received attention and from what perspective. This is important because such an overview paves the way for future research by highlighting the blind spots and inconsistencies as well as fruitful points of reference within current research. In doing so, we have three foci:

1. To cluster the literature on AI technology design, development and implementation along a continuum of sociotechnical perspectives from predominantly social to predominantly technical (Sarker et al., 2019). We chose this perspective because not only the processes of designing, developing and implementing are sociotechnical in nature but also the AI technology as the outcome of this processes is both a technological artefact and developed for use in social settings (Crowley, 2019). A sociotechnical perspective draws attention to both technical issues, such as performance or functionality, and social issues, such as bias or fairness (Sartori and Theodorou, 2022).
2. To distinguish the types of AI technologies in focus of the research conducted so far, in order to reflect about how this may influence what we actually know about this type of digital technology, its design, development and implementation. To that end, we will use the distinctions offered by Russell and Norvig (2010) to differentiate AI technologies that agents that can think or act humanly or rationally.
3. To examine which dimension of the technology lifecycle – from “design”, to “development” and “implementation” has received most attention in the literature.

Our review leads to several key findings: first, conceptual research still dominates empirical research on AI technologies. Second, most sociotechnical research does not balance emphasis on both social and technical elements that relate to AI technologies, rather, research either on the social or the technical aspects of AI technologies dominates. Third, most research to date focuses on AI technologies as agents that act humanly or think rationally. Less is known about how AI technologies could be designed, developed, and implemented that follow rational rather than human agency goals.

We proceed as follows: Section 2 provides a theoretical background, introducing the key concepts that serve us as basis for our scoping review. Section 3 explains the literature review strategy we followed. Section 4 presents our main findings. We discuss our contributions in section 5 and close with some conclusions in section 6.

2 Background

In the following, we provide an overview of three concepts our review is based on. These concepts serve us as analytical lens on the current body of knowledge within research on AI.

2.1 Artificial Intelligence Technologies as Agents that Can Think or Act

For the purpose of our research we follow Russell and Norvig (2010) and define AI technologies as man-made digital agents that perceive and act upon environmental data. We chose this definition because their work is considered the seminal authoritative reference on artificial intelligence and the one most widely used in university courses.

The definition highlights two key points. First, AI technologies have some level of independent agency (Leonardi, 2011) – they can autonomously perceive and in case also act based on some intelligent decision-making schema (Russell, 2019). Second, their intelligence is either concerned with thought processes in contrast to the behavior of intelligent entities that might either emulate human performance, or an ideal idea of intelligence, referred to as rationality.

On this basis, Russell and Norvig (2010) distinguish four basic kinds of AI technologies: First, AI technologies are defined as agents that think humanly. This perspective pertains to the construction of technological systems through cognitive modelling approaches that seek to build human reasoning into a program in order to design machines that are able to make decisions, solve problems or learn just like humans do.

Second, AI technologies are understood as agents that act humanly. This perspective refers to the construction of technological systems that are able to achieve a human-like performance in various cognitive functions, such as communication through natural language processing, storing information through knowledge representation, using this information to draw conclusions through automated reasoning, or detecting patterns in data through machine learning.

Third, AI technologies are described as agents that think rationally. This perspective refers to the development of technological systems based on formal logic that can state problems and knowledge in formal terms and process these through a computational model to find a solution for a given problem that fits to stated goal functions.

Fourth, AI technologies are defined as agents that act rationally. This perspective refers to the development of technological systems that are able to fully automate intelligent behavior through computational processes and thus are capable of achieving their goals in the context of their beliefs in form of correct inferences.

These four perspectives will serve us as an analytical lens in order to examine, which understanding of AI technologies is represented or missing within the current literature and how this might influence what we actually know about this type of technology.

2.2 Technology Lifecycle Stages

Digital information and communication technologies are man-made and undergo different lifecycle stages, namely their *design and development*, their *implementation*, e.g. into organizational structures, their deliberate or unintended *use* in different contexts, as well as the *impacts* (both negative and positive) that they pose for society, organizations or individuals (Sarker et al. 2019).

Our understanding is specifically on the stages design, development, and implementation, because these are essential prerequisites for the actual use of a technology and any impacts use may result in. We use a simple definition of these three stages based on widely accepted textbooks on systems analysis and design (Kendall & Kendall, 2014): With “design” we mean the processes and outcomes of technology analysis and design, including phases such as situational analysis (problems, opportunities, objectives), feasibility analysis, requirements collection, capture, and analysis. With “development” we mean the processes and outcomes of defining, testing, coding and preparing the implementation of a new technology. Referring to AI technologies, this stage also includes training and testing the AI model as well as the respective data preparation and modeling (Chen et al., 2019). With “implementation” we mean the processes and outcomes of introducing a new technology to the organization, including phases such as user training, documentation, system integration, and data transfer.

2.3 The Continuum of Sociotechnical Perspectives

Already the basic definition of AI technologies as agents that think or act rationally or humanly shows that the phenomenon itself combines both social aspects (agents that think or act like humans would) and technical aspects (the agents as technology objects that can think or act rationally). Likewise, as research on AI technology has blossomed across disciplines (Perrault et al., 2019), researchers from different fields have at times focused on the technology whilst blackboxing social aspects such as the human and organizational side or vice versa (Berente et al., 2021).

To differentiate these perspectives, we use the schema developed by Sarker et al. (2019) that defines six different viewpoints that sociotechnical research can incorporate:

Type	**Name**	**Description**
I	Predominantly social	Either the investigation only focuses on the social component, and does not directly address technical component or the investigation mostly focuses on the social component, and the technical component is addressed in an indirect or contextual way.
II	Social imperative on the technical	Technology as a predominant outcome of social structures or processes.
III	Social and technical as additive antecedents to outcomes	Both social component and technical component are antecedents to certain outcomes; however, there is generally no evidence of any interaction between the components themselves while producing these outcomes.
IV	Social and technical as interactive to produce outcomes	Social and technical are both considered as critical to produce outcomes, but the focus is on the interplay between the two components (such as fit/alignment, reciprocal interactions, or entanglement/imbrication) that produce those outcomes.
V	Technical imperative on the social	Technology as the major antecedent to social outcomes, such as those in impact or evaluation studies.
VI	Predominantly technical	Focusing solely on how to develop or improve the technical (e.g., database algorithm) and very limited and direct concern about the role of the social.

