6-18-2022

Ethics and Morality in AI - A Systematic Literature Review and Future Research

Jan-Lukas Selter
University of Siegen, jan-lukas.selter@uni-siegen.de

Katja Wagner
University of Siegen, katja.wagner@marketing.uni-siegen.de

Hanna Schramm-Klein
University of Siegen, schramm-klein@marketing.uni-siegen.de

ETHICS AND MORALITY IN AI – A SYSTEMATIC LITERATURE REVIEW AND FUTURE RESEARCH

Research Paper

Jan-Lukas Selter, University of Siegen, Siegen, Germany, jan-lukas.selter@uni-siegen.de

Katja Wagner, University of Siegen, Siegen, Germany, katja.wagner@uni-siegen.de

Hanna Schramm-Klein, University of Siegen, Siegen, Germany, hsk@marketing.uni-siegen.de

Abstract

Artificial intelligence (AI) has become an integral part of our daily lives in recent years. At the same time, the topic of ethics and morality in the context of AI has been discussed in both practical and scientific discourse. Either it deals with ethical concerns, concrete application areas, the programming of AI or its moral status. However, no article can be found that provides an overview of the combination of ethics, morality and AI and systematizes them. Thus, this paper provides a systematic literature review on ethics and morality in the context of AI examining the scientific literature between the years 2017 and 2021. The search resulted in 1,641 articles across five databases of which 224 articles were included in the evaluation. Literature was systematized into seven topics presented in this paper. Implications of this review can be valuable not only for academia, but also for practitioners.

Keywords: Ethics, Morality, Artificial Intelligence, Systematic Literature Review.

1 Introduction

Applications from voice assistants and platforms such as Siri, Alexa, Netflix or YouTube are based on artificial intelligence (AI) and are increasingly finding their way into everyday life (Antao, 2019). Here, AI has established in almost all application areas, such as construction industry (Pillai and Matus, 2021), telecommunications (Macnish and Ana, 2019) and automotive industry (Cunneen et al., 2020). AI can also be used beneficially in the healthcare sector (Togni et al., 2021) or in the legal system (Simshaw, 2018).

AI offers users a wide range of benefits through increased computing power at lower costs and the usability of big data (Huang and Rust, 2021; Haenlein and Kaplan, 2019). For example, AI can analyze consumer emotions and script an advertisement based on them (Huang and Rust, 2021), generate a personalized playlist on platforms such as Spotify based on collected data (Puntoni et al., 2021), or automate the clothing buy-back process through a chatbot (Schanke et al., 2021). However, the development of AI has been accompanied not only by benefits but also by ethical problems, such as the prioritization and thus discrimination of certain consumers based on demographic and economic aspects in customer relationship management (Libai et al., 2020) or the unequal representation of providers and companies on platforms such as AirBnB or Amazon, leading to unequal market shares (Milano et al., 2021). In addition, with the support of AI, attempts have already been made to manipulate people by placing appropriate political advertisements based on social media data (Kaplan and Haenlein, 2019) and to monitor and log their user behavior (Du, 2021). Entrepreneurs like Elon Musk and Bill Gates also see risks in using AI, such as losing control over its use (Steels and Lopez de Mantaras, 2018). Here, autonomous weapons systems and associated issues regarding human safety (Hynek and Solovyeva, 2021) and ethical concerns with surveillance technologies (Belk, 2021) are mentioned. Thus, addressing ethical implications of the application of AI is important due to the increasingly consequential interaction between humans and AI (McManus and Rutchick, 2019). Due to these controversial aspects, the question for creating an ethical design of AI-based solutions (Morley et al., 2020) arises. The increasing development of ethics-related

projects in technology associations (Adamson et al., 2019) and the development of ethical guidelines and principles by, for example, the EU Commission (EU Commission, 2020) or by China (Wu et al., 2020a) also highlight the international awareness of the relevance of ethics and morality in the context of AI development.

Up to our knowledge, the current development of AI and the accompanying public interest have resulted in various research directions in the last years. However, scientific articles only provide insights into individual research areas in the context of AI and ethics, while a holistic overview of the topic is still lacking (Loureiro et al., 2021). The literature addresses e.g., ethical guidance (Vakkuri et al., 2020), different application areas (Hamilton and Davison, 2021), ethical governance (Shneiderman, 2020), or ethical aspects in programming (Arvan, 2018). However, no paper could be identified that deals with a systematization of the different research foci and an identification of the relevant articles in this context. For future research, however, such a systematic review of the literature is necessary to identify research gaps and new research opportunities (Briner and Denyer, 2012). Implications of such an overview can be valuable not only for academia, but also for other stakeholders involved in the development, application, or commercialization of AI-based products and services. Thus, this research aims to analyze literature regarding ethical and moral aspects in the context of AI in order to: (i) identify and systemize existing literature; (ii) provide an overview about the discussed topics in this context; (iii) highlight future research directions with specific research questions. So, our research questions are: RQ 1: What is the current state of literature for ethics and morality in the context of AI? RQ 2: What are future research areas in the context of ethics and morality with AI?

2 Conceptual Framework

This section provides a brief introduction in the relevant concepts of this study and an overview of the keywords used in this research. Thus, a short definition of AI with the subdivision into machine learning and deep learning will be provided. In addition, the context of morality and applied ethics and here machine ethics, which play a central role in the context of AI, will be introduced.

