ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/359312252

Ethics and AI in Information Systems Research

Article *in* Communications of the Association for Information Systems · June 2022

DOI: 10.17705/1CAIS.05034

CITATIONS
16

READS
2,040

3 authors:

Milad Mirbabaie
Otto-Friedrich-Universität Bamberg

161 PUBLICATIONS **3,241** CITATIONS

SEE PROFILE

Alfred Benedikt Brendel
Indiana University Bloomington

108 PUBLICATIONS **1,968** CITATIONS

SEE PROFILE

Lennart Hofeditz
University Hochschule Niederrhein

29 PUBLICATIONS **536** CITATIONS

SEE PROFILE

All content following this page was uploaded by Lennart Hofeditz on 01 September 2022.

The user has requested enhancement of the downloaded file.

Communications of the Association for Information Systems

Volume 50 Article 38

6-27-2022

Ethics and AI in Information Systems Research

Milad Mirbabaie
Paderborn University, milad.mirbabaie@uni-paderborn.de

Alfred B. Brendel
Faculty of Business and Economics, Technische Universität Dresden

Lennart Hofeditz
Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen

Follow this and additional works at: https://aisel.aisnet.org/cais

Recommended Citation

Mirbabaie, M., Brendel, A. B., & Hofeditz, L. (2022). Ethics and AI in Information Systems Research. Communications of the Association for Information Systems, 50, pp-pp. https://doi.org/10.17705/1CAIS.05034

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Communications of the Association for Information Systems

Special Section DOI: 10.17705/1CAIS.05034 ISSN: 1529-3181

Ethics and AI in Information Systems Research

Milad Mirbabaie
Department of Information Systems, Paderborn University
milad.mirbabaie@uni-paderborn.de

Alfred Benedikt Brendel
Faculty of Business and Economics, Technische Universität Dresden

Lennart Hofeditz
Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen

Abstract:

The ethical dimensions of Artificial Intelligence (AI) constitute a salient topic in information systems (IS) research and beyond. There is an increasing number of journal and conference articles on how AI should be designed and used. For this, IS research offers and curates knowledge not only on the ethical dimensions of information technologies but also on their acceptance and impact. However, the current discourse on the ethical dimensions of AI is highly unstructured and seeks clarity. As conventional systematic literature research has been criticized for lacking in performance, we applied an adapted discourse approach to identify the most relevant articles within the debate. As the fundamental manuscripts within the discourse were not obvious, we used a weighted citation-based technique to identify fundamental manuscripts and their relationships within the field of AI ethics across disciplines. Starting from an initial sample of 175 papers, we extracted and further analyzed 12 fundamental manuscripts and their citations. Although we found many similarities between traditionally curated ethical principles and the identified ethical dimensions of AI, no IS paper could be classified as fundamental to the discourse. Therefore, we derived our own ethical dimensions on AI and provided guidance for future IS research.

Keywords: Ethics, Artificial Intelligence, Discourse Approach, Review Article, Information Systems.

This manuscript underwent peer review. It was received 9/1/2020 and was with the authors for 18 months for three revisions. The Associate Editor chose to remain anonymous.

1 Introduction

While organizations and researchers have repeatedly shown the advantages of Artificial Intelligence (AI)-based systems for humanity (such as self-driving cars, AI-based conversational agents, and process automation), serious AI-related abuses and incidents have raised pressing ethical concerns (Benbya et al., 2021; Berente et al., 2021; Seppälä et al., 2021). While unethical behavior can be intended in some cases due to skewed organizational or managerial values (e.g. during the VW diesel scandal) (Stieglitz et al., 2019), many unintended ethical challenges and moral issues can occur when applying AI (Boddington, 2017). For instance, Amazon's discriminatory human resources (HR) software and Microsoft's racist chatbot provide a strong case for the dangerous and unethical sides of AI that were inadvertent (Dastin, 2018; Horton, 2016; Yampolskiy, 2016). Furthermore, organizations such as Uber increasingly apply AI-based algorithms for exerting autonomous managerial control over employees (referred to as algorithmic control), resulting in constant surveillance, less transparency and possible dehumanization (Wiener et al., 2021). On the one hand, these unethical sides of AI and algorithms are grounded in biased man-made algorithms. The latter are used, for instance, in hiring, and cannot be completely non-discriminatory (Mann & O'Neil, 2016). On the other hand, this is due to the predictive nature of AI, resulting in a non-transparent derivation of outputs (Boddington, 2017).

There have been many different approaches to defining artificial intelligence in the past. Carvalho et al. (2019) considered AI as a group of technologies that rely on techniques such as machine learning, natural language processing, and knowledge representation. However, we do not consider AI to be a single technology or a group of specific technologies. We follow the definition of an AI as *"the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems" (Berente et al., 2021)*.

There is, however, a conflict between AI and ethics. Advances in AI technologies require increasing amounts of data to make AI work properly while, at the same time, the technology is being given more and more autonomy. Normative ethics, in contrast, aim to protect the rights of individuals, including data and autonomy. AI technologies can be applied to many different use cases. It is, therefore, difficult for organizations, researchers, and policymakers to draw up ethical guidelines that are neither too narrowly targeted on a specific use case, nor too vague. In addition, organizations such as IBM have defined ethical principles for themselves, although this does not prevent them from pursuing unethical AI activities (Robin, 2019). Researchers will need to address this conflict between ethics and AI and develop strategies to resolve it.

Accordingly, the increasing influence of AI on society as well as individuals goes hand-in-hand with the increasing pressure on organizations to acknowledge responsibility for their AI products and offerings (Brendel et al., 2021). This includes ethical considerations as to their AI's potential consequences. For leaders to incorporate ethical considerations into their decisions when applying AI in their organizations, they need guidance from research. With knowledge in both normative ethics (e.g., Stahl, 2012) and organizational processes, the information systems (IS) community clearly offers the potential to take an interdisciplinary bridging role in the ethical application of AI-based systems. Previous research proved that examining the effects of digitalization on principles such as human dignity is an area in which IS scholars can contribute valuable artifacts (Leidner & Tona, 2021). IS researchers connect knowledge from different disciplines and provide theories that can be used to understand and interpret emerging phenomena such as AI. An ethical discourse on AI that has been widely acknowledged by researchers from different disciplines lacks such an interdisciplinary link that IS research can provide (Brendel et al., 2021). Research on AI ethics resides within multiple domains, including but not limited to philosophy, IS, computer sciences, and social or management research (Bostrom & Yudkowsky, 2014). The renunciation of this ethical discourse, which by its philosophical and multidimensional nature tends to be controversial, can entail significant and considerable consequences and risks for our society (Boddington, 2017). With AI-based systems, it is, therefore, important to create guidelines for dealing with AI at an early stage.

In the last decade, IS researchers have focused on designing new artifacts, particularly AI applications (Ahsen et al., 2019; Kloör et al., 2018) or on examining AI applications in certain application domains such as healthcare (Mirbabaie et al., 2021a) or media distribution (Hofeditz et al., 2021). However, as AI becomes increasingly more capable, only focusing on AI's positive side can be misleading or even dangerous. Therefore, some IS scholars have begun to establish a discourse related to the ethical challenges of AI (e.g., Mendling et al., 2018; Porra et al., 2019). Primary examples of ethical considerations include the greater complexity of AI and its increasing decision-making autonomy. The complexity makes it harder to understand how and why an AI has come to a certain decision—and what decisions it will make in the future (Gunning, 2017). The increasing decision-making autonomy of AI concerns decisions that an AI is able to take on its own with little or no prior human approval or supervision (Kalenka & Jennings, 1999). A prominent concept in this context is algorithmic aversion (Berger et al., 2021; Kawaguchi, 2021; Renier et al., 2021). This phenomenon, which has been illustrated by various studies (e.g., Dietvorst et al., 2015; Dietvorst et al., 2018), shows that human decision makers tend to consider algorithmic forecasters significantly less than human forecasters, even if the humans repeatedly perform worse in the forecasts. Furthermore, decision makers tend to choose a modifiable imperfect algorithm over a non-modifiable perfect algorithm. One reason for this algorithmic aversion is the desire of individuals to have at least some level of control and autonomy (Dietvorst et al., 2018). However, this possibility for autonomy and indivisibility is not present in every AI-based system. There are also studies showing that laypeople are more likely to trust algorithms than humans for certain predictions, which can be called algorithmic appreciation (Logg et al., 2019). This shows that algorithmic aversion is not always as straightforward as it might seem, and that future research needs to look further into this and other related ethical dimensions and phenomena. Another ethically relevant area in the context of AI-based systems is trust in the system (Hofeditz et al., 2021; Mirbabaie et al., 2021a; Thiebes et al., 2021). One recently published study suggests that a loss of trust in familiar AI-based systems due to perceived errors of a familiar system over time is one possible explanation of algorithmic aversion (Berger et al., 2021).

The discourse on ethical dimensions of AI with the potential of IS to assume a leading role due to its interdisciplinary knowledge seems to be highly unstructured, and we hardly found any established theory papers in this field. We found some promising conference pieces dealing with the implementation of AI ethics in organizations (Mayer et al., 2021), a governance framework for AI regulation (de Almeida et al., 2020), and ethical implications of bias in machine learning (Yapo & Weiss, 2018). However, with an initial search, we neither found often-cited high-quality IS journal publications nor articles providing guidance for IS research on how to systematically examine the ethical dimensions of AI. In addition, some domains, such as healthcare or quality management for materials, could, from an ethical point of view, be considered more important than others. In sensible cases, ethical discourse must be discussed more compellingly compared to less sensitive cases. However, the IS discourse has not elaborated on that so far. To the best of our knowledge, the individual conclusions on the ethical dimensions and implications of AI reside within various domains, hiding a common foundation of what is known and what needs to be addressed in practice and research. The foundations of the ethical dimensions of AI seem to be widely scattered and ambiguous. Therefore, we ask the following research question:

> **RQ: What is the status quo of IS research regarding the discourse on the ethical dimensions of AI?**

Against this background, we aim to gather research from various sources, extending beyond the scope of the AIS basket of eight journals and prominent IS conferences (e.g., ICIS, ECIS, PACIS, AMCIS). To contribute to the discourse with knowledge of the ethical dimensions of AI, we identified and analyzed the domain ecosystem via the application of a discourse approach to corpus construction (Larsen et al., 2019), including consecutive forward and backward searches. Starting from an IS perspective, but also including various works from outside IS in the backward and forward search, we gathered 125 relevant papers from several disciplines and identified 12 fundamental manuscripts on the ethical dimensions of AI. By analyzing the gathered literature, we identified research gaps within the ecosystem and derived directions for IS research. With our review, we aim to provide a base for future research directions on the ethical dimensions of AI inside the IS community, hopefully jumpstarting a rich exchange between disciplines.

This paper is structured as follows. In Section 2, we highlight the importance of ethics in IS research and outline the current state of research on the ethical dimensions of AI. In Section 3, we describe why and how we used the discourse approach, according to Larsen et al. (2019). In Section 4, we summarize our findings and provide an overview of the fundamental manuscripts on the ethical dimensions of AI identified by our adapted discourse approach. We interpret these findings and discuss the role of IS research in the ecosystem of the ethical dimensions of AI in Section 5. We provide concrete contributions to IS research and highlight an avenue for future studies. We conclude with closing thoughts and a call to action in Section 6, reflecting on how scholars may build upon our results.

2 Research Background

Ethics scholarship in IS deals with various questions, such as privacy, intellectual property, employment relationships, design decisions, and the changing role of humans in society (Stahl, 2008). As early as 1985, Moor (1985) distinguished computer ethics from ethics in relation to other technologies. In this context, most research on ethics deals with normative challenges (Stahl, 2008). Thus, illegal, inappropriate, and unethical behavior is researched in the context of information technology (Leonard et al., 2001; Sojer et al., 2014). Recommendations for action, agendas, or frameworks for ethical research and practice are therefore established (Stahl, 2008; Stahl et al., 2014; Walsham, 1996).