Table 1. Summary of six sociotechnical categories following Sarker et al. (2019).

Importantly, these six perspectives are all sociotechnical, that is, they all acknowledge the existence, and interdependencies of technical artifacts as well as the individuals or collectives that develop and use the artifacts in social contexts (Briggs et al., 2010). The perspectives simply vary in emphasis that they give to either the social or the technical.

3 Method

In this section, we explain the methodological procedures we followed and the key decisions we took. First, we classify our review and argue how it connects to the current state of the art. Second, we report on how the review process has been carried out.

3.1 Scope and Classification of the Review

Several reviews of AI technology research exist. We found nine reviews that explicitly deal with aspects concerning design, development, or implementation of AI technologies. These reviews typically focus on specific aspects such as explainability (Markus et al., 2021), fairness (Robert et al., 2020), safety approaches (Dey and Lee, 2021), or social characteristics (Chaves and Gerosa, 2021), on success factors impacting AI technology implementation (Merhi, 2022), on the impact of human factors in designing AI technology (Felmingham et al., 2021), or on ethical risks and opportunities of AI technology design, development and implementation (Floridi et al., 2018). To the best of our knowledge, there is only one literature review that pursues ambitions similar to us. Safaei and Haki (2022) report in a systematic literature review on the question of how research on AI technology within the IS discipline has contributed to our sociotechnical understanding of AI-related phenomena.

To complement and extend the contributions made through these reviews, we decided to (1) cross the boundaries of different scientific communities, and (2) distinguish different types of AI technologies the current research is referring to rather than using “AI” or “machine learning” as an umbrella term with partly inconsistent definitions.

Specifically, we decided to undertake a scoping review (Paré et al., 2015). A scoping review is useful when dealing with broader fields of interest in comparison to well-defined research questions (Arksey & O’Malley, 2005). Scoping reviews are especially suitable to explore the extent and nature of the current body of knowledge in order to identify possible research gaps and relevant research opportunities (Paré et al., 2015).

3.2 Literature Review Strategy

In what follows, we report on how we carried out our review in five stages (Arksey & O’Malley 2005): (1) identifying the research question, (2) identifying relevant studies, (3) study selection, (4) charting the data, (5) collating, summarizing and reporting the results.

Stage 1 – Identifying the research question. In accordance with the aim of this scoping review, our broad research question is: *What do we know about the design, development, and implementation of different AI technologies from a sociotechnical point of view?*

Stage 2 – Identifying relevant studies.

The aim of a scoping study is to be as comprehensive as possible in identifying the available literature on a topic of interest (Arksey & O’Malley, 2005). To keep the scope manageable, we limited the time frame of search to 2012-2022. We used Google scholar as a trans-disciplinary search engine to identify scientific papers matching 18 search terms, constructed by combining terms such as “design”, “development”, “implementation” (plus synonyms) in combination with “AI” or “Artificial Intelligence”. This search led to 5.077.099 results in total. For each search term, we processed the 100 most relevant results, meaning that we identified 1800 potentially relevant articles in total.

Stage 3 – Study selection.

In order to decide which of the 1800 studies to include in our sample, we defined three criteria: (1) as a proxy for quality of the research, we only considered paper published in journals indexed ranked in Journal Impact Factor quartiles one or two. We manually collated all journals of our sample with the Clarivate’s Journal Citation Reports. (2) We only included studies that dealt with AI-related or induced issues in connection with AI technology design, development, or implementation. (3) We only considered studies that were in some form sociotechnical, that is, those that in some way considered social (e.g., practices, beliefs, norms, values) as well as technical (e.g., AI model and its technical capability, data) aspects (section 2.3). We excluded, for example, studies that focus exclusively on the development of the technical artefact (e.g., Yang et al., 2022). Conversely, type of research approach or knowledge contribution (e.g., empirical, theoretical, conceptual research or commentaries) were not used as criteria.

With these filter criteria defined, we screened titles and abstracts of the 1800 publications and selected 27 papers as relevant. We then performed a full-text screening of these papers and excluded two papers due to the lack of conformity with inclusion criteria (3). Next, we applied backwards and forwards reference searching from the remaining 25 papers. This way, we identified 15 additional studies that matched our selection criteria. Thus, we retained 40 paper in our final sample.

Stage 4 – Charting the data.

Charting involves "sorting [the] material according to key issues and themes" (Arksey & O'Malley 2005, p. 26). We used a descriptive-analytical approach by applying a framework to all studies included (Pawson, 2002). The framework we applied for sorting the data, first, captured general information regarding the respective studies such as year of publication, author information, outlet and impact factor ranking. We also coded each paper for research approach, and empirical evidence, if any (Recker et al., 2021). Second, we classified each paper by stage of the technology lifecycle considered, the type of AI technology the studies refer to, and the sociotechnical perspective taken by the researcher(s). Except for the dimension *technology lifecycle*, where papers could be coded under multiple stages, all other dimensions were coded exclusively.

In developing and applying the framework, we proceeded iteratively. First, we developed the categories we used for the review. Through discussions, we developed and refined the concept definitions explained above based on the literature on sociotechnical thinking in IS research (Sarker et al., 2019; Bostrom and Heinen, 1977), research on systems analysis and design (Kendall and Kendall, 2014), and fundamentals of AI technologies (McCorduck, 2004; Russell and Norvig, 2010). Incrementally, we applied, tested and in case revised these coding categories. One of us started by performing the full-text screening and developing a first coding proposal for the complete sample, which included a short justification of the allocation. Based on this, the second author reviewed these proposals, especially commenting on questions, ambiguities or uncertainties that persisted and, if necessary, adding his own perspective on the appropriate coding, which was then returned to the other author for reviewing. We followed this process of iterative evaluation and review for three rounds until we agreed that we have reached an inter-subjective consistent and reliable shared categorization of all studies.