Since AI has been established in several contexts, different definitions are used (Wang, 2019). For example, an early definition of AI indicates that the term is used to refer to the action of a computer when that action, performed by a human, would require intelligent thought. This implies that the capability of intelligence is not limited to humans or other living beings, but is also related to computers (Simon, 1995). Therefore, the term AI is also used to describe the multidisciplinary research in the fields of computer science, systematic neuroscience, and human cognitive science (Cichocki and Kuleshov, 2021). In general AI can be subdivided into machine learning (ML) and deep learning (DL). Machine learning as one part of AI represents the ability of systems to learn and improve automatically and independently of humans as long as the data is structured (Campbell et al., 2020). Deep learning as another part of AI and basically a subarea of ML is based on neural networks, being computational networks that are biologically inspired and operates on the functionality of the human brain (Haenlein and Kaplan, 2019). Furthermore, there are two types of AI. “Weak AI” means it can only outperform humans in specific tasks, being the more common use of AI for now. ”Strong AI” could be more powerful than humans in almost all cognitive tasks. However, it currently has no practical application (Lu et al., 2018).

Considering the second part of this research, the concept of ethics and morality, it encompasses e.g., the field of applied ethics. Applied ethics includes machine ethics (Bauer, 2020), which is of particular importance in the discourse on AI. This can be divided into two research directions. One direction includes robot ethics or AI ethics when talking about the avoidance of harm through the application of robots and AI. This line of research refers to the behavior of developers, manufacturers, and users. The other direction includes the terms of machine ethics, ethical AI, or ethical robots and refers to research into ethical behavior of robots and AI themselves, as well as research into the implementation of ethical values in autonomous systems. Machine ethics is the first research field that deals with artificial morality (Winfield et al., 2019). Artificial morality on the other hand explores the possibilities of moral capabilities of artificial systems (Misselhorn, 2018). Therefore, the discourse addresses the moral status of AI discussing moral theories that are directed toward moral patients and moral agents. The concept of morality, as distinct from the concept of

ethics, is mostly directly related to decisions and actions (Verbeek, 2008). In the context of ethics, the question of what is 'right' action is considered (Fritz et al., 2019), whereas morality comprises rules, values, and norms that influence a person's behavior (Bartneck et al., 2019). Nevertheless, it should be pointed out that in the scientific discourse there is not always a clear conceptual separation between ethics and morality. The adjectives ethical and moral are used as synonyms in contrast to fixed terms such as AI ethics and moral agency. For instance, the expressions ethical machines (Tolmeijer et al., 2020) and moral machines (Magrani, 2019) correspond analogously.

3 Methodology

The systematic literature analysis was conducted in accordance with Briner and Denyer (2012) and subsumed in five steps. After identifying the research questions, search strings, search and selection criteria were defined for selected databases, the inclusion and exclusion protocol was set which included e.g., the publication period, and type of articles, followed by the analysis and synthesis and report (see Figure 1).

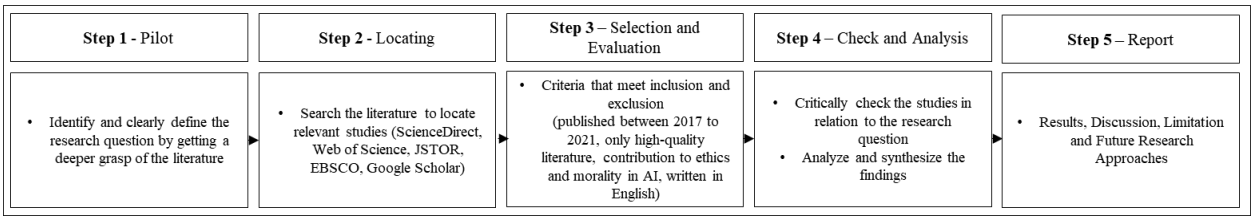

Figure 1. Research process of systematic literature review

The databases and Boolean Operators can be found in Table 1. For a current and at the same time comprehensive overview of the literature, the publication period of the years from 2017 to 2021 was chosen, because almost 90 percent of the articles relevant to the topic were published from 2019 to 2021. Between 2020 and 2021 alone, 155 of the final 224 papers selected were published. According to RQ1, this study aims to indicate the current state of literature of ethics and morality in the context of AI. Since especially AI is a fast-growing sector and a fast-developing technology, it is important to include the latest trends.

Database	Search String	Results
ScienceDirect	Title, abstract or author-specified keywords: artificial intelligence OR ai; Title: moral OR morality OR morals OR morally OR ethic OR ethically OR ethics OR ethical	66
	Title, abstract or author-specified keywords: moral OR morality OR morals OR morally OR ethic OR ethically OR ethics OR ethical; Title: artificial intelligence OR ai	107
Web of Science	moral* OR ethic* (Title) and ai OR artificial intelligence (Topic) or ai OR artificial intelligence (Title) and moral* OR ethic* (Topic)	807
JSTOR	(ti:(moral* OR ethic*) AND (ai OR artificial intelligence)) AND la:(eng OR en)	152
	(ti:(ai OR artificial intelligence) AND (moral* OR ethic*)) AND la:(eng OR en)	74
EBSCO	TI (ai or artificial intelligence) AND AB (ethic* or moral*) OR TI (ethic* or moral*) AND AB (ai or artificial intelligence)	235
Google Scholar	allintitle: ai moral OR ethic OR morally OR morality OR morals OR ethics OR ethical OR ethically	100
	allintitle: artificial intelligence moral OR ethic OR morally OR morality OR morals OR ethics OR ethical OR ethically	100

Table 1. Search functions and results of the selection of research articles.