Computer and algorithm biases are a curated ethical issue in IS research. There are several types of biases in AI technologies, such as sampling bias, which produces models relying on training data that is not representative of future cases, and performance bias, which examines performance distortion in predictions by AI (Abbasi et al., 2018). In addition, confirmation bias can lead to machine learning searches that reinforce biases, and anchoring bias can lead to incorrect assumptions about initial information provided by AI. An established classification of computer bias is a framework provided by Friedman and Nissenbaum (1996). They defined criteria such as reliability, accuracy, and efficiency by which the ethical quality of computer systems should be judged. Ethical principles were also established for specific methods of IS research. One example can be found in the ethical principles for design science research, which are: public interest, informed consent, privacy, honesty and accuracy, property, and quality of the artifact (Myers & Venable, 2014). It is important to discuss these principles because violating them can cause harm to individuals or society. Furthermore, compliance with these principles can improve social coexistence or reduce discrimination against individuals. The same principles can also be found in other contexts, such as ethical guidelines for internet communities (King, 1996). More recent research on information privacy in organizations has also considered the constructs of control, justice, and ethical obligation (Greenaway et al., 2015). These principles were transferred to concepts such as nudging (the guiding of individuals' behavior toward a beneficial choice for themselves or society) and have been expanded accordingly (Renaud & Zimmermann, 2018). The ethical principles for nudging are 1) respect (including retention and transparency), 2) beneficence, 3) justice, 4) scientific integrity, and 5) social responsibility. In the current discourse on the ethical dimensions of AI, these principles have again been used and extended to transfer them to autonomous computer systems (Floridi & Cowls, 2019). The discussion of these principles in the context of AI technologies is highly unstructured, especially in IS research, and has so far only scratched the surface of an important social challenge. There are no fundamental IS works that provide directions for future research on the ethical dimensions of AI. The conflict between ethics and AI has not been addressed sufficiently in IS research, leaving ethical issues unresolved which could impact people's lives. Problems such as privacy abuses or hate speech that have arisen in connection with social media technologies show that it is important to create ethical frameworks prior to the widespread utilization of new technology. As AI will penetrate more and more areas of professional, public, and private life in the future, it is important to prevent possible damage to society and individuals, to maximize its benefits, and to guide developments ethically. IS scholars can take a leading role in this quest due to their expertise in understanding socio-technical phenomena.

IS research has a long history of examining and ensuring the ethical use of computers and curating this knowledge (Chatterjee et al., 2009; Kallman, 1992; Stahl, 2008). Various frameworks, principles, and guidelines have been established to support researchers and practitioners in the ethical use of computers (Ess, 2009; Harrington, 1996; King, 1996; Sojer et al., 2014). Nonetheless, Stahl noted in 2008 that there were only a small number of IS papers dealing with ethics. Although the research interest in the ethics aspects of IS grows continuously, there remains a dearth of fundamental articles on emerging technologies such as AI-based technologies.

2.1 Ethics and AI

Currently, research on the ethical dimensions of AI is trending. AI ethics differs from the debate on other technologies, as AI raises different ethical questions in relation to other technological trends, such as blockchain, big data, or virtual reality (Boddington, 2017). As summarized by Russel and Norvig (2016), AI can be defined as a research stream that includes all technologies that can think or act like a human or that can think or act rationally. However, not only do the capabilities of AI-based systems continue to evolve, but so does what can be defined as AI. Currently, AI can be considered a frontier of computational advancements, capable of solving more and more complex decision problems that were once reserved for humans (Berente et al., 2021). In practice, and in most IS case studies, AI is usually considered to be a technology with self-learning abilities via machine learning, neuronal networks, or deep learning and thus performs better than a human in narrow tasks (Brynjolfsson & Mitchell, 2017; Kotsiantis, 2007). AI can thereby relieve people from repetitive processes (Dias et al., 2019). However, unlike most other technologies, AI not only threatens the jobs of employees who perform many repetitive tasks but can also replace the work of knowledge workers by being designed to make independent decisions (Boddington, 2017). Furthermore, studies show that many people already perceive AI as an independent individual (Araujo, 2018; Feine et al., 2019; Mirbabaie et al., 2021b; Seeber et al., 2020), which also raises ethical questions. In addition, the use of AI is not an exact science, since AI learns and builds on predictions (Boddington, 2017). AI technologies are usually trained on huge datasets that can hardly be traced by a human. This means that the output of AI cannot always be easily explained. Due to the complex algorithms and the huge amount of training and test data, AI takes on the form of a certain "black box" like character for humans, resulting in difficulties tracing back the outputs of AI predictions. In particular, when AI has to make important decisions that impact directly on a person’s life (such as getting credit approval or health insurance), major ethical challenges arise (Aversa et al., 2018; McNamara et al., 2018). The research on Explainable AI addresses this topic (Barredo Arrieta et al., 2020; Gunning, 2017; Miller et al., 2017). Furthermore, ethical dilemmas arise when an AI makes a challenging decision and a human is directly or indirectly harmed by it (Coppersmith, 2019). For example, in the case of an accident involving a self-driving car, the question arises as to whether responsibility for the damage lies with the developer, the supplier, the customer, or even the technology itself.

In order to address these socially relevant ethical problems with AI, governments and organizations have established guidelines and policies on how AI should be used. One example can be found in the Ethics Guidelines for Trustworthy AI, which were formed by a European Union expert committee to regulate the use of AI in European countries (EU HLEG, 2019). Other countries, such as China or Canada, also have their own guidelines for the use of AI (BAAI, 2019; Floridi et al., 2018). In addition, large organizations such as Google, Microsoft, and IBM have published guidelines (Vakkuri et al., 2019).

Research on this topic is in its early stages, needing guidance and a clear understanding of the cumulative tradition of related domains and what needs to be addressed in future research. An initial attempt to synthesize the various principles was carried out by Floridi et al. (2018). The authors identified four core risks of AI: devaluing human skills, removing human responsibilities, reducing human control, and eroding human self-determination. Furthermore, they established a framework and recommendations for a good AI society, considering the AI guidelines of various governments (e.g., The Montreal Declaration for Responsible AI) and institutions (e.g., Asilomar AI Principles) (Floridi et al., 2018). As core principles for the ethical use of AI, the same researchers identified beneficence, non-maleficence, autonomy, justice, and explicability.

2.2 Discourse on the Ethical Dimensions of AI in IS Research

Research on the ethical dimensions of AI is a broad field, including a wide range of disciplines. de Almeida et al. (2020) provided an overview of frameworks and guidelines on the ethical dimensions of AI. They carried out a systematic literature review including peer-reviewed articles from several relevant databases using keywords such as “ethics”, “how to regulate”, “risk”, and “framework”. Although they offered a broad overview of frameworks on AI ethics in IS research and beyond, their review was limited to peer-reviewed articles published in journals in a given time period using specific keywords for their identification (de Almeida et al., 2020).

However, even within IS research, empirical and theoretical works hardly differ in terms of their viewpoints on the ethical dimensions of AI. Thus, the implementation of values such as power, achievement, hedonism, stimulation, self-direction, universalism, benevolence, tradition, conformity, and security were examined (Robbins & Wallace, 2007). In addition, the problem of bias in machine learning is a trending focus in IS research (Ahsen et al., 2019; Kordzadeh & Ghasemaghaei, 2021; Yapo & Weiss, 2018) and in management practice (Martin, 2019). Other IS works on ethical dimensions of AI focus exclusively on specific application domains such as hiring (Hofeditz et al., 2022) or healthcare (Mirbabaie et al., 2021a).

Furthermore, Teodorescu et al. (2021) highlighted failures of applying fairness in human-AI augmentation, resulting in unintentional discrimination. They argued that IS scholars’ knowledge on how to address the principle of fairness for AI-based systems is limited and call for further research. On another level, Etzioni and Etzioni (2016) suggested that a new variety of AI technologies should ensure that existing AI-based systems meet ethical standards by monitoring, auditing, and holding operational AI systems accountable. Porra et al. (2019) argued that it will most likely turn out not to be beneficial for our societies if AI becomes increasingly anthropogenic. They predicted that digital assistants would outnumber humans by 2021, and, therefore, the ethical dimensions of AI should be discussed philosophically (Porra et al., 2019). In 2021, the market for digital assistants continues to grow strongly (Research and Markets, 2021).

In sum, the discourse on the ethical dimensions of AI in total, and especially in IS, is taking place on different levels of abstraction and from various empirical and theoretical angles. Previous reviews reflect parts of the big picture. However, it is unclear which manuscripts are fundamental for the research domain and which future research directions scholars should address.

3 Research Design: An Adapted Discourse Approach

To the best of our knowledge, the ethical dimensions of AI are ambiguous and the discourse on how to address the ethical issues of AI is unstructured. A systematic literature review is a method to reveal the current state of the art on a theory and to point out gaps or define a research agenda (vom Brocke et al., 2015). However, according to Larsen et al. (2019), it is hardly possible to identify and properly consider all relevant works due to the constant growth in knowledge and the sometimes very high number of publications on a topic or a theory. Therefore, we use a discourse approach, which starts from fundamental theory-building papers (L1) that derived a fundamental theory, framework, or model or that shed light on a phenomenon or a new research domain. As a second step, theory-contributing papers and papers that cited the L1 articles will be identified (L2). The third type of papers (L3) are those that influenced the L2 papers. L1, L2, and L3 form the interconnected ecosystem of a theory or domain.

We have adopted this approach for our review of the research field of the ethical dimensions of AI in order to structure the discourse and understand the ecosystem behind it. Larsen et al. (2019) did not describe in their work how they identified fundamental manuscripts (L1 papers). In our case, the fundamental manuscripts were not apparent at first. Therefore, we developed a method to be able to identify L1 papers. Our research approach consists of three phases, following the recommendations of Larsen et al. (2019). An overview of the applied research approach is provided in Figure 1 and will be presented in the following sub-sections.

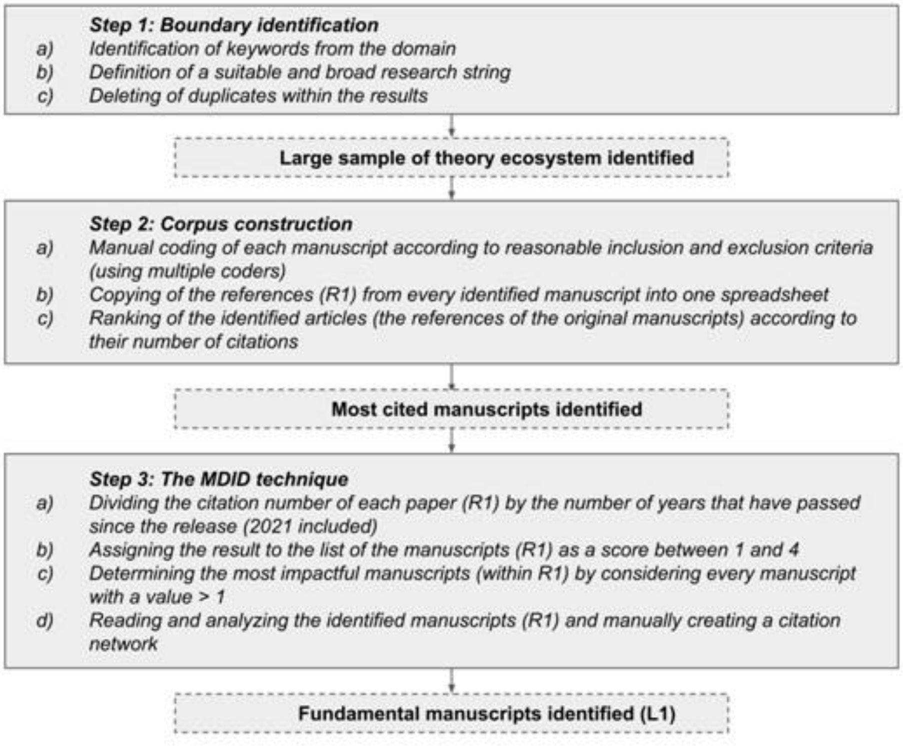
Step 1: Boundary identification
a) Identification of keywords from the domain
b) Definition of a suitable and broad research string
c) Deleting of duplicates within the results
Large sample of theory ecosystem identified
Step 2: Corpus construction
a) Manual coding of each manuscript according to reasonable inclusion and exclusion criteria (using multiple coders)
b) Copying of the references (R1) from every identified manuscript into one spreadsheet
c) Ranking of the identified articles (the references of the original manuscripts) according to their number of citations
Most cited manuscripts identified
Step 3: The MDID technique
a) Dividing the citation number of each paper (R1) by the number of years that have passed since the release (2021 included)
b) Assigning the result to the list of the manuscripts (R1) as a score between 1 and 4
c) Determining the most impactful manuscripts (within R1) by considering every manuscript with a value > 1
d) Reading and analyzing the identified manuscripts (R1) and manually creating a citation network
Fundamental manuscripts identified (L1)

Figure 1. Adapted Discourse Approach In Three Steps (Source: Larsen et al., 2019).