Stage 5 - Collating, summarizing and reporting the results.

We present the findings from the review in both summative and narrative ways (Arksey and O'Malley 2005; Pare et al., 2015). To summarize, we constructed descriptive cross-sectional tables aggregating the distribution of papers across our analytical dimensions. To narrate, we describe our findings alongside our analytical dimensions, thereby also considering the type of research conducted, in order to expose the main fields of interest as well as to highlight relevant research gaps.

4 Findings

In this section, we present the main findings of our scoping review. Table 2 summarizes how our sample of literature is distributed over the analytical dimensions we used in our review. Table 3 classifies the literature we considered by technology lifecycle stage and type of study conducted.

Sociotechnical Focus / **Type of AI Technology**	**Predominantly social**	**Social imperative**	**Social and technical (additive)**	**Social and technical (inter-active)**	**Technical imperative**	**Predominantly technical**
Agents that think humanly		1 (DES)			1 (DEV)	
Agents that act humanly	2 (IMP)	3 (DES) 1 (IMP)	1 (DES) 6 (IMP)		4 (DES)	2 (DES) 3 (DEV) 1 (IMP)
Agents that think rationally	1 (DES) 1 (DEV)	1 (DES) 2 (DEV) 1 (IMP)	1 (DES) 1 (IMP)	1 (DES)	2 (DES) 1 (DEV)	1 (DES) 1 (DEV)
Agents that act rationally	2 (DES)	3 (DES)	1 (DES)	1 (IMP)	1 (DES)	
Other: Not specified					1 (DES)	

Table 2. Distribution of sample alongside the three analytical dimensions technology lifecycle stage DES = design, DEV = development, IMP = implementation, sociotechnical focus and type of AI technology.

Technology Lifecycle Stage / **Study Type**	**Design**	**Development**	**Implementation**
Case Study	1	2	2
Survey		1	1
Experiment	2		
Other - Empirical	5	2	1
Literature Review	6	1	2
Commentary	5	2	1
Theoretical	7	1	6
Other – Non-empirical			

Table 3. Distribution of sample alongside technology lifecycle stage and study type.

4.1 Design of AI Agents

With "design" we mean the processes and outcomes of technology analysis and design, which includes phases such as the situational analysis asking for the context-specific problems, opportunities and objectives related to the development of a new technology, the feasibility analysis as well as the collection, capture, and analysis of requirements.

We assigned 26 studies to this category. Of these, 16 studies perused a perspective in which one of the components – either the social or the technical – have a dominant impact on the other one. Research that referred to a point of view where *the technical is imperative on the social*, dealt with the impact of certain design aspects of AI technologies, such as its capacity for sensing, thought and action autonomy, or communication style, on humans or particular desired social outcomes (e.g. Hu et al., 2021; Kim et al., 2021; Baek et al., 2021; Robert et al., 2020; Asatiani et al., 2020; Chaves and Gerosa, 2021; Apiola and Sutinen, 2020; Seo et al., 2021). Research that incorporated a perspective where *the social is imperative on the technical*, dealt with the role and the influence of human or social factors, such as personality traits or human values, on the process and outcome of AI technology design (e.g., Umbrello et al., 2021; Felmingham et al., 2021; Martin, 2019; Rohlfing et al., 2021; van den Broek et al., 2021; Truby, 2020; van de Poel, 2020; Morley et al., 2021).

Six studies in our sample incorporated technical as well as social components into their analysis but conceptualized either the social or the technical factors as rather contextual to the other ones. These studies addressed the development of principles or guidelines for how humans should design AI technology in order to ensure a beneficial societal and humanistic outcome (*predominantly social*, e.g. Floridi et al., 2018; Li, 2021; Sloane, 2022), specific AI-based solutions based on design science approaches, or concrete frameworks governing the technical process of AI technology development and giving specific technical recommendations on how to design the AI model in order to ensure a transparent, explainable and safe output (*predominantly technical,* e.g., Markus et al., 2021; Dey and Lee, 2021; Villegas-Ch et al., 2021).

Three studies focused on understanding and analyzing social and technical components as additive factors. Research from this point of view has dealt with the question of how technical mechanisms and design techniques are able to ensure fair outcomes, given the human perception of fairness, or with social processes, e.g. decision-making, in relation to or within the design process of AI technology (*social and technical as additive antecedents to outcome,* e.g., Morse et al., 2022; Umbrello, 2022; Choi et al., 2020).

Finally, one study of our sample assigned to the stage of “design” was based on a sociotechnical perspective that understands the technical and the social components as interactive factors in the production of outcomes. Fügener et al. (2021) examined how AI technology should be designed in order to enable a sensible delegation of tasks between human and AI technology (*social and technical as interactive to produce outcomes*).

Taking into consideration the type of AI technology that is referred to within the research on AI technology design, first of all, all types of AI technologies were featured in the literature. Most studies (10) referred to AI technologies as *agents that act humanly*. These publications predominantly discussed either the social component as a dominant influence on the technical component (e.g., Umbrello, 2021), or the technical component as imperative to the social (e.g., Kim et al., 2021).

Seven studies each referred to AI technologies as *agents that think rationally* and *as agents that act rationally*. The research that considered AI technologies as agents that think rationally was distributed evenly over the continuum of sociotechnical perspectives, whereas the publications which approached AI technologies from the perspective of a rational agent focused on the social side of the continuum by considering predominantly social aspects, or by understanding social factors as imperative on technical factors (e.g., Floridi, 2018).