The type of article was filtered in ScienceDirect for review articles, research articles, mini reviews, in Web of Science for proceeding papers, review articles, early access, in JSTOR for journal articles and in EBSCO for scientific journals. The search resulted in a total of 1,641 articles across all databases, taking into account all search criteria. Followed by a two-stage selection process for the inclusion or exclusion of articles in the literature analysis, articles that were found several times, were not peer-reviewed, and could not contribute to answering the research questions on the basis of the title and abstract were excluded. Besides, duplicate articles, book reviews, chapters, case reports, discussions, editorials, and news articles are also not included

as well as pure reproductive paper with no own contribution. Thus, 468 articles were selected for a review of the full text. In the second selection process, the impact factor which is understood as a quality indictor of the scientific journals was examined (Briner and Denyer, 2012) and articles with a two-year impact factor below 0.5 (Analytics, 2018) were sorted out. Of the 468 articles, a further 66 were excluded. During the full-text review, care was taken to ensure that the papers considered the aspects of ethics and/or morality as well as AI, and especially that both aspects were considered in relation to each other. Thus, 99 articles with a lack of focus on ethics and morality, 5 articles with insufficient focus on AI, and 22 articles with a lack of linkage between the topics were eliminated. In a final step, 52 technical articles that did not provide concrete research results were eliminated. So, 224 articles meet the criteria and form the basis for the presentation of the current state of the scientific literature on the topic of ethics and morality in the context of AI. Each paper was read by two reviewers in order to ensure the required quality and combats potential bias. We checked for an intercoder-reliability for the selected criteria throughout the process with about 100 abstracts. These were discussed, hold against the set criteria, and compared. In case of disagreement the issues were solved (Miles and Hubermann, 1994). Furthermore, it can be noted that the selected articles were published in a total of 109 different journals. In order to provide an overview of some of the dominant journals for researching the topic of this study, the journals in which more than three of the selected articles were published are listed: AI & SOCIETY (28), Ethics and Information Technology (17), Science and Engineering Ethics (14), Minds and Machines (13), Big Data & Society (6), The ORBIT Journal (6), Computer Law & Security Review (5), International Journal of Social Robotics (4), IEEE Technology and Society Magazine (4), and AI Magazine (4). In order to systematize the research articles, research methods, regions, results and focal points of the 224 contributions were protocolled and systematically analyzed for similarities and differences. From the protocol, it was found that systematization of the articles based on an amalgamation of identical or similar research foci is best suited to provide an understandable overview of the current state of the literature and to answer the underlying research question. During the review of the literature, it became apparent that the topic of ethics and morality in the context of AI is discussed from a wide variety of perspectives and in different application areas. For a suitable systematization that reflects this heterogeneity of the literature, seven main topics were identified, which in turn are subdivided into different contents and are presented in table 2.

4 Results

The protocol indicated that disadvantages and concerns as well as advantages and opportunities of AI were addressed from an ethical and moral perspective. Since these corresponding articles depict immediate implications of the development and application of AI, they are considered in the chapter "Ethical and moral implications of AI". In the second category, findings to an ethical approach to AI that have been highlighted and discussed in recent years are explained, and in the third category, demands for governance and regulation of AI are reflected. In addition, requirements for programming that have been made in recent years were frequently addressed. The following category includes articles in which human moral judgment was explored more deeply in the discourse around AI. Other research foci resulting from the protocolling includes discussions of moral responsibility for consequences of an AI decision or action and discussions of the moral status of AI. Finally, articles that address law or policy against the backdrop of ethics, morality, and AI were identified (see Table 2).

Topic	Subtopics and Content
Ethical and Moral Implications of AI	ethical concerns and benefits in the context of AI
	ethical and moral implications in the context of AI in different application areas (businesses, health sector, public sector)
	implications of physical products, digital products and implications without specific application area, use or technology
	Implications on relationship of humans to machines
Approaches of an Ethical Handling of AI	recommendations about the use of ethics in the context of AI
	principles of dealing with AI
	further approaches and human rights

Demand for Governance and Regulations	models and principles to implement existing principles as a bias for ethical governance
	governance for AI on political level
	need for regulation of AI
AI Programming Requirements	approaches of moral and moral justification as well as user and stakeholder values to be programmed in AI
	programming of a friendly AI, recognizing moral actions
	integrating rebel agents
	further technical and social challenges
Human Moral Judgement and Moral Responsibility	moral judgement of humans and human moral values
	moral responsibility of developers, providers and users
	moral responsibility of AI itself
Moral Status of AI	Discussion about moral status of AI
	Moral status of AI in the future
Importance of Law and Politics	Importance of legal system to enforce regulations
	Legislative gap and policies of various countries around the world

Table 2. Systematization of the literature on ethics and morality in the context of AI.

Ethical Implications of AI

The review of the literature has shown that many articles deal with direct implications of the development and application of AI and specifically with advantages and disadvantages as well as problems and opportunities through AI from an ethical and moral perspective. Ethical concerns that are addressed include invasion of privacy, unauthorized data collection (Truby and Brown, 2021), hacking (Swarte et al., 2019) or disadvantages to social groups as a result of biased data and a lack of transparency of AI (Pillai and Matus, 2021). In turn, literature is also concerned with ethics-related benefits in the use of AI, such as the satisfaction of customer needs through its use in marketing (Hermann, 2021b), the potential effectiveness of using AI to prevent suicide (Luk et al., 2021), and the stress reduction associated with replacing human soldiers with autonomous unmanned aircraft (Swarte et al., 2019).