3.1 Boundary Identification

The first step in every literature review should be the identification of boundaries. Systematic literature reviews have been regularly criticized for not providing a comprehensive picture of a discourse (Larsen et al., 2019; vom Brocke et al., 2015). The discourse approach of Larsen et al. (2019), however, considers a research domain less as a set of characteristics but rather as a discourse between scholars. As the aim of our work was to gather a literature base and to identify the origin of the current discourse on the ethical dimensions of AI, we applied the approach in order to derive directions for IS and IS-related research. To do this, however, we did not limit our search to IS journals and conference papers alone, but also to relevant manuscripts from outside IS research. The approach is centered on so-called L1 papers, which represent the best, most cited, or most well-known papers in their respective research stream. These L1 papers are fundamental manuscripts about a theory, a model, a framework, a research domain, or a (trending) topic. For instance, Davis (1989) was mentioned as an example of a fundamental L1 paper on the Technology Acceptance Model (TAM).

Papers that want to contribute to academic discourse and develop a theory or domain should cite fundamental manuscripts (Larsen et al., 2019). From the citations of the theory-contributing papers and the papers on which these manuscripts are based, a citation network can be developed, which can be called a theory or domain ecosystem.

To define this network, the L1 papers must be identified first. However, Larsen et al. (2019) do not describe an exact process for tracking fundamental papers. In their example, the L1 paper was presented as generally known (the TAM by Davis (1986)). For new research domains and emerging fields and phenomena, however, there often is no consensus on the origin of a discourse. Thus, in the context of the ethical dimensions of AI, a predefined set of fundamental papers (L1) has yet to be identified in order to identify contributing articles (L2). Therefore, we had to modify the discourse approach in order to identify papers that can be considered fundamental for the discourse of the ethical dimensions of AI. Hence, we decided to commence by applying a “traditional” systematic keyword search but with a broad search query. We used as many synonyms as possible for terms from our domain of interest in order to follow Larsen et al. (2019), who recommended not limiting the search to a too narrow search string. However, our aim was to identify the status quo in IS research regarding the ethical dimensions of AI; therefore, our starting point for our search was based in IS research. The following search string was run through Scopus (with the help of Litbaskets.io[1] to identify relevant IS journals and IS-related interdisciplinary journals) and AISeL databases (mainly to include IS conference pieces):

(“artificial intelligence” OR “AI”) AND (“ethics” OR “ethical” OR “ethic”)

We know that there are synonyms for both artificial intelligence (such as “machine learning” or “neural networks”) and ethics (such as “morals” or “morality”), but through an upstream keyword search, we found that all articles that really discussed ethical dimensions of AI and not only one facet (such as ethics in IS or AI) contained the keywords of “artificial intelligence” and “ethics” or “ethical”. We started our initial search by choosing AISeL (mainly for IS conferences) and Scopus as a meta database (for the Basket-of-Eight journals, general IS journals, and IS-related journals) as we wanted to contribute to the ongoing discourse on the ethical dimensions of AI in IS, and these databases include all manuscripts such as journal articles, conference proceedings, and books that can be considered IS research. However, only the starting point of our search was focused on the IS discipline to identify the status quo in IS research. The further steps of the systematic search, including a forward and backward search according to Webster and Watson (2002), identified articles published outside IS. However, we deem those papers relevant as they are related to the discourse on the ethical dimensions of AI.

We performed our literature search between June and July 2020, and it was updated during the revision process. According to Larsen et al. (2019), we did not apply any filter or limitations by year. As a next step, we identified duplicates within the results. Our initial search resulted in 381 papers. After deleting duplicates, we ended up with 175 results. As none of these articles stand out by the number of citations per year, a holistic view, or the connectedness within the results, we assumed that these articles can be labeled as L2 or even L3 articles. With these results, we were still not able to understand the discourse on the ethical dimensions of AI or even determine the center of the discourse by identifying L1 manuscripts. Although the keyword search was a necessary first step to shed light on the discourse on the ethical

[1] Litbaskets is an information technology artifact supporting exploratory literature searches for information systems research (Boell & Wang, 2019).

dimensions of AI, we also assumed that our initial keyword search did not precisely cover the whole picture for a corpus of literature. Therefore, we proceeded with a more comprehensive cross-disciplinary search to understand the discourse on the ethical dimensions of AI and identified the fundamental (L1) manuscripts.

3.2 Corpus Construction

To build a corpus of literature, the manuscripts of the initial search must be considered in more detail. As is the case for all literature reviews, not all manuscripts are relevant (Larsen et al., 2019). For the next step, two independent coders filtered the papers for relevance and fit to our topic, applying inclusion and exclusion criteria. The coders manually scanned abstracts and keywords from the population identified. We included articles that added new knowledge to the discourse on AI and ethics or contributed to existing guidelines, models, or frameworks. We excluded articles that mentioned ethical aspects or AI just as a side note. To measure the intercoder reliability, we used Cohen's κ. We calculated an intercoder reliability of κ = 0.91. This step led to 125 relevant papers from the initial search. Since none of the papers still stood out, and we did not find a close connection between those articles, but because they all addressed the ethical dimensions of AI, we classified these papers as L2 (discourse contributing).

As with most research, these articles contributed to a research domain by citing and discussing certain previous works. We concluded that if all articles contribute to the discourse on ethical dimensions of AI but none of the articles within the corpus could be considered as fundamental manuscripts, fundamental works need to be among the references of those articles. Therefore, we copied all references of these 125 papers from the identified population into a list, which led to a total of 5,077 references that were no longer limited to IS research and contained manuscripts of various disciplines. L1 papers are manuscripts that should be cited in many articles addressing the discourse on a research domain. Therefore, we ranked the identified manuscripts in the reference table according to how often they were cited by the initially identified papers (which was not equal to the total number of citations, e.g., on Google Scholar or Scopus). After checking these references manually, we came up with the results presented in Table 1. These papers can be considered highly relevant for our research domain, although most of them cannot be allocated to the IS discipline. However, our aim was to understand the current interdisciplinary discourse on the ethical dimensions of AI to derive directions for IS research, and we neither knew a threshold for which articles need to be discussed in more detail nor did we consider the year of publication in relation to the number of citations within the 125 identified articles. Inspired by Larsen et al. (2019), we, therefore, developed a manual detection of implicit domain (MDID) technique to identify articles that came closest to what Larsen and his colleagues described as fundamental manuscripts for discourse in research.

Table 1. Ranking of Identified Articles According to their Number of Citations[2]

Number of citations within the 125 identified articles	Number of papers
14	1
12	2
8	4
7	5
6	1
5	6
4	19
3	64

[2] Papers that were quoted less than five times throughout the 5,077 references were omitted from this initial count due to time constraints.

Table 1. Ranking of Identified Articles According to their Number of Citations[2]

2	226
1	2105

3.3 A Manual Detection of Implicit Domain (MDID) Technique

To identify theory-contributing manuscripts in an ecosystem, Larsen et al. (2019) used an automated detection of implicit theory technique based on machine learning. However, we were not able to detect at least one L1 paper for the discourse on the ethical dimensions of AI accurately. In this article, we, therefore, developed a manual technique based on a ranking of citations among articles identified by a systematic keyword search to then be able to identify fundamental manuscripts and highlight the literature ecosystem (Larsen et al., 2019) of the ethical dimensions of AI across disciplines to guide IS research.

After reviewing the most cited manuscripts within the references of our initially interdisciplinary searched papers (not the total number of citations), we found that there was a wide time span between the publication dates of the manuscripts. However, we aimed to understand the current discourse on the ethical dimensions of AI, as for such an emerging field, a discourse can change its focus over time. Therefore, we divided the number of times articles occurred within the reference lists of the initially identified 125 articles by the number of years that have passed since the release of their first version, 2020 included. This led to a value between 0 and 4 citations per year, with only a few papers in the range of 1 to 4 and many papers at 1 or below. We considered these values as a score that describes the impact of the manuscripts on the current discourse on the ethical dimensions of AI. To be able to determine a threshold for the most relevant articles, we visualized the number of times these articles were cited by the 125 initial articles on a graph. An excerpt of this graph is provided in Figure 2, which shows the distribution of the scores of the manuscripts and some examples of paper titles.

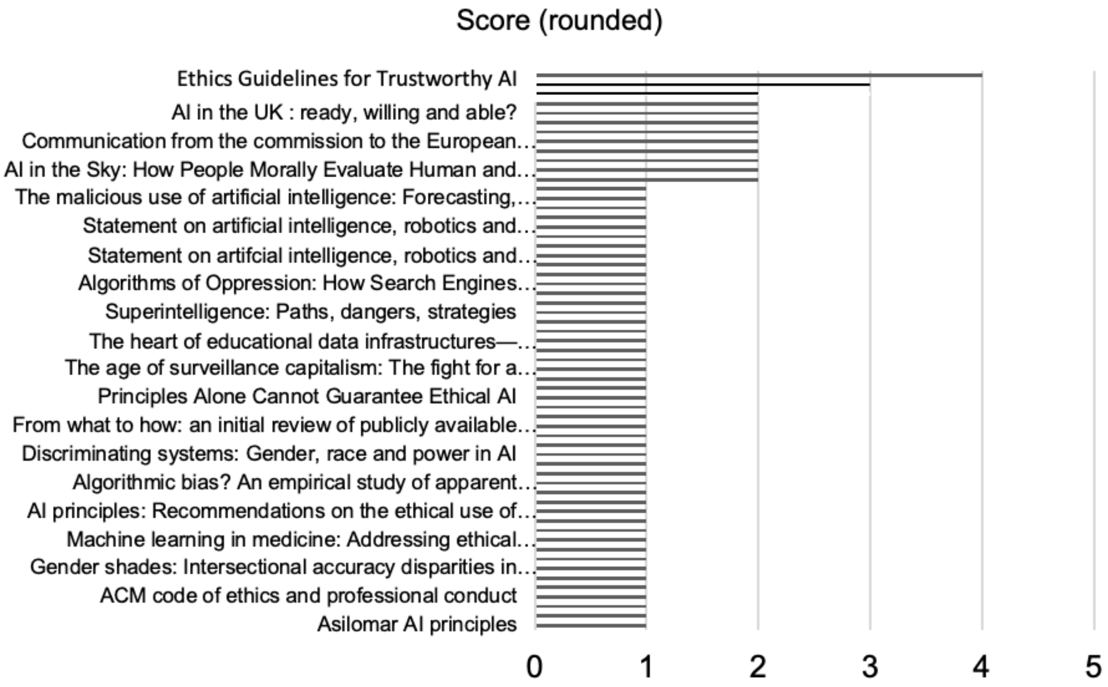

Figure 2. Visualization of an Extract from the Distribution of the Scores of the Identified Papers.