One study defined AI technologies as *agents that think humanly* (Rohlfing et al., 2021) and one study lacked a clear and consistent definition (Seo et al., 2021).

This distribution is also of significance with regard to the type of research that was done within the distinct analytical dimensions. Overall, eight of the 26 publications included and assigned to the “design” stage took an empirical approach (e.g., van den Broek et al., 2021; Villegas-Ch et al., 2021), 18 papers were based on theoretical or conceptual research or on commentaries (e.g., Morse et al., 2022; Umbrello et al., 2021). Three studies each and thus most *empirical research* on the design of AI technology was done from the perspective of AI technologies as agents that act humanly (e.g., Baek et al., 2021; Kim et al., 2021) or that think rationally (e.g., Asatiani et al., 2021; Fügener et al., 2021), only one paper defined AI technologies as agents that are acting rationally (Hu et al., 2021). Analyzing the *conceptual research* in this section, seven publications and thus most conceptual research was done from the perspective of AI technologies as agents that act humanly (e.g., Robert et al., 2020; Felmingham et al., 2021), which consequently seems to be the most fruitful point of view in current research on the design of AI agents. In contrast to the empirical research, the conceptual research almost equally often drew on the perspective of AI technologies as agents that act rationally (six studies, e.g., Martin, 2019; Li, 2021).

Examining the type of research done so far with regard to the *sociotechnical perspective* that is incorporated, the majority of the *conceptual research* was done from the social side of the continuum. Especially the perspective of social components as imperative on the technical has been utilized in seven publications (e.g., van de Poel, 2020; Rohlfing et al., 2021) and is thus the point of view used most frequently in conceptual research.

In contrast, the *empirical research* on the design of AI agents most frequently drew on technical perspectives, especially on the perspective of technical components as imperative on the social (five studies, e.g., Seo et al., 2021; Hu et al., 2021). The current body of research involves some empirical approaches that focus on the social aspects of AI technology design as well as on the additive or interactive combination of technical and social factors. Especially the perspective of technology and humans, organizations and society as interactively entangled is not strongly represented in current works, be they conceptual or empirical.

4.2 Development of AI Agents

With "development" we mean the processes and outcomes of defining, testing, coding and preparing the implementation of a new technology. Referring to AI technologies, this stage also includes training and testing the AI model as well as the respective data preparation and modeling. This stage is often discussed jointly with the "design" stage and the transitions are fluent.

We assigned nine studies to this category. In these studies, the development processes of AI technologies are not as often examined from a sociotechnical point of view than the design processes and outcomes. Three studies incorporated *a predominantly social perspective* (Sloane, 2022) or *a perspective in which the social is understood as imperative on the technical* (Asatiani et al., 2021; Henriksen and Bechmann, 2020). Research from these perspectives dealt with the role and influence of social processes and practices on the development of AI technologies as well as with principles that technologists can apply in order to develop ethical AI technologies. Six studies incorporated a perspective in which *the technical is understood as imperative on the social* (Asatiani et al., 2020; Montes and Goertzel, 2019) or *a predominantly technical point of view* in which social factors were taken into consideration as the context or setting of the research (Villegas-Ch et al., 2021; Dey and Lee, 2021; Gupta et al., 2021; Reddy et al., 2021). Research from these perspectives dealt with the specific development of AI-based solutions or with concrete frameworks governing the technical process of AI technology development, giving explicit technical recommendations on how to build an AI model to ensure a responsible, explainable and safe output.

Next, we into consideration the type of AI agent that is referred to within the current research on AI technology development. Five publications drew on the perspective of *AI technologies as agents that think rationally* (e.g., Gupta et al., 2021; Henriksen and Bechmann, 2020), three conceptualized *AI technologies as agents that act humanly* (e.g., Dey and Lee, 2021; Reddy et al., 2021) and one defined *AI technologies as agents that think humanly* (Montes and Goertzel, 2019). None of the studies included referred to *AI technologies as agents that act rationally*.

Considering the type of research conducted, empirical and conceptual approaches were almost evenly represented in our sample. Five studies took an empirical approach (e.g., Asatiani et al., 2020; Henriksen and Bechmann, 2020; Gupta et al., 2021). Four studies took a conceptual approach (e.g., Sloane, 2022; Dey and Lee, 2021; Reddy et al., 2021). The *empirical studies* distribute as follows: one study referred to AI technologies as agents that act humanly (Villegas-Ch et al., 2021); four studies referred to AI technologies as agents that think rationally (e.g., Asatiani, 2021; Gupta et al., 2021). In contrast, the *conceptual publications* were almost evenly distributed over three of the types of AI agents discussed: one study defined AI technologies as agents that think humanly, two studies referred to AI technologies as agents that act humanly, and one study conceptualized AI technologies as agents that think rationally.

Examining the type of research done so far on the issue of the development of AI technologies with regard to the *sociotechnical perspective* applied in the research, the majority of the conceptual as well as of the empirical research was approached from a technical side of the sociotechnical continuum (e.g., Asatiani et al., 2020; Dey and Lee, 2021), though also conceptual (one publication, Sloane, 2022) and empirical research (two publications, e.g., Henriksen and Bechmann, 2020) exists that focused on the social end of the continuum.

4.3 Implementation of AI Agents

With "implementation" we mean the processes and outcomes of introducing a new AI technology to the organization, including phases such as user training, documentation, system integration, and data transfer.

We assigned 13 papers to this category. The research assigned to this category accounted for the sociotechnical perspective in a narrow sense: seven publications incorporated a perspective in which social and technical components have been understood as equally critical factors in the process of AI technology implementation, both influencing the outcome through the combination of their specific characteristics (*social and technical as additive*, e.g. Daye et al., 2022; Liew, 2018; He et al., 2019). Research from this perspective dealt with the technical (e.g., data, model) and social aspects (e.g., expectations of AI technology's potential added value, presence of innovation strategies, trust) that need to be considered when implementing AI technologies in organizational settings.