Furthermore, systematization of the literature has shown that three application areas particularly are discussed for ethical and moral implication of AI. These are implications by businesses, for example for back-end or on-site tasks in the construction sector or for supportive tasks in HR management (Hamilton and Davison, 2021; Hermann, 2021a, 2021b; Pillai and Matus, 2021; Macnish and Ana, 2019; Mark, 2019b), implications by the health sector, like comparing CT images in the radiology (Luk et al., 2021; Montemayor et al., 2021; Sand et al., 2021; Stewart et al., 2021; Brady and Neri, 2020; Fenech and Buston, 2020; Morley et al., 2021; Stokes and Amitabha Palmer, 2020; Fiske et al., 2019; Kretzschmar et al., 2019; Park et al., 2019; SFR-IA Group and CERF, 2018) and the public sector, for understanding complaints of citizens or using chatbots for parking permits (Chounta et al., 2021; Szocik and Abylkasymova, 2021; Calvo, 2020; Wakunuma et al., 2020; Swarte et al., 2019; Liu et al., 2019; Mark, 2019a; Mark and Anya, 2019; Vallejos et al., 2017). In addition to application areas, the implications of specific applications or technologies, divided into physical and digital products, are also discussed in the literature. Themed physical products include caregiver robots (Yew, 2021) and robots for police applications (Szocik and Abylkasymova, 2021), autonomous vehicles (Vrščaj et al., 2020; Cunneen et al., 2019a) and devices or toys for children (Antle and Kitson, 2021; McStay and Rosner, 2021). Digital applications, for example, include online surveillance systems (Nigam et al., 2021) and biometric facial recognition (Smith and Miller, 2021), recommender systems (King, 2020; Milano et al., 2020), digital twins (Popa et al., 2021; Truby and Brown, 2021), intelligent assistance technologies (Wangmo et al., 2019) or risk assessment tools (Liu et al., 2019).

Moreover, some of the articles considered are directly dealing with implications without referring them to application areas or technologies. They discussed the impact of AI on several aspects like environment (Murdock, 2018), human rights (Livingston and Risse, 2019), human moral status (Danaher, 2019) and further referred to ethical doubts and benefits when using AI. One study identified AI to raise major concerns about an increase in unemployment and inequality (Ghotbi et al., 2021). Further, it has been argued that the decision-making power of an AI can change the moral norms of humans (Gill, 2020). Although this study raised doubts (Novak, 2020), it was suggested that AI can promote unethical human

behavior (Köbis et al., 2021). In contrast, it has been highlighted that AI can train moral cognitive abilities as well as human moral action, thus promoting moral decision making (Lara and Deckers, 2020). Moreover, the possible collaboration of AI with humans has been demonstrated (Seeber et al., 2018), with AI being able to change the relationship of humans to machines at affective, relational, and physical levels, as well as expand the physical, mental, and physical scope of humans (Togni et al., 2021). Affective relationships between humans and AI were further affirmed with regard to emotional AI. Here, the importance of being aware of the differences with human relationships was emphasized (Weber-Guskar, 2021).

Approaches of an Ethical Handling of AI

Analysis of the literature has shown that articles deal with different directions such as recommendations or principles that enable a handling of ethical and moral aspects in the context of AI. For example, Floridi et al. (2018) suggested recommendations for European policy makers called "Recommendations for a Good AI Society". Further publications have formulated recommendations explicitly for companies on how to consider ethical aspects when developing AI (Vakkuri et al., 2020) and specifically on how to solve ethical problems and develop their corporate social responsibility (Du and Xie, 2021). Few authors have been identified who have focused their research entirely on making recommendations for specific application contexts of AI (Jotterand and Bosco, 2020; Asaro, 2019; McKernan et al., 2018). For the most part, recommendations for specific application contexts have been made (Hamilton and Davison, 2021; Brady and Neri, 2020; Calvo, 2020; Yew, 2021; Fiske et al., 2019; Kretzschmar et al., 2019).

Recommendations can be distinguished from established principles in the discourse about AI. Principles for dealing with AI are most often mentioned in papers that address an application in the health sector (Stewart et al., 2021; Abràmoff et al., 2020; Larson et al., 2020; Joerin et al., 2020; SFR-IA Group and CERF, 2018). Principles for the entire context of AI ethics have also been established in recent years. In their research, Floridi et al. (2018) brought together existing principles and identified following main principles for an ethical treatment of AI: beneficence, harm avoidance, autonomy, justice, and explicability. These principles have also been taken up in follow-up research (Floridi et al., 2021) and were discussed by other authors (Milossi et al., 2021). Scholars have also expressed criticisms about the principles, such as contradictions or dependencies (Hermann, 2021a; Robbins, 2019). Similarly, other recent contributions have pointed out the limitations of using principles for the context of AI ethics (Mittelstadt, 2019; Whittlestone et al., 2019) and the challenge of implementing ethical principles in the context of AI (Stix, 2021; Peters et al., 2020), as well as the contradiction of ethical principles with the concept of ethics itself (Rességuier and Rodrigues, 2020).

While these principles have similarities globally, the approaches and motivation to implement them are rather heterogeneous (Jobin et al., 2019). Following these guidelines, it is not binding (Buruk et al., 2020) and it is merely a response to public demands (Kerr et al., 2020) without concrete contextual reference (Hagendorff, 2020). The literature has already criticized models regarding the applicability of the principles (Bærøe et al., 2020) and shown that guidelines need to be adapted for the needs of the organization (Vakkuri et al., 2020) and for regional and cultural ethical preferences of stakeholders (Segun, 2021a; Gabriel, 2020). In recent years, efforts have been made in academia to help organizations to understand guiding principles and how to consider ethical issues with AI (Mark and Stahl, 2021; Clarke, 2019).