After reviewing the data and visualizing the distribution of referred manuscripts on a graph, we assessed every paper with a score above 1 to be impactful enough to be called a fundamental manuscript (as they visually stood out on the graph), which led to a total of 12 papers. Table 2 provides an overview of the manuscripts that came closest to what Larsen et al. (2019) described as fundamental L1 manuscripts. We described which artifacts were discussed in these manuscripts and compared the calculated scores with the overall citations on Google Scholar.

It was not our aim to extract the complete ecosystem by classifying every single paper in a citation network. We focused on the origin of the current discourse on the ethical dimensions of AI. However, within this process, we also identified some contributing L2 manuscripts.

Table 2. Literature Classified as L1 by Applying MDID Technique, Sorted by Score. (Status: February 2022)

ID	Consideration/Artifact	Author & Year	Outlet	Score (rounded)	Cit. in sample	Schol. Cit.
#01	Ethics Guidelines for Trustworthy AI	(EU HLEG, 2019)	EC Europe	4	8	0
#02	Ethical Framework for a Good AI Society	(Floridi et al., 2018)	Minds and Machines	3	8	679
#03	“Weapons” of Math Destruction	(O’Neil, 2016)	Broadway Books	2	12	4411
#04	AI recommendations for the UK	(House of Lords, 2018)	House of Lords (UK parliament)	2	7	0
#05	ACM’s Code of Ethics	(McNamara et al., 2018)	ESEC/FSE 2018 (conference)	2	5	105
#06	Metareview on researching algorithms	(Kitchin, 2017)	Information, Communication & Society	2	6	836
#07	Case studies for AI in military	(Malle et al., 2019)	Robotics and Well-Being	2	3	35
#08	Recommendations for AI in healthcare	(Yu et al., 2018)	Nature Biomedical Engineering	2	3	670
#09	Beijing AI principles	(BAAI, 2019)	BAAI	2	3	0
#10	Industry viewpoint and an empirical study on ethically aligned design of autonomous systems	(Vakkuri et al., 2019)	Computers & Society	2	3	22
#11	Overview of AI ethics tools, methods and research to translate principles into practices	(Morley et al., 2020)	Science and Engineering Ethics	2	3	199
#12	Ethically Aligned Design (EAD v1 & v2)	(Shahriari & Shahriari, 2017)	2017 IEEE Canada International Humanitarian Technology Conference (IHTC)	2	8	34

Furthermore, supplementing the numeric analysis, we manually checked each paper for its relevance and its role in the domain ecosystem. We extracted frameworks, guidelines, models, theories, and theory-contributing work in order to illuminate the discourse on AI ethics. In addition, we visualized how our fundamental manuscripts were cited and cited each other.

4 Results

Although we commenced our discourse approach from an IS perspective using IS databases (before we expanded our search to other disciplines through both forward and backward search), we did not find one fundamental paper published in an IS journal or in IS conference proceedings among the most-cited articles in our cross-disciplinary systematic search. Many of the most frequently mentioned manuscripts among the papers of our identified corpus were reports, books, or white papers from governmental or research institutions. With our interdisciplinary MDID technique, we also found research papers from other disciplines that could be considered fundamental for the discourse on the ethical dimensions of AI.

Before we applied our weighting of the citations, one article stood out, as it was cited 14 times by the papers that we identified with our keyword search. Turing's seminal paper on AI addressed the question of whether AI can or will ever be able to think like humans (Turing, 1950). Within his work, he introduced the "imitation game," also known as the Turing Test. Although the paper was the first seminal work on the ethical dimensions of AI, we did not consider it a fundamental L1 manuscript for the current discourse due to the score we used to weight the identified papers. Below, we discuss those manuscripts that we classified as L1 papers after applying our MDID technique.

One of the most frequently cited manuscripts we identified in our domain ecosystem was the EAD guidelines (v1 & v2) published by a committee of the IEEE Global Initiative (Shahriari & Shahriari, 2017). The document, of which there now exists an updated version, was developed based on the knowledge of several hundred leaders from six continents from academia, industry, civil society, politics, and government. Their aim was to enable ethical and social implementations of AI technologies in accordance with human values and ethical principles. Furthermore, the guidelines were intended to encourage researchers to develop new standards. Fundamental principles include the embodiment of the highest ideals of human beneficence as a superset of human rights and the prioritization of people and the natural environment when applying AI. In addition, risks and negative influences, as well as misuse, should be mitigated through transparency and accountability. As these IEEE guidelines were one of the two most prominent artifacts within our ecosystem, we classified the manuscript as L1.

The "Ethics Guidelines for Trustworthy AI" were quoted very frequently and achieved the highest score overall. The guidelines were established by the EU Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) as part of the European AI strategy (EU HLEG, 2019). The manuscript contains the Framework for Trustworthy AI, which we classify as one of the fundamental L1 frameworks for the considered research domain of AI ethics. The framework is based on four basic principles: 1) respect for human autonomy, 2) prevention of harm, 3) fairness, and 4) explicability. In addition, eight key requirements should be fulfilled before an AI can be implemented: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, and accountability. These requirements are of high importance in the field of AI, as AI technologies tend to have more autonomy in decision making and can, therefore, cause greater harm to humans than most other technologies. AI is constantly evolving, and its outputs are hardly traceable for humans, which can result in errors being detected very late.

As another fundamental manuscript on the ethical dimensions of AI, we identified the article by Floridi et al. (2018) that we already highlighted in the research background. The manuscript reports the results of the AI4People initiative, which aims to create a foundation for a good AI society. The researchers identified beneficence, non-maleficence, autonomy, justice, and explicability as basic principles for the ethical use of AI. They also formulated 20 concrete recommendations for the development, incentives, and support of good AI. The paper lists more than 200 Google Scholar citations and showed a very high relevance within the domain ecosystem (Floridi et al., 2018). Therefore, we also classified it as a fundamental L1 paper.

In "Weapons of Math Destruction," O'Neil (2017) argues that decisions affecting people's lives will increasingly be made using mathematical models (Verma, 2019). This results in less fairness, as these models are opaque, unregulated, and incontestable. The book was difficult to categorize in the domain's ecosystem, as it primarily addresses Big Data rather than AI. However, since the book has been cited frequently as a basis for further IS research and achieved a high score, we classified it as an L1 work. The two manuscripts #05 and #10 were reports and recommendations of the British and Chinese governments, respectively, on the use of AI. In #05, the recommendations of the British AI Council, the Centre for Data Ethics and Innovations, the Alan Turing Institute, and the Government Office for AI were merged into one document of guidelines (House of Lords, 2018). The recommendations for action in #10 were divided into three areas: 1) research and development, 2) use, and 3) governance. These principles were developed by the Beijing Academy of Artificial Intelligence (BAAI) and are being used by leading research institutions and organizations in China (BAAI, 2019). Therefore, we classified both #5 and #10 as L1 manuscripts.

We also found the ACM's code of ethics to be a fundamental framework (McNamara et al., 2018). We classified the code and the conference paper identified in our search as L1 manuscripts, as it achieved a score of 2. The ACM's code of ethics primarily aims at guiding researchers and practitioners in the field of computer science. The principles are divided into three sections: 1) general principles, 2) professional leadership principles, and 3) compliance with the code. They are formulated very broadly and include, for example, the following phrase: “Be fair and take action not to discriminate.” Although a study has implied that consideration of the ACM’s code of ethics has no effect on decision making, it is a fundamental manuscript for the domain ecosystem (McNamara et al., 2018).

Manuscripts #07, #08, and #09 did not consider fundamental theories, frameworks, or models of ethical AI, nor did they contribute to any of the fundamental manuscripts already identified. Rather, they discussed subdomains such as AI ethics in healthcare (Yu et al., 2019) and narrow challenges for ethical AI such as AI's ethical dilemmas in military operations (Malle et al., 2019). Nevertheless, important ethical challenges and issues regarding the use of AI were addressed, and the manuscripts achieved a high score according to the weighted citations. We found many articles (L2) that build on the findings of these manuscripts. These manuscripts address key areas that are not covered in the other L1 manuscripts. Just among the other fundamental manuscripts, they were not discussed. As they can be considered pioneering work on the ethical dimensions of AI, we classified these papers as manuscripts that came close to L1 manuscripts.

Vakkuri et al. (2019) conducted a large empirical study. They conducted a multiple case study with five organizations to demonstrate a gap between research and practice on AI ethics, further providing recommendations for closing this gap. They referred to ACM’s code of ethics, the Ethics Guidelines for Trustworthy AI, and the guidelines on Ethically Aligned Design. But even within the population of the initially identified articles, the manuscript was cited frequently and reached a score of 2. Therefore, we classified the manuscript as a theory- and domain-contributing L1 paper.

In the last manuscript we identified as fundamental, Morley et al. (2020) argue that the discourse on AI ethics focuses too much on principles and too little on practices. They also attempt to close the gap between principles and practices, referring to the conclusions and recommendations of the British House of Lords (House of Lords, 2018), the Ethical Framework for AI (Floridi et al., 2018), the Ethical Guidelines for Trustworthy AI (EU HLEG, 2019), and the framework of Ethically Aligned Design (Shahriari & Shahriari, 2017). They also refer to further guidelines and principles such as Asilomar’s AI Principles and IBM’s Everyday Ethics for AI (Morley et al., 2020). In addition, their work has been cited frequently and reached a score of 2. Therefore, we also classified this manuscript as theory contributing L1 work.

In total, we were able to classify all 12 manuscripts we identified by our adapted discourse approach as fundamental L1 papers since they either provide guidelines, principles, or frameworks on the ethical dimensions of AI or address them. However, only four of the identified fundamental manuscripts were peer-reviewed journal articles, and two were conference proceedings. Furthermore, no article published within the IS community could be recognized. Seven of the articles did not establish new frameworks but rather discussed existing guidelines and frameworks or narrow subdomains. Except for five of the papers, the manuscripts referred to at least one other L1 article or report. These five manuscripts did not refer to other fundamental papers but discussed AI ethics either on a meta level or addressed practical challenges or AI dilemmas. The 12 identified L1 manuscripts are visualized as a chronologically sorted citation network in Figure 1. The arrows indicate how the manuscripts cited each other. The citations mentioned in the figure are the Google Scholar citations from August 2020.