Two publications were found to incorporate a perspective in which the technical aspects have been considered in a contextual way to the social aspects (*predominantly social*, Ongena et al., 2020; Pachidi et al., 2021) and two further publications characterized the social aspects as decisive influence on the technical factors involved in the process of AI technology implementation (*social imperative on technical*, Felmingaham et al., 2021; Asatiani, 2021). Consequently, much of the research assigned to the "implementation" stage was done from the social end of the sociotechnical continuum, dealing with social practices and methods for AI technology implementation and governance as well as with the influence of human factors such as norms or preferences on the process. Moreover, current research seems to focus on the social processes that surround the implementation of AI technology, while the technical pitfalls and especially their interplay with the social forces in a situated context find less consideration.

One study analyzed the implementation of AI technology from a point of view that recognizes an interaction between the social and the technical components involved (*social and technical as interactive*, Makarius et al., 2020). This paper derived a framework on how employees and AI technology can and should collaborate in order to achieve a competitive advantage for an organization, depending on the scope and the level of novelty of the respective AI technology.

Finally, one study analyzed the "implementation" stage from the technical end of the sociotechnical continuum by understanding the technical components as decisive and embedded in a social context, providing an evaluation framework which was based on the assessment of the technical capability of the AI technology, its contextual relevance and efficiency, potential use cases as well as the technical integration of the AI technology into related workflows (*predominantly technical*, Reddy et al., 2021).

Taking into consideration the type of AI technology that is addressed by the research on the implementation of AI agents, it is remarkable that 10 out of 13 and thus most of the studies included in this section related to *AI technologies as agents that act humanly* (e.g., Choi et al., 2020; Strohm et al., 2020; DeCamp and Lindvall, 2021). Two publications referred to *AI technologies as agents that think rationally* (Asatiani et al., 2021; Mehri, 2022) and only one understood *AI technologies as agents that act rationally* (Makarius et al., 2020).

Regarding the type of research, nine out of 13 publications were based on *conceptual or theoretical research* (e.g., Daye et al., 2022; Reddy et al., 2021). Four studies deployed an *empirical approach* to the phenomena of interest in this section (e.g., Pachidi et al., 2021; Ongena et al., 2020).

The conceptual research was mostly done based on the understanding of *AI technologies as agents that act humanly* (e.g., Choi et al., 2020; DeCamp and Lindvall, 2020). This also holds true for the empirical research included (e.g., Pachidi et al., 2021; Strohm et al., 2020). Thus, this perspective can be described as the central understanding of AI technology within current research on its implementation.

Comparing the type of research from the point of view concerning the sociotechnical focus that was incorporated in the respective studies, it appears that, while the empirical research mostly focused on the socially determined perspectives by analyzing the social factors in a technical context (*predominantly social*, Pachidi et al., 2021; Ongena et al., 2020) or by understanding the social aspects as decisive influence on the technical process and outcome (*social imperative on technical*, Asatiani et al, 2021), the conceptual research largely drew on a sociotechnical perspective in the narrower sense by understanding *technical and social components as additive antecedents* to the implementation of AI technologies (six out of nine studies, e.g. Liew, 2018; He et al., 2019; Daye et al., 2022).

5 Discussion

In this section, we present the contributions of our research and identify implications for research and practice. We also address aspects in which our review is limited and describe opportunities for further research.

5.1 Contributions and Implications for Research

Our review shows that there is a substantial body of work that examines AI technology design, development, and implementation processes and their outcomes from a sociotechnical perspective. We found all types of AI technologies being covered, as well as evidence for all lifecycle stages being covered and a range of sociotechnical perspectives being taken. Still, our scoping review allows us to identify several blind spots and points of reference, which can become fruitful starting points for further sociotechnical research:

First, relating to the agency perspective, our review demonstrates that most of the current research understands AI technologies as agents that either act humanly or think rationally. This means that current research has focused on understanding the design, development and implementation of AI technologies that are capable of performing cognitive tasks human-like or deducting correct inferences based on formal logic. Both perspectives are accompanied by restrictions. On the one hand, assuming that technology is artificially intelligent when able to perform tasks human-like means deliberately attempting to model human fallibility. On the other hand, assuming that artificial intelligence is given when AI technologies draw correct conclusions ("know-what") based on correct premises involves the difficulty of how to make the necessary knowledge explicit. Especially informal knowledge ("know-how") is often not precise enough to be processed based on the rules of formal logic (Lebovitz et al., 2021). Thus, the spectrum and depth of problems that can be solved by such kinds of AI technology might be limited. Furthermore, there is also a difference between solving a problem principally and in practice (Russell & Norvig, 2010).

Moving forward, the question emerges in how far the current body of knowledge can account for AI technologies that are designed, developed and deployed as agents that act rationally. For example, consider AI technologies that act as "human assistants" that capture the user's needs and act on their behalf, thereby performing tasks in line with their preferences autonomously and goal-oriented (Hu et al., 2021; Han and Yang, 2018). AI technologies as agents that act rationally expand on the ideas of AI technologies as agents that act humanly or think rationally because they embed human-like abilities of AI agents and their capabilities to make correct inferences into a broader ability for goal-oriented action, based on which the AI agent is able to make own determinations as well as to develop its determinations over time. Research should focus more on this broader conception of AI technologies as smart agents in complex environments because they fundamentally differ from other advanced technologies (Murray et al. 2020; Russell & Norvig, 2010).

Second, even though the literature we reviewed demonstrates that research on AI technology design, development and implementation from a sociotechnical perspective unfolds alongside a continuum in which the influence of various social and technical factors is varying, there seems to be a tendency in the literature to focus on the ends of the continuum – the rather social or the rather technical. This could be balanced with more research on the intersection of the technical and the social in a narrower sense to "ensure a healthy distribution of papers" (Sarker et al., 2019, p. 708) alongside the sociotechnical continuum.