In addition to these principles and guidelines, other approaches have been discussed in literature to integrate various ethical aspects in different AI applications. For example, Sekiguchi and Hori (2020) have developed a tool for engineers to support the implementation of ethical design of technologies through comprehensibility and standardization. They have shown that this can promote engineers' creativity and long-term projects. Furthermore, literature has also explicitly considered how users can be protected. For example, countering AI bias technically (Tomalin et al., 2021) and ensuring compliance with ethical principles in the context of social networks and countering the spread of fake news (Salem et al., 2020). Kelley and Atreides (2020) developed an assessment protocol that takes into account the specificity of the emotions and emotional structures of this AI. In contrast to approaches targeting engineers and researchers already working in this field, a paper has been identified that targets teachers of AI ethics and offers case studies helping students to understand ethical and moral aspects of AI (Burton et al., 2017). Other research

is not exclusively addressed to researchers or developers, but also to users of AI. For example, an ethical model has been created as a thinking note approach to support AI development and implementation (Schrader and Ghosh, 2018). Ethics-based auditing, and thus a third party, has been used to check the behavior of an AI for compliance with standards and principles (Mökander and Axente, 2021; Mökander and Floridi, 2021). Furthermore, models for AI-related management decisions (Brendel et al., 2021), business analysis (Vidgen et al., 2020), and ethical decision making (Anshari et al., 2021) have been developed. One model already in use in the field is the "Data Ethics Decision Aid", which focuses on the ethical use of data (Franzke et al., 2021). In addition, approaches have been developed to fundamentally sensitize stakeholders to ethical considerations, e.g., through an ethical checklist in the context of mental health (Mörch et al., 2020), a questionnaire in radiology (Geis et al., 2019), in the context of cyber-physical-human systems (Khargonekar and Sampath, 2020), or in the military domain (Wasilow and Thorpe, 2019).

While these approaches deal with ethical aspects, there are a few proposals to base the conception, development and use of AI on human rights principles rather than ethics. The articles justify the reference to human rights with the fact that there are already agreements on these formulated in laws worldwide and thus a uniform level for regulating AI already exists (Gibbons, 2021). Moreover, this approach considers accountability and power asymmetries and is complementary to ethics (Fukuda-Parr and Gibbons, 2021). Therefore, taking into account human rights, Mantelero (2018) has developed an assessment model and, based on it, a management tool to assess AI development and its impact. This is already being used in companies and could serve as a legal tool in the regulation of AI (Mantelero and Esposito, 2021).

Demand for Governance and Regulation of AI

Another key theme identified in the systematic literature review is that ethical models and values are difficult to understand and implement in the complex context of the technologies and therefore governance and regulations are required. These problems are said to be overcome by social and organizational governance mechanisms supported by external actors (Raab, 2020). For example, Shneiderman (2020) or Winfield and Jirotka (2018) have developed models to implement existing principles as a basis for ethical governance. Moreover, it has been emphasized that governance models are needed not only in relation to AI development, but also in relation to its application (Lepri et al., 2021). This is evident in the development of governance models for AI use in healthcare (Reddy et al., 2020) and formulated principles to support data governance (Janssen et al., 2020). In addition, it has become clear that governance for AI also occurs at a political level (Radu, 2021; Wu et al., 2020b) and thus the role of government in the governance process should be considered alongside the scientific community (Almeida et al., 2021).

Although governance has been called upon to solve implementation problems of other models and tools, a variety of models can be observed (Liu and Maas, 2021; Stahl et al., 2021a). It has been pointed out that for governance development, organizations should first understand their responsibility for ethical behavior (Abrams et al., 2019) and that multiple stakeholders should be involved in the development process (Liu et al., 2019; Floridi, 2018). Complementary governance development requirements exist in terms of political governance and the need for international cooperation among governments. There have been calls for political governance not to be dominated by the United States, Europe, and China, but to consider all societies affected by AI (ÓhÉigeartaigh et al., 2020).

Even there is a link between governance and regulation, governance does not necessarily go hand in hand with regulation (Winfield and Jirotka, 2018), so developed models are not binding (Stahl et al., 2021b). Thus, there is increasing emphasis on the need for regulation of AI (Almeida et al., 2021; Fernandes et al., 2020), criticism of the difficulties in implementing principles (Larsson, 2020; Iphofen and Kritikos, 2021), and recommendations to follow regulations from the beginning (Floridi, 2018). Furthermore, the importance of iteration and public discussion has been identified (Cunneen et al., 2019a), emphasizing that public interests cannot be protected by market forces alone (Ho et al., 2020), thus clarifying expressed calls for AI regulation (Truby and Brown, 2021; Kriebitz and Lütge, 2020; McStay, 2020; Pillai and Matus, 2021; Truby, 2020; Gruson et al., 2019; Wirtz and Müller, 2019).

Additionally, Robbins (2020) has noted, that responsible control and regulation of AI is not possible until there is sufficient knowledge about the nature of AI's functions, benefits, learning data, inputs and outputs, and limitations. This may justify that a balance between bottom-up and top-down approaches, between ethical approaches and regulations, has already been recommended (Hermann, 2021a).