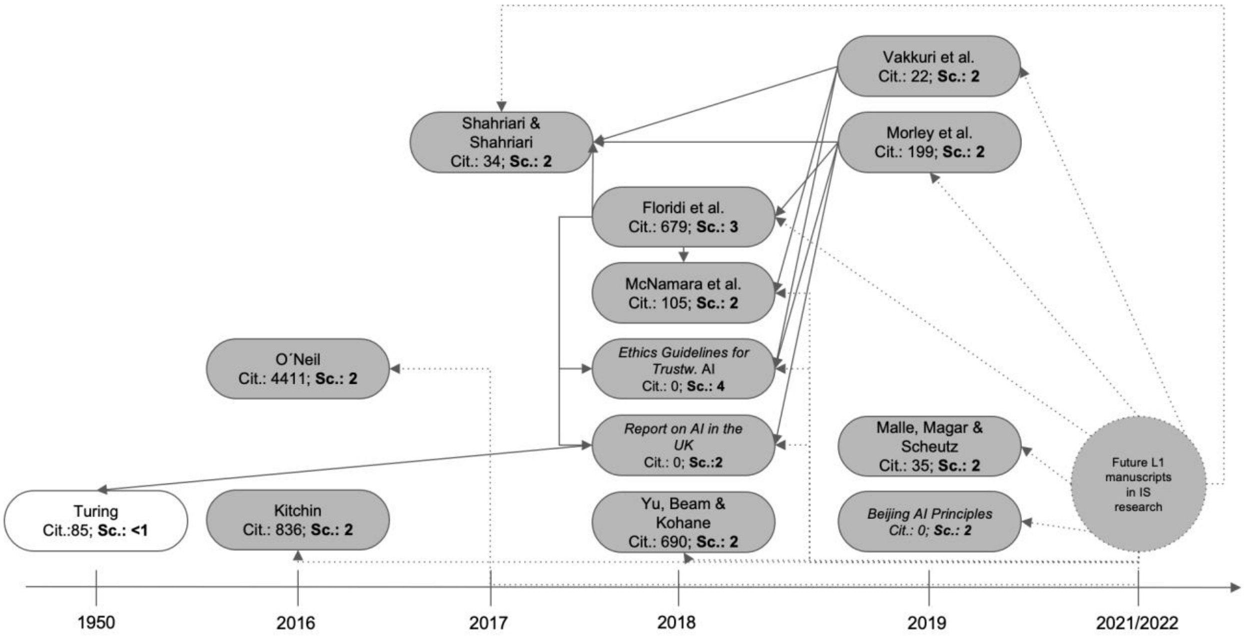

Figure 3. Domain Ecosystem for the Current Discourse on the Ethical Dimensions of AI.

The gray boxes in Figure 3 represent the manuscripts extracted from our identified corpus. The manuscripts also discuss other AI principles, such as Google's AI principles, IBM's Everyday Ethics for AI, Microsoft's guidelines for conversational bots, Intel's recommendations for public policy principles on AI, the Montreal Declaration for Responsible AI, and the Future of Life's Asilomar AI principles (Morley et al., 2020). In addition, Turing's article on the imitation game was cited the most among the considered manuscripts. However, it did not achieve a high enough score to be classified as L1, which is why we visualized the paper in a white box. The circle on the bottom right in the figure highlights possible future fundamental IS papers on the ethical dimensions of AI.

5 Discussion

Literature reviews are essential to structure an ongoing discourse or to provide research directions. Nevertheless, the method of the literature review needs to be developed further (Larsen et al., 2019; Rzepka & Berger, 2018; vom Brocke et al., 2015). The discourse approach of Larsen et al. (2019) is one of the latest methods to structure a discourse on a theory using reverse citations. In this approach, a network of citations is built from fundamental L1 manuscripts. However, as described by the authors, there is not always such a clearly defined point of origin. The discourse on the ethical dimensions of AI is such a discourse without a clear origin. Larsen et al. (2019) did not provide information on how the approach can be applied in such a case. However, since the discourse approach is based on citations, we followed this argument and offered a solution to identify fundamental manuscripts when they are initially unclear.

5.1 Discussing the Ethical Dimensions of AI

Our adapted discourse approach was well suited to identifying fundamental manuscripts on the ethical dimensions of AI. Overall, the MDID technique worked quite well to identify the most important manuscripts in the domain ecosystem. Interestingly, the papers we identified were quite different from the most cited papers on AI ethics in a simple Google Scholar keyword search. Some of the articles found by Google Scholar may also be important manuscripts; however, they rarely or never appear in the core ecosystem of the ethical dimensions of AI. It should also be noted that Google Scholar, as well as other literature databases such as Scopus, do not contain all important manuscripts for a comprehensive theory or domain ecosystem. That is why the relevance of articles within a research domain cannot be determined by citations in a database. Furthermore, often applied exclusion criteria in keyword searches, such as limitations by certain years, (specific) journal articles, or peer-reviewed articles only, lead to an incomplete picture of a discourse. Within our corpus, documents such as the Beijing AI principles and the report on AI in the UK were highly relevant, despite not being listed in the common literature databases. Thus, we agree with Larsen et al. (2019) that it is important that no manuscripts are excluded from the initial literature search. However, the score we developed not only enabled us to illustrate the discourse on a research domain and to identify L1 articles, but we could also identify the most relevant manuscripts for the current discourse.

Although we initially started our literature search in the corpus of IS, we did not find any IS journals or conference proceedings among the manuscripts we identified as fundamental. The only IS-related research article we could classify as L1 was published in a philosophy journal (Floridi et al., 2018). Overall, no single discipline can be identified that forms the origin of the current discourse on the ethical dimensions of AI. Nevertheless, we have found that most of the fundamental manuscripts originate from the disciplines of philosophy and computer science. Although one of the most important fundamental works is still Alan Turing's work on the imitation game (Turing, 1950), a new generation of fundamental manuscripts is now emerging in the domain.

We found that many of the manuscripts we classified as L1 were reports and recommendations from governments, institutions, or organizations. These contained guidelines, frameworks, principles, or recommendations for action. According to Larsen et al. (2019), we included conference proceedings and preprints in our corpus, which proved to be very valuable. We identified two fundamental manuscripts that were conference proceedings and one preprint published on arXiv that would most likely be excluded in a traditional systematic literature review process such as the one described by vom Brocke et al. (2015).

Although the identified manuscripts from our domain ecosystem refer to each other, there is no superordinate L1 paper covering the entire spectrum of the domain. The most relevant manuscripts among the 12 fundamental papers were those of Floridi et al. (2018), the Ethically Aligned Design (EAD v1 and v2), the Ethical Framework for AI, and the ACM's code of ethics. These documents have many similarities. The principles of explainability, prevention of harm, and respect for human rights are used as basic principles in most guidelines. In addition, benefits, autonomy, and justice are often mentioned, referring to the traditional principles of bioethics (Floridi et al., 2018). Some frameworks also refer to the practical readiness for AI ethics of organizations (Floridi et al., 2019). Interestingly, the AI ethics principles of the Chinese government are also strongly aligned with the values of Western cultures.

Although IS literature was not found among the fundamental manuscripts for the ethical discourse on AI, it indirectly contributed to its development. Except for the principle of non-beneficence, we found a similar counterpart for each ethical dimension of AI within the IS literature. Non-beneficence or prevention of harm also appears in a more moderate IS beneficence principle (Renaud & Zimmermann, 2018) or is described as a general "control" principle (Myers & Venable, 2014). However, the principle is particularly relevant to AI, as AI technologies are now and will continue to be given significantly more decision-making power than other technologies have ever had in the past (Floridi et al., 2018). We transferred the IS ethics principles and the ethical principles of AI into the ethical dimensions of AI that aim to guide future research and development of AI. The dimensions are visualized in Table 3.

Table 3. Comparison of IS Ethics Principles and Ethics Principles for AI.

Traditional IS ethics principles	Ethics principles for AI	Ethical dimensions of AI
Beneficence Renaud & Zimmermann (2018)	**Beneficence** Floridi et al. (2018), McNamara et al. (2018), Shahriari & Shahriari (2017) **For Humanity** (BAAI 2019)	Researching and developing AI should **contribute to the common good** and should consider privacy, dignity, freedom, autonomy, and rights of users.
Beneficence Renaud & Zimmermann (2018)	**Non-Maleficence** EU HLEG (2019), Floridi et al. (2018) **Prevent Harm to Humans** BAAI (2019), McNamara et al. (2018)	When researching and developing AI, **misuse should be prevented,** and caution should be implemented to avoid harm to humans.
Justice/Transparency/Respect Greenaway et al. (2015), Renaud & Zimmermann (2018)	**Justice/Explicability** EU HLEG (2019), Floridi et al. (2018), House of Lords (2018), McNamara et al. (2018), Shahriari &	Research and development of AI should be as fair as possible and **reduce possible discrimination**. Transparency and explainability

Table 3. Comparison of IS Ethics Principles and Ethics Principles for AI.

	Shahriari (2017) **Debiasing** BAAI (2019)	should be as high as possible in order to **prevent biases**. Make AI more explainable, predictable, traceable, auditable, and accountable.
Public Interests King (1996), Myers & Venable (2014)	**Do Good** BAAI (2019), EU HLEG (2019), Floridi et al. (2018), McNamara et al. (2018)	Researchers and developers of AI should enhance the well-being of society and ecology. Therefore, **stakeholders who may be affected need to be identified**. Security, autonomy, health, democracy, empowerment, and anticipation should be placed above features and capabilities.
Control Greenaway et al. (2015), Myers & Venable (2014)	**Autonomy** Floridi et al. (2018), EU HLEG (2019), Shahriari & Shahriari (2017), EU HLEG (2019), Floridi et al. (2018), Shahriari & Shahriari (2017)	Researchers and developers should **ensure that users have a certain level of control** when interacting with an AI.
Quality of the Artifact Greenaway et al. (2015), Myers & Venable (2014)	**Control Risks** BAAI (2019), House of Lords (2018)	Researchers and developers should **improve the maturity, robustness, reliability, and controllability** of AI systems through rigorous testing.
Responsibility King (1996), Myers & Venable (2014)	**Be Responsible** EU HLEG (2019), BAAI (2019), Floridi et al. (2018), Shahriari & Shahriari (2017), McNamara et al. (2018), House of Lords (2018)	Researchers and developers should **consider potential ethical, legal, and social impacts and risks** brought in by AI.
Scientific Integrity Renaud & Zimmermann (2018)	**Be Diverse and Inclusive** EU HLEG (2019), BAAI (2019), Floridi et al. (2018), Shahriari & Shahriari (2017), McNamara et al. (2018), House of Lords (2018)	Researchers and developers of AI should **reflect diversity and inclusiveness** and benefit as many people as possible.
Property King (1996), Myers & Venable (2014)	**Open and Shared Data** EU HLEG (2019), BAAI (2019)	Researchers and developers should **make sure that there is an agreement about the ownership** of an AI. In addition, they should establish open AI platforms to avoid data/platform monopolies.
Informed Consent Myers & Venable (2014)	**Informed Consent** BAAI (2019)	Researchers and developers should ensure that users' own rights and interests are not infringed. Therefore, **the informed consent of users should be obtained**.
Trust(worthiness) Rousseau et al. (1998)	**Trustworthiness** Morley et al. (2019), Floridi et al. (2018), AI HLEG (2019)	Researchers and developers should ensure that users **perceive a high level of trust in the AI** by meeting the seven key requirements suggested by the EU HLEG.

Despite there being no fundamental theoretical IS article on the ethical dimensions of AI, we found many similarities between ethical principles from IS research and those provided in the fundamental manuscripts on the ethical dimensions of AI. In sum, the L1 papers seem to follow ethics principles from IS research, such as those for nudging (Renaud & Zimmermann, 2018), privacy (Greenaway et al., 2015), design science research (Myers & Venable, 2014), and Internet communities (King, 1996), without directly referring to them. Floridi et al. (2018) found that the already established principles for the use of AI differed only slightly and simply added the principle of explicability to their framework. We go a step further and conclude that the ethical principles established for the ethical dimensions of AI hardly differ from the existing ethical guidelines in IS. To demonstrate this, we provide an overview of IS ethics principles for researchers and the principles contained in our L1 papers in Table 3. However, these principles are of high importance, as AI, on the one hand, is constantly evolving and, therefore, needs ethical observation. On the other hand, AI may soon permeate nearly every aspect of our lives, which is different from other technologies. Furthermore, the perception of AI differs; it can be perceived either as a tool or as a moral agent. As there are many synonyms for certain ethical principles, it is important to provide aggregated ethical dimensions of AI as a starting point for further research.