Understanding the technical and the social aspects coming into effect in these stages as additive or even as interacting components could be valuable for several reasons. First, the processes surrounding the design, development and implementation are social processes themselves, influenced by values, norms, beliefs, knowledge and social relations which affect the way decisions on technical aspects and issues are made (Johnson and Powers, 2008; Umbrello, 2022; Mittelstadt et al., 2016). Second, these processes also lead to an outcome (the AI technology as an agent) that is constructed to interact with human users (Johnson and Powers, 2008; Mittelstadt et al., 2016). Therefore, designing, developing and implementing also means defining and shaping the way in which social and technical entities will interact, e.g., how they will collaborate, how tasks will be allocated or how decisions will be made and related responsibilities are delegated (Martin, 2019b; Fügener et al., 2021b; Onnasch et al., 2014; Tatasciore et al., 2021). As Taylor et al. (2001) or Leonardi (2011) argue, it is always a human decision, how material agency – in this case, an AI agent – is designed and implemented and thus becomes intertwined with human goals.

A case in point are the efforts toward "explainable AI" (Gunning et al., 2019) or "responsible AI" (Werder et al., 2022), both of which are essential sociotechnical imperatives in which socially constructed requirements, such as demands for explainability or responsibility, are meant to interact in a balanced way with technical functionality implemented in AI agents, such as data or algorithm constructions. These examples highlight a need to understand the design, development and implementation processes of AI technology from a sociotechnical point of view as a process of human agency constructing material agency and thereby configuring how these two components will become entangled in use. This entanglement in turn might lead not only to the change of social practices but also to the change of technical components or specifications. Thus, social factors and technical components should not be seen as independent entities within the context of each other; rather they are interwoven based on the technical constraints and affordances in a given social context (Leonardi, 2011).

Third, our review indicates that the research to date involves some empirical insights but more could be done. While there is value in conceptual research done from a sociotechnical perspective, as well as a in the available normative contributions in the form of commentaries, guidelines, or frameworks on how the design, development and implementation of AI technologies could or should be conducted in safe and ethically sound ways, and why doing so would be important, even more empirical research on how such advice is reflected in the actual practices would be meaningful, as would be research that evaluates the efficacy of the many guidelines and frameworks. Demand continues for researchers to empirically examine the application and utility of normative research on AI technologies in practice and to contrast these insights with practice-based views on how practitioners in fact design, develop, and implement AI technologies (Mittelstadt, 2019; Vakkuri et al., 2021).

5.2 Potential Implications for Practice

Our scoping review may also be valuable to practitioners in charge of the design, development and/ or the implementation of AI technology. Our review provides a broad overview over the amount, scope, and findings of academic literature on these topics. It can thereby serve as a point of orientation and entry into the available knowledge around the design, development and/ or the implementation of AI technology. While our review does not explicitly reflect all possible context- or industry-specific particularities, it still allows professionals to gain an in initial impression of how much and what type of evidence is available for their decision-making. Furthermore, our review also raises awareness about "blind spots", where no robust insights are yet available to practitioners, such as, for example, about selected social aspects or ethical issues that might surface during the practical design, development and implementation of AI technologies.

5.3 Limitations

Some limitations apply to our review. First, we used a structured literature search approach with pre-defined search terms. These search terms do not cover all relevant literature, which is why we added a more flexible backwards and forwards referencing search strategy. Moving forward, we plan to extend our literature selection by applying a broader set of keywords, such as the terms “AI”, “Artificial Intelligence”, “AI system”, as well as further backwards and forwards reference searches.

Second, our review did not explicitly refer to research streams that are involved with “explainable” or “responsible AI”. However, this might be an interesting and important extension; consequently, we aim to expand our literature search in this direction as well.

Third, we used a limited timeframe (2012-2022). Thus, it would be valuable to extend the timeframe to broaden the overview given within this review.

Fourth, our review had a specific scope. We focused on research examining issues related to the design, development, and implementation of AI technologies, analyzing it in the direction of the sociotechnical perspective it has incorporated, the type of AI technology it focuses on and the type of research that is applied. These dimensions are not exclusive. Thus, there might be different points of view such as specific design aspects (e.g., different development methodologies) that could provide a meaningful extension. For example, future research could make use of explicit AI lifecycle models (e.g., the CRISP-DM model, see Chapman et al., 2000) for analyzing and classifying the literature of interest. This could be valuable for two reasons: first, it would allow for more specific conclusions that relate directly to the technology of interest; second, in doing so, it would unpack some specific processes of AI development, such as training and testing as well as the related significance of data within the development process, in more detail than our coarse differentiation allows. Currently, these aspects are not explicitly covered in our broad view on AI development, and we see the value in unearthing these specific aspects in more detail.

Fifth, our lifecycle stages design, development, and implementation of AI technologies are also neither disjoint nor complete. Moreover, several studies exist that address some of these aspects but also cover AI technology use and impact – categories that we excluded from our review. For example, Lebovitz et al. (2021) study the changes in evaluating AI technologies during their use but their study also addresses some aspects of implementation (e.g., the rationales used by decision-makers to select one technology over another).

Sixth, in alignment with our scoping review approach, we decided to take a broad focus regarding the different types of AI technologies as well as the areas of research we are referring to. This might lead to the question in how far our contributions, based on a limited sample of papers, allow to make general claims about the design, development and implementation processes of AI technologies overall. This point of caution is valid. For example, our conclusions about the scope of available empirical research, or the balance in social, technical, or sociotechnical characteristics in research on AI technologies should always be interpreted in light of the sample of papers we considered here. Continuing from here, it might be interesting to analyze further how research from a specific application context (e.g., healthcare) conceptualizes the processes discussed in this review in order to allow for a more nuanced understanding.