AI Programming Requirements

Considerations and approaches to programming ethics into AI are another attempt to respond to ethical and moral concerns and issues and have also emerged as a research focus in literature. Arvan (2018) criticizes previous approaches in three ways. Either they are too flexible, too inflexible, or allow AI to pick up and replicate human misbehavior. As a solution, Arvan (2018) suggests programming AI the ability to simulate the past and future. Other scholars also called for developing a version of imagination for AI programming (Pinka, 2021; Umbrello, 2020). Tajalli (2021) pointed out that developing AI's ability of thinking would reduce the chance of AI acting badly. Because no moral theory is able to address the full complexity of the issue, it has been recommended that the focus should be on avoiding immoral actions for now, that a pluralistic approach is preferable to a single theory (Gordon, 2021), and that various moral preferences and decision rules of society should be incorporated into AI (Martinho et al., 2021). The difficulty of resolving disagreements about ethics has led to calls to develop a system that takes into account people's different values (Bogosian, 2017), thus integrating different values of users and other stakeholders into AI (Umbrello, 2020; Umbrello et al., 2021; McDougall, 2019). Such approaches require human monitoring and continuous revision, according to van de Poel (2020). In contrast to these value-based approaches, other proposals for programming ethics into AI consist of combining existing approaches (Cvik, 2021; Noothigattu et al., 2019), e.g., using an extended utility function of AI to choose between multiple actions (Vamplew et al., 2018). Another consideration is enabling AI to evaluate the specific situation in terms of ethics (Benzmüller et al., 2020), which can also map the flexibility of human moral judgment of a situation (Dubljević, 2020).

Fröding and Peterson (2020) suggested a design approach which can be distinguished from the approaches shown and makes values and utility functions secondary by being based on programming a friendly AI that mimics aspects of friendship and evolving through learning (Fröding and Peterson, 2020)(Fröding and Peterson, 2020)(Fröding and Peterson, 2020). Rather than pre-programming all aspects of action, these evolve through learning processes (Fröding and Peterson, 2020). This approach has already met with criticism, as the programming of this AI would need to be adapted to the humans acting with it and would require stricter, situation-specific legal regulations (Li, 2021). In addition, designers have been criticized for making decisions about the AI's ability to learn, and thus incorporate their ethical views into the programming (Baum, 2020; Sood, 2018). Furthermore, literature is dealing with artificial agents that recognize moral actions and link them to rewards (Haas, 2020), with simulating ethical situations (Arnold and Scheutz, 2018), and with software reflecting personal user preferences regarding ethics (Autili et al., 2019). However, a long-term approach to protect against manipulation has not yet been found (Osório and Pinto, 2019). Other technical and social challenges which consider the programming requirements include establishing trust in artificial moral agents, as well as responsible research and development and public involvement (Cervantes et al., 2020a). Thus, the need to equip machines with the ability of explaining themselves was highlighted (Nallur, 2020; Khrais, 2020). In addition, van Berkel et al. (2020) showed that users have different preferences in setting standards for programming. This was also illustrated by a study by Awad et al. (2018) with preferences of people from over 200 countries. However, the last-mentioned study has been criticized for neglecting fundamental ethical principles, basic law, human dignity, equality, and legal standards (Kochupillai et al., 2020). In relation to such thought experiments, it has also been noted that the future environment of AI and its behaviour cannot be predicted in the design phase because the design is based only on incomplete information (Héder, 2020; Borenstein et al., 2019). The selected literature does not provide concrete approaches to enable the required involvement of users (Martin, 2017) and ethicists (Gordon, 2020; Segun, 2021b), except for that of Awad et al. (2018). However, proposals have already been made to consider selected human rights norms in the design of AI (Aizenberg and van den Hoven, 2020).

Human Moral Judgement and Moral Responsibility

Another research direction that has emerged from literature review is the study of human moral preferences, as well as moral responsibility. For example, Frank et al. (2019) showed that moral preferences depend on the decision situation and the person's personal perspectives. These results are partially reflected in the study by Wolff et al. (2019), who found that moral decisions depend on neural and psychological factors and are, e.g., influenced by reaction time and the subjectively perceived distress of a situation. Furthermore, an influence of moral preferences and individual characteristics on an AI's perception of a moral dilemma was found (Rhim et al., 2021). Similarly, a qualitative study found that human moral values may depend on context, culture, and emotional state (van Berkel et al., 2020), showing that human judgments about decision situations and preferences for particular decisions of an AI depend on various factors.

In recent years, the discussion of moral responsibility for the consequences of actions enabled or performed by AI has also been taken up in scientific articles. Burton et al. (2020) referred to these considerations as the responsibility gap, which arises from the risk that neither the developers nor the providers or users can be held morally accountable for harm. Lauwaert (2021),on the other hand, argued that this gap does not exist, and that one person always bears responsibility. According to Martin (2019) and Héder (2020), this responsibility belongs to the companies and designers who developed the algorithms. In contrast, Rochel and Evéquoz (2021) argue that developers can only be held responsible for active misconduct and that a new definition of responsibility is required. Developers also see moral responsibility on themselves, but beyond that, on people providing input on AI, on users, and on the machine itself (Orr and Davis, 2020; van der Waa et al., 2021). However, few scholars have considered the possibility that AI might bear moral responsibility in the future (Sebastián and Rudy-Hiller, 2021; Tigard, 2021). It has been pointed out that only strong AI meets the requirements to be morally responsible itself (Smith and Vickers, 2021), and that only humans can bear this responsibility (D'Acquisto, 2020). In contrast, studies by Hohenstein and Jung (2020) and Shank and DeSanti (2018) have shown that humans consider AI responsible and attribute the blame to it. Thus, insight into the reasoning of scientists and research findings in recent years suggest that the current literature does not provide a clear answer to the question of moral responsibility. The difficulty of providing a clear answer to the question was also recognized by Cunneen et al. (2019b).