However, the principles are not clearly delineated in the literature. Even though we found overall criteria for differences in the identified principles, we also found distinct overlaps in the literature. For example, Floridi et al. (2018) concluded explicability – which has been used synonymously with explainability – AI would enable the principles of beneficence, non-maleficence, justice, and autonomy. We have highlighted these overlaps in Figure 4. Achieving trustworthy AI was described as the overarching goal in three of the fundamental manuscripts (Morley et al., 2019; Floridi et al., 2018; AI HLEG, 2019) or as one of the greatest challenges (Shahriari & Shahriari, 2017), we also consider it the most important dimension that can be enabled by respecting the other principles. Trustworthiness in AI can be achieved, for example, according to Floridi et al. (2018), if the five main criteria of beneficence, non-maleficence, justice, autonomy, and explicability are fulfilled. Moreover, these criteria were discussed by all fundamental manuscripts that addressed ethical principles for AI (see Table 3). The principles in the inner part of Figure 4, in contrast, were used in these papers to describe the principles in more detail.

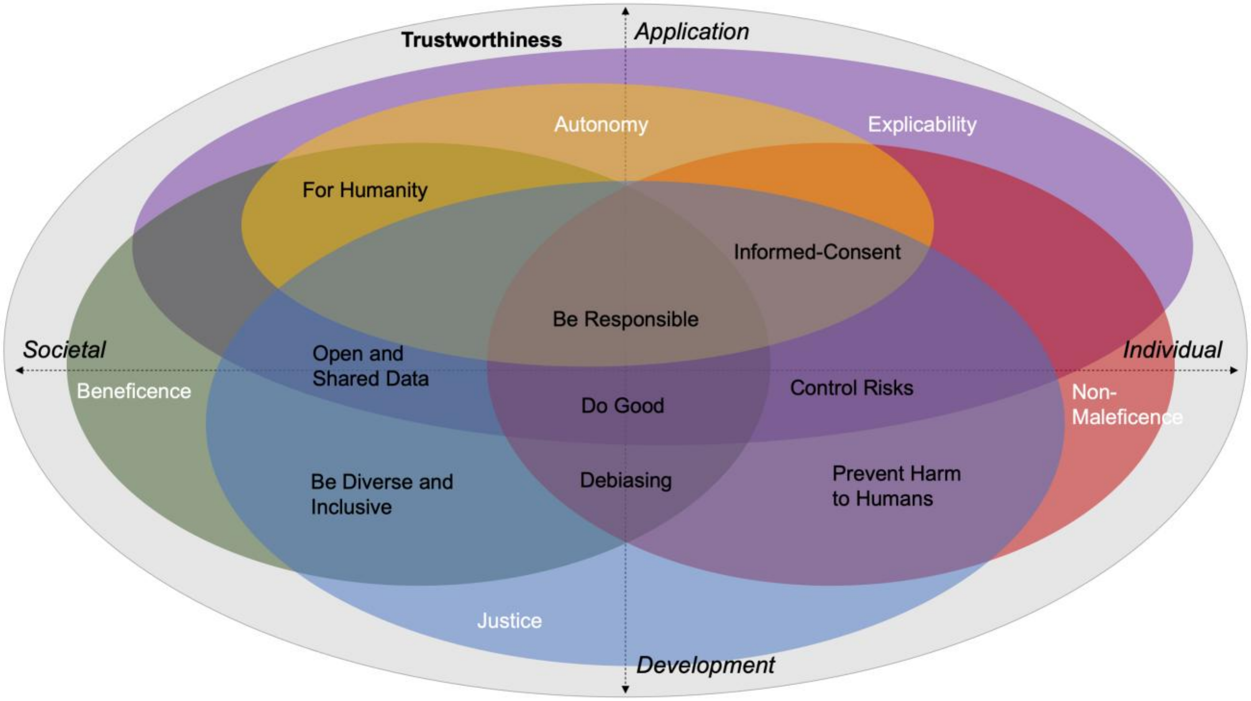

Figure 4. Classification of the Identified Ethical Principles for AI in the Dimensions Of Application, Development, Society, and Individual.

We also noticed that the ethical dimensions of AI were discussed from different perspectives. For example, some ethical principles (e.g., debiasing) refer to the development of AI-based systems and some to the application (e.g., autonomy). Also, some moral principles relate more to the impact on society (e.g., for humanity), whereas others relate more to the impact of individuals (informed consent). In Figure 4, we, therefore, classified all principles into the four dimensions “societal,” “individual,” “application,” and “development.” Even if existing ethical principles can never be unambiguously assigned to one of these dimensions, they tend to address either societal aspects or individual aspects. Even if trustworthiness can be regarded as the overriding ethical principle, subordinate principles relate either more to the applications of AI-based systems (e.g., explicability) or more to the development of AI-based systems (e.g., be diverse and inclusive).

The classified principles can also be further discussed in the context of existing literature. For example, algorithmic bias, which, according to Kordzadeh and Ghasemaghaei (2021), has not yet been investigated enough empirically, can be classified under the dimension of development and concerns both societal and individual issues. This is covered in Figure 4 with debiasing and should also be further investigated in our opinion. Phenomena such as algorithmic aversion, which was raised by Dietvorst et al. (2015, 2018), can be more closely allocated to the dimensions of individual and application, as it is related to the principle of autonomy. In contrast, algorithmic appreciation, as studied by Logg et al. (2019), can unequivocally be classified under the principle of informed consent, but also to the principle of explicability, as, for example, laypeople need to be informed of what exactly they are agreeing to when interacting with an AI-based system. The work of Leidner and Tina (2021) can rather be classified in the dimensions of individual and development, as it deals with preventing harm to humans and, thus, non-maleficence. This classification not only provides material for further discussion but also helps future research to focus on specific dimensions and explore them in more depth.

In addition to these principles of the ethical dimensions of AI for research, we identified further principles for the practical use of AI by organizations and governments. Organizations should educate and train their employees in order to improve the adaption of AI on the psychological, emotional, and technical levels (BAAI, 2019; EU HLEG, 2019; Shahriari & Shahriari, 2017). Governments should optimize employment to give full play to human advantages in order to avoid job losses and unemployment (BAAI, 2019). The Beijing AI Principles call for more cooperation, interdisciplinary work, and continuous improvement and rethinking of the principles (BAAI, 2019). Even if these aspects originally refer to governments, they can also be applied to research. Our results showed the interdisciplinary nature of research on the ethical dimensions of AI. Nevertheless, this research needs better coordination and collaboration between the different disciplines.

One question that arises is whether there are L1 papers on the ethical dimensions of AI that integrate the identified principles, guidelines, and frameworks. A clear agenda for future research on AI and ethics would also be extremely valuable. There is a lack of clear definitions and conceptualizations of what constitutes AI ethics. IS research, which otherwise addresses ethics in detail, seems disengaging and not very visible in the ecosystem of this research domain. Articles such as one by de Almeida et al. (2020) only scratch the surface of the overall discourse and offer hardly any concrete principles for the ethical dimensions of AI. Other IS articles focus more on a practical contribution rather than on a contribution to the research discourse (Martin, 2019; Robbins & Wallace, 2007). Although Porra et al. (2019) point out the importance of theoretical discourse on the ethical dimensions of AI, they do not provide concrete guidance for future research.

Therefore, we derived research questions for each ethical dimension of AI in section 5.2 to guide future IS research.

5.2 Implications for IS Research

The following implications can be derived from the interpretation of our results. First, our adapted discourse approach can be used to identify fundamental manuscripts of a current discourse based on citations and their weighting. Although we started from an IS point of view, other disciplines would find a very similar basis of L1 manuscripts in their search. Our approach provides a good starting point to identify an ecosystem of L1, L2, and L3 manuscripts.

Second, Google Scholar citations and citations in other databases are not decisive for the importance of a paper in a certain discourse, such as the ethical dimensions of AI. We found many fundamental manuscripts that had no or few citations. Other manuscripts with a high number of citations on Google Scholar or Scopus, however, could not be identified as fundamental to the considered discourse.

Third, to avoid biases, it is important that non-peer-reviewed manuscripts, conference articles, and other forms of documents are included in the search. Among the fundamental manuscripts, we found conference papers, reports, and white papers from governments and institutions. Thus, a literature search should not only focus on selected journals such as the Basket-of-Eight or a specific time period; otherwise, important papers cannot be identified.

Fourth, the discourse on the ethical dimensions of AI in IS remains fragmented and without a clear structure. So far, there are no fundamental manuscripts from IS that are directly linked to the general interdisciplinary discourse. IS literature refers to publications from the fields of philosophy and computer science as fundamental manuscripts. However, there are many similarities between the traditional ethics principles in IS research and the ethical principles of AI.

Fifth, most fundamental manuscripts on the ethical discourse in relation to AI refer to each other. However, there is no research article that links all existing principles and guidelines and discusses them in a scientifically sound manner, although Floridi et al. (2018) are very close to that. Other fundamental manuscripts, however, are not connected to other relevant papers and opened their own sub-discussions within the discourse.

Sixth, since AI technologies are constantly evolving, there cannot be universally valid and permanent principles that adhere to all ethical dimensions of AI. Existing principles and guidelines need to be continuously revised and supplemented.

5.3 Directions for Future Research

Following Pienta et al. (2020), we identified research questions and directions for IS research for each ethical dimension of AI. With these research questions, we do not claim to create an exhaustive list. Rather, we offer initial questions referring to each dimension that can be used by IS scholars as a starting point for discussion and further questions. We derived the questions from an interpretation of the future research chapters of the traditional IS literature on ethical principles and from the 12 fundamental manuscripts that we were able to identify using our manual detection method. As an example, one important question regarding the dimension of informed consent of users could be how AI can be designed by internal parties and third parties to ensure that users' rights and interests are recognized. The research questions and research directions were classified according to our identified ethics principles for developing and using AI-based systems. With related ethical themes, we provided a higher level of abstraction, which relates back to the classification in Figure 4. Figure 4 focuses primarily on the visual classification of the principles and themes in the four dimensions: societal, individual, application, and development, as well as the relationship of the principles to each other, and offers material for further scientific discourse. In accordance with Figure 4, we also show the main and tendency dimensions for the principles in a table. Table 4, in contrast to Figure 4, goes a step further and offers concrete research questions and directions for future research. The ethical themes of beneficence, non-maleficence, justice, autonomy, and explicability build on the principle of classification by Floridi et al. (2018), and the overarching principle of trustworthiness was derived from AI HLEG's (2019) discussion on trustworthy AI. The research questions and directions for IS research are shown in Table 4.

Table 4. Guiding IS Research on the Ethical Dimensions of AI by Providing Exemplary Research Questions and Directions.

Ethical Theme	Ethics principles for AI	Dimensions	Possible research directions and questions
Beneficence	Benefit Humanity	Societal, Application, & Development	**Example research questions:** • *What positive effects can be achieved for society using self-driving shuttle services in smart cities?* • *How can intelligent assistance systems be used in hospitals to relieve nurses of their workload and allow them to spend more time with their patients?* **Directions for IS research:** • Conduct design science research on new societal AI applications in healthcare or governance. • IS lecturers need to teach their students not only commercial AI applications, but also societal applications.
Non-Maleficence	Prevent Harm to Humans	Individual, Application, & Development	**Example research questions:** • *Which tasks and decision-making functionalities should not be delegated to AI-based systems to prevent harm to humans?* • *What are the design principles for AI in recruiting that help to prevent harm to applicants?* **Directions for IS research:** • Conduct quantitative research on misuse of AI applications through organizations and highlight how harm to humans can be prevented and human digital dignity can be preserved. • IS lecturers need to increase awareness of possible misuse of AI and teach how caution can be implemented into AI-based systems.