6 Conclusion

Though our review is not exhaustive, we hope that the focus of our analysis, drawing on sociotechnical systems theory, theory on AI technologies, and lifecycle models from systems analysis and design, helps shedding light on how the sociotechnical perspective is represented within research on the processes and outcomes of AI technology design, development and implementation. We hope for our review to provide a point of orientation for IS researchers regarding areas and perspectives that have been active or underemphasized in research.

References

Abbasi, A., Li, J., Clifford, G., & Taylor, H. *Make "Fairness by Design" Part of Machine Learning.* Retrieved November 13, 2022, from https://hbr.org/2018/08/make-fairness-by-design-part-of-machine-learning.

Apiola, M.-V., & Sutinen, E. (2020). Towards Constructivist Design of Artificial Intelligence: Perspectives and Ideas. *Constructivist Foundations*, *16*(1).

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19–32.

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2020). Challenges of Explaining the Behavior of Black-Box AI Systems. *MIS Quarterly Executive*, 259–278.

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. Journal of the Association for Information Systems, 22(2), 325–352.

Baek, T. H., Bakpayev, M., Yoon, S., & Kim, S. (2021). Smiling AI agents: How anthropomorphism and broad smiles increase charitable giving. *International Journal of Advertising*, 1–18.

Balakrishnan, T., Chui, M., Hall, B., & Henke, N. *The State of AI in 2020*. Retrieved November 07, 2022, from https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*(3), 1433-1450.

Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio-Technical Perspective, Part II: The Application of Socio-Technical Theory. *MIS Quarterly*, *1*(4), 11–28.

Briggs, R. O., Nunamaker, J. F., & Sprague, R. H. (2010). Special Section: Social Aspects of Sociotechnical Systems. *Journal of Management Information Systems, 27*(1), 13–16.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T. P., Shearer, C. et al. (2000). CRISP-DM 1.0: Step-by-step data mining guide, from *https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf*

Chaves, A. P., & Gerosa, M. A. (2021). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758.

Chen, P.-H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. *Nature Materials*, *18*, 410–427.

Choi, H. H., Chang, S. D., & Kohli, M. D. (2020). Implementation and design of artificial intelligence in abdominal imaging. *Abdominal radiology (New York), 45*(12), 4084–4089.

Crowley, J. (2019). Toward AI Systems that Augment and Empower Humans by Understanding Us, our Society and the World Around Us, from *https://www.humane-ai.eu/wp-content/uploads/2019/11/D21-HumaneAI-Concept.pdf*

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42.

Daye, D., Wiggins, W. F., Lungren, M. P., Alkasab, T., Kottler, N., Allen, B., et al. (2022). Implementation of Clinical Artificial Intelligence in Radiology: Who Decides and How? *Radiology*, from https://doi.org/10.1148/radiol.212151.

DeCamp, M., & Lindvall, C. (2020). Latent bias and the implementation of artificial intelligence in medicine. *Journal of the American Medical Informatics Association : JAMIA*, *27*(12), 2020–2023.

Dey, S., & Lee, S.-W. (2021). Multilayered review of safety approaches for machine learning-based systems in the days of AI. *Journal of Systems and Software, 176*, 110941.

Felmingham, C. M., Adler, N. R., Ge, Z., Morton, R. L., Janda, M., & Mar, V. J. (2021). The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World. *American journal of clinical dermatology, 22*(2), 233–242.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and machines, 28*(4), 689–707.

Fu, R., Huang, Y., & Singh, P. V. (2021). Crowds, Lending, Machine, and Bias. *Information Systems Research*, 32(1), 72–92.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. *MIS Quarterly, 45*(3), 1527–1556.

Gill, T. G. (1995). Early Expert Systems: Where Are They Now? *MIS Quarterly, 19*(1), 51–81.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2022). XAI - Explainable Artificial Intelligence. *Science Robotics, 4*(37).

Gupta, S., Kamboj, S., & Bag, S. (2021). Role of Risks in the Development of Responsible Artificial Intelligence in the Digital Healthcare Domain. *Information Systems Frontiers*.

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants. *Industrial Management & Data Systems, 118*(3), 618–636.

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature medicine, 25*(1), 30–36.

Henriksen, A., & Bechmann, A. (2020). Building truths in AI: Making predictive algorithms doable in healthcare. *Information, Communication & Society, 23*(6), 802–816.

Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management, 56*, 102250.

Johnson, D., & Powers, T. (2008). Computers as Surrogate Agents. In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy* (pp. 251–269). Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo: Cambridge University Press.

Kendall, K. E., & Kendall, J. E. (2014). *Systems Analysis and Design* (9th ed.). London: Pearson.

Kim, J., Merrill, K., Xu, K., & Sellnow, D. D. (2021). I Like My Relational Machine Teacher: An AI Instructor's Communication Styles and Social Presence in Online Education. *International Journal of Human–Computer Interaction, 37*(18), 1760–1770.

Kroes, P., Franssen, M., van Poel, I. de, & Ottens, M. (2006). Treating socio-technical systems as engineering systems: some conceptual problems. *Systems Research and Behavioral Science, 23*(6), 803–814.

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly, 45*(3), 1501–1526.

Lee, A. S. (2001). Editors Comments: Research in Information Systems: What We Haven't Learned. *MIS Quarterly, 25*(4), v–xv.

Leonardi, P. M. (2011). When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. *MIS Quarterly, 35*(1), 147–167.

Li, O. (2021). Problems with "Friendly AI". *Ethics and Information Technology, 23*(3), 543–550.

Liew, C. (2018). The future of radiology augmented with Artificial Intelligence: A strategy for success. *European journal of radiology, 102*, 152–156.

Makarius, E. E. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*. (120), 262–273.

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of biomedical informatics, 113*, 103655.

Martin, K. (2019). Designing Ethical Algorithms. *MIS Quarterly Executive, 18*(2), 129–142.

Martin, K. (2019b). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics, 160*(4), 835–850.