Moral Status of AI

In addition, the review has shown that the moral status of AI is another focus of research in recent years. While Gordon (2020) and Wareham (2021) argue that the attribution of moral status should not be rejected in principle, the majority does not assign moral status to AI. For example, AI does not meet the criterion of sentience (Gibert and Martin, 2021) and consciousness (Mosakas, 2021; Nath and Sahu, 2020). Moreover, AI is unable to uphold norms by not being able to consciously violate them (Swanepoel, 2021). The fact that AI cannot be a moral agent is also justified by the fact that it lacks emotion-driven behavior (Brożek and Janik, 2019) and ethically relevant properties (Farisco et al., 2020). Furthermore, literature discusses that AI cannot be a moral agent because AI is defined and controlled by humans and should not be a subject of rights and duties (Roff, 2019) and the attempt to develop ethical machines contradicts the concept of ethics itself (Sparrow, 2021). Arguments such as those of Sparrow (2021), which rule out the moral status of AI even for the future, are countered by the possibility that moral status could be ascribed to AI in the future (Livingston and Risse, 2019; Podschwadek, 2017), for example if the cognitive abilities of AI resemble those of other living beings that are considered moral patients (Shevlin, 2021).

Although arguments show that it is currently not technically feasible for artificial agents to replace human actions in morally relevant situations, the development of ethical mechanisms and cognitive structures for artificial agents is not seen as an option, but as an important and feasible condition due to the interaction between humans and AI (Cervantes et al., 2020b; Scheutz, 2017; Mabaso, 2021). Besides, research indicates that strong AI could experience emotions like humans and become an equal part of the society in the future (Kelley and Twyman, 2020). Given the development of cognitive structures and emotions for AI, Anthis and Paez (2021) have also recommended to expand the circle of those to whom humans pay moral attention

to all sentient beings. With respect to the above-mentioned possibility of granting moral status to AI in the future, literature has not only advised caution (Herzog, 2021; Formosa and Ryan, 2021). Accordingly, avoiding such status for AI would be the most ethical decision (Bryson, 2018), as humans would have difficulty controlling and understanding moral actions of a machine (Fabre et al., 2021) and criminal actions. A moral transgression by an AI, would only be effectively sanctioned if humans, rather than the AI itself, are responsible for its actions (Iphofen and Kritikos, 2021). There is also the risk of malware abusing the granted moral status (Mowbray, 2021). Empirical studies have shown that people perceive an AI differently depending on the situation (Hohenstein and Jung, 2020; Hong et al., 2020) and that the perception of the AI's moral status depends, among other things, on its behavior (Banks, 2020) or its visual presentation (Laakasuo et al., 2021). In contrast, other scientific studies have shown that AI is sometimes not perceived as acting morally at all (Gamez et al., 2020; Shank and DeSanti, 2018). While a study by Gupta et al. (2021) shows that racist or gender discriminatory decisions made by an AI are not equally questioned by all humans, a study by Borau et al. (2021) has argued that the perceived moral status of an AI may depend on the specific design of the AI, for example, in the form of a female or male chatbot.

Importance of Law and Politics

A further research focus in the academic discourse around ethics and morality of AI are discussions about the importance of the legal system in preventing or sanctioning the misuse of AI. Appropriate regulations are binding and can be legally and judicially enforced (Carillo, 2020), so AI developers are also primarily guided by legal frameworks (Orr and Davis, 2020). Also, according to Kochupillai et al. (2020), AI programming in particular should be evaluated not only from an ethical perspective, but also from a legal one, thus protecting compliance with legal principles and rules of human values, dignities, and responsibilities. The high responsibility of legislation, as well as the need for new legal norms, has also been highlighted in other studies (Huang, 2019; Pagallo, 2017) and discussed, for example, with regard to the application of AI in schools (Mantovani et al., 2020). In recent years, both judicial application of AI and existing laws have been discussed and criticized (Miller, 2019). While Schönberger (2019) argues in the context of AI in healthcare that current European laws and ethical approaches are sufficient, Miguel et al. (2020) write that the General Data Protection Regulation (GDPR) is incomplete and insufficient for the application of AI in the healthcare sector. According to the literature, other application areas of AI also exhibit inconsistency, a lack of scientific justification (McStay and Rosner, 2021), and a neglect of the specifics of the technologies and contexts (Felzmann et al., 2019). In addition to the gaps in EU legislation, these gaps have also been identified in U.S. legislation (Truby and Brown, 2021) and the influence of large corporations on AI legislation has been criticized (Nemitz, 2018). Legislation and e.g., issues of ownership of AI-created innovations in relation to the patent system are also discussed in literature (Feng and Pan, 2021). It was summarized that the EU, the U.S., and international law do not currently provide a basis for the validity of these moral rights, and it is unclear when AI will be legally considered a personality (Miernicki and Ng, 2021). To this end, it has been voiced that rights can only refer to humans (D'Acquisto, 2020) and robots are not legal persons and do not meet moral criteria (Gordon, 2020). In addition, scholarly articles have also discussed the policies of various countries (Vesnic-Alujevic et al., 2020). Criticisms include the lack of a long-term strategy (Cath et al., 2018) or that the formulation of trustworthy AI is used (Mark, 2020). According to Mark (2020), the formulation does not reflect the current definition of trust, and it is not AI but the organizations and corresponding individuals using AI that require trustworthiness. A scholarly examination of Chinese policy shows that it strongly promotes the development and use of AI because of its potential but disregards the negative consequences. There is no balance between the ethical norms and standards for AI and its implementation (Roberts et al., 2021). Moreover, research discusses the relation to politics not only through the analysis of political actions, but also through the analysis of the integration of AI in political decision-making processes (Sætra, 2020), as well as the manipulation of elections (Landon-Murray et al., 2019; Kane, 2019) or in terms of cyber-attacks (Timmers, 2019). However, possible measures are always associated with ethical challenges (Timmers, 2019).