Table 4. Guiding IS Research on the Ethical Dimensions of AI by Providing Exemplary Research Questions and Directions.

Justice Explicability	Act Debiasing	Societal, Individual, & Development	**Example research questions:** • *What explanations lead to the understanding of an AI-based conversational agent by elderly people?* • *How do journalists in media organizations need to be trained to avoid data bias from an AI being used?* **Directions for IS research:** • Conduct qualitative research on mechanisms that lead to more fairness and earlier detection of biases in data used by an AI and ensure a high level of transparency, explainability (explicability), predictability, traceability, and accountability for study participants and your paper's audience. • Lecturers need to teach strategies and approaches for reducing possible discrimination (e.g., through training data) of AI-based systems.
Beneficence Non-Maleficence Justice Explicability	Do Good	Societal, Individual, Application, & Development	**Example research questions**: • *How can AI-based systems be used on social media platforms to detect and counteract fake news and misinformation?* • *What can we learn from green IS to develop green AI applications that support sustainable use cases?* **Directions for IS research:** • Identify stakeholders such as employees or customers that could be affected by (future) AI introductions (in qualitative studies) and develop targeted applications for these groups (in design science studies). • IS lectures and seminars should not be limited to the features and capabilities of AI, such as certain machine learning or deep learning algorithms, but also teach awareness of ethics and the most important application fields for societal issues.
Explicability	Ensure Autonomy	Societal, Individual, & Application	**Research questions:** • *What functionality needs to be built into self-driving vehicles to enable manual occupant intervention?* • *How can remote organizations mitigate algorithmic control to provide more autonomy for their employees?* **Guidance for IS research:** • Conduct behavioral research on the effects of algorithmic control, algorithmic aversion, and algorithmic appreciation on employees and provide guidelines to mitigate negative effects. • IS lecturers need to teach how students can design AI-based systems that provide a high level of user control.
Non-Maleficence Justice Explicability	Control Risks	Individual, Development	**Research questions:** • *What precautions can organizations take to provide the highest possible level of security and prevent cyberattacks on an AI-based system?* • *Which robustness checks do emergency management organizations need to apply before using an AI-based system in crisis communication?* **Guidance for IS research:** • Before applying an AI-based system in a study or in practice, conduct a risk analysis to control the maturity, robustness, reliability, and controllability of AI systems. • Modules for controlling AI risks and cyber threats need to be created in study programs at universities and technical colleges.
Beneficence Non-Maleficence	Be Responsible	Societal, Individual, & Application	**Research questions:** • *Which preconditions need to be established by institutions before applying AI-based systems in education?* • *How can uncertainty among employees in organizations*

Table 4. Guiding IS Research on the Ethical Dimensions of AI by Providing Exemplary Research Questions and Directions.

Justice Autonomy Explicability			*be mitigated before and after an AI-based change process?* **Guidance for IS research:** • Conduct qualitative and quantitative research on the social effects of the introduction of AI-based systems and provide guidance on how to mitigate risks. • IS decision makers (such as professors or heads of departments) should provide supplementary lectures and seminars on legal and social responsibility in organizations to establish grounding knowledge among students.
Beneficence Justice	Be Diverse and Inclusive	Societal, Development	**Research questions:** • *How can an AI-based system be used by governments to distribute information in a wide range of languages in order to better include minority groups in society?* • *How does an AI-based system need to be designed to enable people with speech disorders to comfortably communicate with non-disabled people?* **Guidance for IS research:** • Conduct qualitative and design science research on how AI needs to be trained to reflect diversity and inclusiveness and benefit as many people as possible, e.g., by recruiting study participants of minority groups or by designing targeted AI solutions. • Lecturers need to teach their students how they can reflect on diversity and inclusiveness when designing and developing AI-based systems. In addition, lectures need to address accessibility criteria for AI-based systems.
Beneficence Justice Explicability	Open and Share Data	Societal, Application, & Development	**Research questions:** • *How can blockchain technologies be used to share the ownership of AI-based systems in order to avoid data monopolies?* • *How can digital nudging be applied to engage researchers and developers of AI to establish open AI platforms and share AI-related data?* **Guidance for IS research:** • When researching or developing AI-based systems, build on open-source solutions and share your data. • Lecturers need to teach open access and open-source AI frameworks instead of teaching commercial solutions to increase awareness of open data and open science.
Non-Maleficence Justice Autonomy Explicability	Obtain Informed Consent	Individual, Application	**Research questions:** • *How can AI policies of third parties be intertwined with informed consent for AI use?* • *Which criteria do hospitals need to include in their consent forms for applying AI-based systems for supporting treatment decisions?* **Guidance for IS research:** • When conducting qualitative or quantitative research on AI, ensure that informed consent of participants is obtained and develop templates for informed consent for AI applications. • Lecturers need to teach students how to design consent forms for applying AI-based systems.
Trustworthiness	Achieve Trustworthiness	Societal, Individual, Application, & Development	**Research questions:** • *How can ethical principles be applied to conversational agents to increase the trustworthiness of public institutions during crisis events?* • *How can the seven key requirements suggested by the AI HLEG be implemented in AI-based systems to achieve a high level of trust in AI?* **Guidance for IS research:**

Table 4. Guiding IS Research on the Ethical Dimensions of AI by Providing Exemplary Research Questions and Directions.

			• Conduct interdisciplinary research on how ethical cues can be implemented in AI-based systems (such as conversational agents) to achieve a high level of trust in the system. • As trustworthiness is an overarching principle for ethical AI, lecturers need to establish courses on how to increase trust in AI-based systems.

5.4 Limitations

There are some limitations to our work. Overall, we mainly analyzed 12 manuscripts in detail. Since our primary goal was to identify fundamental manuscripts on the ethical dimensions of AI, we did not further examine L2 and L3 papers. There is also a chance that there are a few more L1 papers pertaining to the current discourse that we could not identify with our approach. For AI, there are many synonyms, and our initial keyword search was limited to rather broad search terms.

We used our adapted discourse approach for the first time and determined the threshold for the manuscripts that we classified as fundamental by visualizing the distribution curve of all calculated values. While this could lead to a small number of unknown L1 manuscripts, we were able to identify a very high percentage of L1 papers.

In addition, ethics and AI is a rapidly and constantly evolving research topic in IS research and beyond. Our work reflects the state of research from July 2020 and does not contain literature that was published after that time. In addition, we limited our initial literature search to the IS databases litbaskets.io and AISeL. Future research is needed to confirm whether we are correct in assuming that the identified L1 papers can also be considered fundamental manuscripts for other disciplines.

6 Conclusion

Every theory or emerging domain needs fundamental manuscripts marking the origin of a research field and enabling a critical discourse. For the ethical dimensions of AI, we were able to identify 12 fundamental manuscripts following an adapted discourse approach according to Larsen et al. (2019). We identified the manuscripts using a broad keyword search and a score based on weighted citations of the initially retrieved papers. We found not only journal publications, but also reports, white papers, and conference proceedings that we classified as relevant to the current discourse. None of these fundamental papers were based in IS research. Therefore, we derived concrete directions for future IS research and exemplary research questions. Nevertheless, many concepts from IS ethics research overlap with the various ethical principles of AI. Transparency, beneficence, autonomy, responsibility, justice, and scientific integrity were often attributed to ethical AI conduct. However, in IS, these principles have been examined in non-AI contexts such as nudging, research on privacy issues, or virtual collaboration for decades. Therefore, we derived the ethical dimensions of AI based on IS ethics principles and the ethical principles for AI in order to guide researchers and developers.

When carrying out research on AI, we recommend following the depicted principles of AI ethics. Our research agenda in Table 4 could serve as a starting point for this. As an interdisciplinary discipline, IS could provide a valuable L1 manuscript, synthesizing and extending the existing principles and frameworks not only for the IS community, but also for related disciplines such as economics, social science, computer science, cognitive science, and psychology. Future research should refer to and critically examine the fundamental manuscripts we have identified. For this, AI development, research, use, and its impact on different stakeholders should be considered more closely by IS scholars.

Furthermore, the IS community has the potential to contribute additional relevant key artifacts. It is especially important to increase the number of peer-reviewed research articles and to ensure that the fundamental manuscripts are not limited to government or corporate documents. IS research could utilize its fundamental knowledge on normative ethics that has already been gathered to discuss the ethical dimensions of AI in more detail. IS scholars could use previous knowledge, for example, from the fields of nudging (Renaud & Zimmermann, 2018), from research on ethics in Internet communities (King, 1996), or from research on privacy issues (Greenaway et al., 2015). Thus, future IS research could produce further fundamental papers that provide guidance for scholars of different disciplines, considering the 12 fundamental manuscripts we identified in this article. In sum, there is a lack of a general theory that explains the complex ethical dimensions of AI. Based on the identified fundamental manuscripts, IS scholars could derive such a theory.

Another important direction for future research is to further identify the ecosystem of the current discourse. Further research could uncover L2 and L3 manuscripts and their connections to L1 papers. To this end, the discourse approach of Larsen et al. (2019) could be continued.

Furthermore, there is a lack of frameworks to guide the ethical management of AI in profit and non-profit organizations. Here, again, the IS community could draw on its previous knowledge in the areas of IT strategy and the management of digital processes to create a scientific foundation. When AI is applied, it usually impacts the environment, and therefore, people and societies. If, for example, AI is used by NGOs or media organizations, the effects on society and people need to be examined more closely.

As a supplementary direction, the ethical dimensions of AI should be further investigated at a detailed level. Future research should investigate how and whether people are influenced by AI that behaves unethically. In experiments and field studies, preventive measures could be derived to prevent unethical behavior and negative effects on society and individuals.

Overall, it can be concluded that IS research on the ethical dimensions of AI is still in its infancy. Nevertheless, based on the existing knowledge on (computer) ethics in IS, there is great potential for future research, which should be exploited.

References

Abbasi, A., Jingjing, L., CLifford, G., & Taylor, H. (2018). *Make "fairness by design" part of machine learning*. Harvard Business Review. Retrieved from https://hbr.org/2018/08/make-fairness-by-design-part-of-machine-learning

Ahsen, M. E., Ayvaci, M. U. S., & Raghunathan, S. (2019). When algorithmic predictions use human-generated data: A bias-aware classification algorithm for breast cancer diagnosis. *Information Systems Research*, *30*(1), 97–116.

Almeida, P., Santos, C., and Farias, J. S. (2020). Artificial intelligence regulation: A meta-framework for formulation and governance. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 5257–5266).

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189.

Aversa, P., Cabantous, L., & Haefliger, S. (2018). When decision support systems fail: Insights for strategic information systems from Formula 1. *Journal of Strategic Information Systems*, *27*(3), 221–236.

BAAI. (2019). *Beijing AI principles*. BAAI. Retrieved from https://www.baai.ac.cn/news/beijing-ai-principles-en.html

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115.

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems*, *22*(2), 281–303.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433–1450.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch me improve—Algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering*, *63*(1), 55–68.

Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer International Publishing.

Boell, S., & Wang, B. (2019). www.litbaskets.io, an IT artifact supporting exploratory literature searches for information systems research. In *Proceedings of the Australasian Conference on Information Systems 2019*.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (Vol. 316, pp. 316–334). Cambridge University Press.

Brendel, A. B., Mirbabaie, M., Lembcke, T. B., and Hofeditz, L. (2021). Ethical management of artificial intelligence, sustainability. *Sustainability 13*(4), 1–18.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534.

Carvalho, A., Levitt, A., Levitt, S., Khaddam, E., & Benamati, J. (2019). Off-the-shelf artificial intelligence technologies for sentiment and emotion analysis: A tutorial on using IBM natural language processing. *Communications of the Association for Information Systems*, *44*(1), 918–943.