McCorduck, P. (2004). *Machines Who Think* (2nd ed.). New York: Taylor & Francis.

Metcalf, L., Askay, D. A., & Rosenberg, L. B. (2019). Keeping Humans in the Loop: Pooling Knowledge through Artificial Swarm Intelligence to Improve Business Decision Making. *California Management Review, 61*(4), 84–109.

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

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2).

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. (2021). Algorithmic Management of Work on Online Labor Platforms: When Matching Meets Control. *MIS Quarterly, 45*(4), 1999–2022.

Montes, G. A., & Goertzel, B. (2019). Distributed, decentralized, and democratized artificial intelligence. *Technological Forecasting and Social Change, 141*, 354–358.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and machines*, *31*(2), 239–256.

Morse, L., Teodorescu, M. H. M., Awwad, Y., & Kane, G. C. (2021). Do the Ends Justify the Means? Variation in the Distributive and Procedural Fairness of Machine Learning Algorithms. *Journal of Business Ethics*.

Murray, A., Rhymer, J., & Sirmon, D. G. (2021). Humans and Technology: Forms of Conjoined Agency in Organizations. *Academy of Management Review, 46*(3), 552–571.

Ongena, Y. P., Haan, M., Yakar, D., & Kwee, T. C. (2020). Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. *European radiology, 30*(2), 1033–1040.

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*(3), 476–488.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing. *Organization Science, 32*(1), 18–41.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199.

Pawson, R. (2002). Evidence-based Policy: In Search of a Method. *Evaluation*, *8*(2), 157–181.

Pearl, J. (Ed.) (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco (CA): Morgan Kaufmann.

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B. et al. (2019). The AI Index 2019 Annual Report. AI Index Steering Committee, Human-Centered AI Institute, Stanford University.

Pillai, R., & Sivathanu, B. (2020). Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. *Benchmarking: An International Journal*, *27*(9), 2599–2629.

Raisch, S., & Krakowski, S. (2021). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review, 46*(1), 192–210.

Recker, J., Briel, F. von, Yoo, Y., Nagaraj, V., & McManus, M. (2023). Orchestrating Human-Machine Designer Ensembles during Product Innovation. *California Management Review, forthcoming*.

Recker, J., Lukyanenko, R., Jabbari, M., Samuel, B. M., & Castellanos, A. (2021). From Representation to Mediation: A New Agenda for Conceptual Modeling Research in a Digital World. MIS Quarterly, 45(1), 269–300.

Reddy, S., Rogers, W., Makinen, V.-P., Coiera, E., Brown, P., Wenzel, M., et al. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health Care Informatics, 28*(e100444).

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: a review, critique, and design agenda. *Human–Computer Interaction, 35*(5-6), 545–575.

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., et al. (2021). Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(3), 717–728.

Russell, S. J. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. London: Allen Lane.

Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River: Prentice Hall.

Safaei, D., Haki, K., & Morin, J.-H. (2022). Artificial Intelligence in Information Systems Research: A Socio-technical Perspective. Retrieved October 28, 2022, from https://www.researchgate.net/publication/364359668_Artificial_Intelligence_in_Information_Systems_Research_A_Socio-technical_Perspective.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly, 43*(3), 695–719.

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology, 24*(1).

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research, 32*(3), 736–751.

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner-instructor interaction in online learning. *International journal of educational technology in higher education, 18*(1), 54.

Sloane, M. (2022). Here's what's missing in the quest to make AI fair. *Nature*. (605).

Strohm, L., Hehakaya, C., Ranschaert, E. R., Boon, W. P. C., & Moors, E. H. M. (2020). Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *European radiology, 30*(10), 5525–5532.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science (New York, N.Y.), 361*(6404), 751–752.

Tatasciore, M., Bowden, V. K., Visser, T. A. W., & Loft, S. (2021). Should We Just Let the Machines Do It? The Benefit and Cost of Action Recommendation and Action Implementation Automation. *Human Factors, 64*(7), 1121–1136.

Taylor, J. R. (2001). Toward a Theory of Imbrication and Organizational Communication. *The American Journal of Semiotics, 17*(2), 269–297.

Terry, N. (2019). Of Regulating Healthcare AI and Robots. *Yale Journal of Law and Technology, 21*(3), 133–190.

Truby, J. (2020). Governing Artificial Intelligence to benefit the UN Sustainable Development Goals. *Sustainable Development, 28*(4), 946–959.

Tsolakis, N., Zissis, D., Papaefthimiou, S., & Korfiatis, N. (2022). Towards AI driven environmental sustainability: an application of automated logistics in container port terminals. *International Journal of Production Research, 60*(14), 4508–4528.

Umbrello, S. (2022). The Role of Engineers in Harmonising Human Values for AI Systems Design. *Journal of Responsible Technology, 10*, 100031.

Umbrello, S., Capasso, M., Balistreri, M., Pirni, A., & Merenda, F. (2021). Value Sensitive Design to Achieve the UN SDGs with AI: A Case of Elderly Care Robots. *Minds and machines, 31*(3), 395–419.

Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E., & Abrahamsson, P. (2021). ECCOLA — A method for implementing ethically aligned AI systems. *Journal of Systems and Software, 182*, 111067.

van de Poel, I. (2020). Embedding Values in Artificial Intelligence (AI) Systems. *Minds and Machines, 30*(3), 385–409.

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

Villegas-Ch, W., García-Ortiz, J., Mullo-Ca, K., Sánchez-Viteri, S., & Roman-Cañizares, M. (2021). Implementation of a Virtual Assistant for the Academic Management of a University with the Use of Artificial Intelligence. *Future Internet, 13*(4), 97.

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

Yang, M., Moon, J., Yang, S., Oh, H., Lee, S., Kim, Y., & Jeong, J. (2022). Design and Implementation of an Explainable Bidirectional LSTM Model Based on Transition System Approach for Cooperative AI-Workers. *Applied Sciences, 12*(13).