Chatterjee, S., Sarker, S., & Fuller, M., (2009). A deontological approach to designing ethical collaboration. *Journal of the Association for Information Systems*, *10*(10), 138–169.

Coppersmith, C. W. F. (2019, April 10). *Autonomous weapons need autonomous lawyers*. The Reporter. Retrieved from https://reporter.dodlive.mil/2019/04/autonomous-weapons_law/

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003.

de Almeida, P. G. R., dos Santos, C. D., & Silva Farias, J. (2020). Artificial intelligence regulation: A meta-framework for formulation and governance. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Dias, M., Pan, S., & Tim, Y. (2019). Knowledge embodiment of human and machine interactions: Robotic-process-automation at the Finland government. In *Proceedings of the 27th European Conference on Information Systems*.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*, *144*(1), 114–126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

Ess, C. (2009). Floridi's philosophy of information and information ethics: Current perspectives, future directions. *The Information Society*, *25*(3), 159–168.

Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. *Ethics and Information Technology*, *18*(2), 149–156.

EU HLEG. (2019). *Ethics guidelines for trustworthy AI*. European Commission. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, *132*, 138–161.

Floridi, L., & Cowls, J. (2019). *A unified framework of five principles for AI in society*. Harvard Data Science Review. Retrieved from https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347.

Greenaway, K. E., Chan, Y. E., & Crossler, R. E. (2015). Company information privacy orientation: A conceptual framework: Company information privacy orientation. *Information Systems Journal*, *25*(6), 579–606.

Gunning, D. (2017). *Explainable artificial intelligence* (XAI) (pp. 1–17) [Distribution Statement "A"]. DAPPA. Retrieved from https://sites.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf

Harrington, S. J. (1996). The effect of codes of ethics and personal denial of responsibility on computer abuse judgments and intentions. *MIS Quarterly*, *20*(3), 257.

Hofeditz, L., Mirbabaie, M., Holstein, J., & Stieglitz, S. (2021). Do you trust an AI-journalist? A credibility analysis of news content with AI-authorship. In *Proceedings of the European Conference on Information Systems 2021*.

Hofeditz, L., Mirbabaie, M., Luther, A., Mauth, R., & Rentemeister, I. (2022). Ethics guidelines for using AI-based algorithms in recruiting: Learnings from a systematic literature review. In *Proceedings of the Hawaii International Conference on System Sciences*.

Horton, H. (2016, March 24). *Microsoft deletes "teen girl" AI after it became a Hitler-loving sex robot within 24 hours*. The Telegraph. Retrieved from https://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/

House of Lords. (2018). *AI in the UK: Ready, willing and able?* Authority of the House of Lords. Retrieved from https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Kalenka, S., & Jennings, N. R. (1999). *Socially responsible decision making by autonomous agents*. Cognition, Agency and Rationality (pp. 135–149). Springer.

Kallman, E. A. (1992). Developing a code for ethical computer use. *Journal of Systems and Software*, *17*(1), 69–74.

Kawaguchi, K. (2021). When will workers follow an algorithm? A field experiment with a retail business. *Management Science*, *67*(3), 1670–1695.

King, S. A. (1996). Researching internet communities: Proposed ethical guidelines for the reporting of results. *The Information Society*, *12*(2), 119–128.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29.

Kloör, B., Monhof, M., Beverungen, D., & Braäer, S. (2018). Design and evaluation of a model-driven decision support system for repurposing electric vehicle batteries. *European Journal of Information Systems*, *27*(2), 887–927.

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388-409.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engeneering*. IOS Press.

Larsen, K. R., Hovorka, D. S., Dennis, A. R., & West, J. D. (2019). Understanding the elephant: The discourse approach to boundary identification and corpus construction for theory review articles. *Journal of the Association for Information Systems*, *20*(7), 887–927.

Leidner, D. E., & Tona, O. (2021). The CARE theory of dignity amid personal data digitalization. *MIS Quarterly*, *45*(1), 343–370.

Leonard, L., & Cronan, T. (2001). Illegal, inappropriate, and unethical behavior in an information technology context: A study to explain influences. *Journal of the Association for Information Systems*, *1*(1), 1–31.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

Malle, B. F., Magar, S. T., & Scheutz, M. (2019). AI in the sky: How people morally evaluate human and machine decisions in a lethal strike dilemma. *Robotics and well-being* (pp. 113–133).

Mann, G., & O'Neil, C. (2016). *Hiring algorithms are not neutral*. Harvard Business Review. Retrieved from https://hbr.org/2016/12/hiring-algorithms-are-not-neutral#:~:text=Don't%20let%20the%20software%20screen%20out%20good%20candidates.&text=More%20and%20more%2C%20human%20resources,pool%20of%20potential%20job%20candidates

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

Mayer, A., Haimerl, A., Strich, F., & Fiedler, M. (2021). How corporations encourage the implementation of AI ethics. In *Proceedings of the 29th European Conference on Information Systems*.

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development? In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018* (pp. 729–733).

Mendling, J., Decker, G., Reijers, H. A., Hull, R., & Weber, I. (2018). How do machine learning, robotic process automation, and blockchains affect the human factor in business process management? *Communications of the Association for Information Systems*, *43*(1), 297–320.

Miller, T., Howe, P., & Sonenberg, L. (2017). Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences. In *IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI)*.

Mirbabaie, M., Hofeditz, L., Frick, N. R. J., & Stieglitz, S. (2021a). *Artificial intelligence in hospitals: Providing a status quo of ethical considerations in academia to guide future research*. AI & Society. Retrieved from https://link.springer.com/article/10.1007/s00146-021-01239-4

Mirbabaie, M., Stieglitz, S., Brünker, F., Hofeditz, L., Ross, B., & Frick, N. R. J. (2021b). Understanding collaboration with virtual assistants – The role of social identity and the extended self. *Business & Information Systems Engineering*, *63*(1), 21–37.

Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, *16*(4), 266–275.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168.

Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management*, *51*(6), 801–809.

Porra, J., Lacity, M., & Parks, M. S. (2019). Can computer-based human-likeness endanger humanness? A philosophical and ethical perspective on digital assistants expressing feelings they can't have. *Information Systems Frontiers*, *22*, 533-547.

Renaud, K., & Zimmermann, V. (2018). Ethical guidelines for nudging in information security & privacy. *International Journal of Human-Computer Studies*, *120*, 22–35.

Renier, L. A., Schmid Mast, M., & Bekbergenova, A. (2021). To err is human, not algorithmic – Robust reactions to erring algorithms. *Computers in Human Behavior*, *124*, 106879.

Research and Markets. (2021). *Global digital assistant market* (2021 to 2026)—*Featuring Amazon, Apple and Baidu among others*. Research and Markets. Retrieved from https://www.globenewswire.com/news-release/2021/12/20/2355021/28124/en/Global-Digital-Assistant-Market-2021-to-2026-Featuring-Amazon-Apple-and-Baidu-Among-Others.html

Robbins, R. W., & Wallace, W. A. (2007). Decision support for ethical problem solving: A multi-agent approach. *Decision Support Systems*, *43*(4), 1571–1587.

Robin, E. (2019). *Artificial intelligence: Conflicts of interest between ethics and the needs of an adapted regulation*. Curiosity. Retrieved from https://medium.com/emmanuelle-robin/artificial-intelligence-conflicts-of-interest-between-ethics-and-the-needs-of-an-adapted-d2c1512e0bc9

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited.

Rzepka, C., & Berger, B. (2018). User interaction with AI-enabled systems: A systematic review of IS research. In *Thirty Ninth International Conference on Information Systems*.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174.

Seppälä, A., & Mäntymäki, M. (2021). From ethical AI principles to governed AI. In *International Conference on Information Systems*.

Shahriari, K., & Shahriari, M. (2017). IEEE standard review — Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)* (pp. 197–201).

Sojer, M., Alexy, O., Kleinknecht, S., & Henkel, J. (2014). Understanding the drivers of unethical programming behavior: The inappropriate reuse of internet-accessible code. *Journal of Management Information Systems*, *31*(3), 287–325.

Stahl, B. C. (2008). Researching ethics and morality in information systems: Some guiding questions. In *Proceedings of the International Conference on Information Systems*.

Stahl, B. C. (2012). Morality, ethics, and reflection: A categorization of normative IS research, *Journal of the Association of Information Systems, 13*(8), 636–656.

Stahl, B. C., Eden, G., Jirotka, M., & Coeckelbergh, M. (2014). From computer ethics to responsible research and innovation in ICT. *Information & Management*, *51*(6), 810–818.

Stieglitz, S., Mirbabaie, M., Kroll, T., & Marx, J. (2019). “Silence” as a strategy during a corporate crisis – The case of Volkswagen’s “Dieselgate.” *Internet Research*, *29*(4), 921–939.

Teodorescu, M., Morse, L., Awwad, Y., & Kane, G. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly*, *45*(3), 1483–1500.

Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, *31*(2), 447–464.

Turing, A. (1950). *Computing machinery and intelligence*. Mind.

Vakkuri, V., Kemell, K.-K., Kultanen, J., Siponen, M., & Abrahamsson, P. (2019). Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study.

Verma, S. (2019). Weapons of math destruction: How big data increases inequality and threatens democracy. *Vikalpa: The Journal for Decision Makers*, *44*(2), 97–98.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37*.

Walsham, G. (1996). Ethical theory, codes of ethics and IS practice. *Information Systems Research*, *6*, 69–81.

Webster, J., & Watson, T. R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2), xiii-xxiii.

Wiener, M., Cram, W., & Benlian, A. (2021). Algorithmic control and gig workers: A legitimacy perspective of Uber drivers. *European Journal of Information Systems*, 1–23.

Yampolskiy, R. V. (2016). Taxonomy of pathways to dangerous artificial intelligence. In *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*.

Yapo, A., & Weiss, J. (2018). Understanding the impact of policy, regulation and governance on mobile broadband diffusion. In *46th Hawaii International Conference on System Sciences* (pp. 2852–2861).

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(1), 717–731.

About the Authors

Milad Mirbabaie is junior professor for Information Systems & Digital Society at Paderborn University and team leader for Sociotechnical Systems at the University of Duisburg-Essen, Germany. He studied Information Systems at the University of Hamburg and received his PhD from the University of Münster, Germany. He has published in reputable journals such as Journal of Information Technology, Business & Information Systems Engineering, Electronic Markets, Journal of Decision Systems, Internet Research, Information Systems Frontiers, International Journal of Information Management, and International Journal of Human Computer Interaction. His work focuses on Sociotechnical Systems, AI-based Systems, Social Media, Digital Work, and Crisis Management.

Alfred Benedikt Brendel is associate professor for business information systems, esp. intelligent systems and services, at the Technische Universität Dresden. Alfred holds a Doctor's degree in management science, specializing in Business Information Systems, from the University of Goettingen. His research focuses on exploring the human-like design of conversational agents and its effect on users' perception, affection, cognition, and behavior. His main areas of research are digital health, smart mobility, and digital work. His research is forthcoming or has been published in leading information systems journals, including *Journal of Information Technology*, *Journal of the Association for Information Systems*, *Information Systems Journal*, and *European Journal of Information Systems*.

Lennart Hofeditz is a research associate at the research group of Professor Stefan Stieglitz at the University of Duisburg-Essen, Germany. He studied Applied Cognitive and Media Science (M.Sc.). Now, he is a PhD candidate in Information Systems at the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen in Germany. In his research, he focusses on socio-technical systems and ethical issues related to the application of artificial intelligence and anthropomorphic machines in organizations. He also works in a research project funded by the German Research Foundation (DFG) on research data management and open science.

Copyright © 2022 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints are via e-mail from publications@aisnet.org.

View publication stats