In the following a short overview is provided which summarizes and visualizes the main information based on the results (see table 3).

Topic	Main Content
Ethical and Moral Implications of AI	- focus of literature on three application areas (implications by businesses, the health sector and the public sector) and only little research on implications of AI for private use - further research considers implications of physical and digital applications as well as implications without referring them to application areas and technologies
Approaches of an Ethical Handling of AI	- recommendations have to be distinguished from principles and literature focuses on principles (in particular in the health sector) as well as on further approaches
Demand for Governance and Regulations	- need for governance models in relation to AI development and its applications - need for governance for AI on a political level and need of binding regulations
AI Programming Requirements	- discussion about programming ethics into AI, programming a friendly AI or rebel agents as well as further technical and social challenges in this context
Human Moral Judgement and Moral Responsibility	- different moral preferences and human judgement leading to different moral decisions - moral responsibility for consequences of actions enabled or performed by AI now and in the future is intensively discussed in literature
Moral Status of AI	- moral status of AI now and in the future is discussed controversially
Importance of Law and Politics	- legal regulations of using or programming AI in different application areas - ethical challenges and different policies around the world

Table 3. Key content of the different research topics.

5 Discussion and conclusion

According to RQ1, this systematic literature review study provides a structured understanding of the current literature of ethics and morality in the AI research. Out of 1,641 studies 224 related articles were included in the evaluation between 2017 and 2021. Regarding RQ2, the systematization of the relevant literature into research priorities also enables the identification of research gaps and thus, contributes to theoretical and practical implications. So, table 4 highlights various additional ideas for future research approaches. All in all, the literature review has shown that existing technologies are applied in a variety of ways in different fields (Hamilton and Davison, 2021; Reiss, 2021) and can thus result in different implications for each derived category:

Ethical and Moral Implications of AI

There is a need for a detailed overview that delineates ethical and moral implications, considering the technologies of AI, the application of these technologies, and their fields of application. It is also noticeable that the research focuses primarily on applications by companies or other organizations. Only few papers have focused on the implications of AI for private use, although the rapid development of technologies in recent years (McAleenan, 2020) and the dynamic nature of ethical concepts and principles (Carillo, 2020) call for a closer examination of the implications for private applications.

Approaches of an Ethical Handling of AI

This literature review showed that there is little focus in the scientific discourse on formulating recommendations for an ethical handling of AI and mostly, they are formulated for specific application contexts. Principles in the scientific discourse about AI can be mainly address applications in the health sector (Stewart et al., 2021). While several main principles are discussed in literature, homogeneous approaches to implement them are required and should be further discussed. Human rights principles have already been discussed as a common aspect for a homogeneous approach to the use of AI across countries.

Demand for Governance and Regulations

Many articles considered in the literature review include calls for governance or the development of governance models by scholars (Lepri et al., 2021). However, the non-binding nature of these is also discussed and regulations are called for in response (Almeida et al., 2021). No articles were identified that present an analysis of the effectiveness of existing regulations, being an important topic for future research.

In addition, efforts by academia to provide alternative approaches and support for the implementation of guidelines (Mark and Stahl, 2021) underscore the criticism of principles and ethical guidelines and highlight the need for empirical testing of these approaches for practical applicability.

AI Programming Requirements

Another identified area of research are the requirements and approaches for programming ethics in AI. Since different application contexts of a machine require different design approaches (Nallur, 2020), the literature is also characterized by arguments of individual authors and different approaches and models. Furthermore, the literature lacks an overview of which applications of AI require particular consideration of ethics and morality in the first place, and the inclusion of user preferences in the discussion of programming (Awad et al., 2018). The analysis showed that the research focus can be divided into approaches to programming (Umbrello et al., 2021) and challenges to programming (Coman and Aha, 2018).

Human Moral Judgement and Moral Responsibility

The subfield of research on human moral judgment considering the decision-making situation of an AI is comparatively under-researched in terms of number of articles and is mostly studied with reference to the trolley dilemma (Frank et al., 2019). However, this thought experiment does not reflect the current technological possibilities of an AI and, in the sense of autonomous driving, represents a currently limited application area of AI (Cunneen et al., 2020). A review of the literature on moral responsibility has revealed disagreement on the distribution of responsibility (Martin, 2019; Rochel and Evéquoz, 2021) and highlighted the question of what exactly is meant by moral responsibility. The responsibility gap poses a major risk for future harm (Burton et al., 2020) and, accordingly, should be discussed further in future.

Moral Status of AI

Disagreement within the literature was also identified in the discourse on the moral status of AI. This literature is heavily influenced by arguments from individual authors and shows disagreement on both whether cognitive structures should be developed for AI (Shevlin, 2021) and what arguments should be used to ascribe moral status to AI (Gordon, 2020).

Importance of Law and Politics

Furthermore, it has become apparent that in this discourse, the consideration of strong AI may be of particular interest in the future (Kelley and Twyman, 2020). In the scholarly discourse on the roles of law and policy, both legislation and policy decisions have been criticized. Strikingly, mostly EU or U.S. law has been mentioned and the focus of the research on European policy has been placed (Truby and Brown, 2021; Mark, 2020). This underrepresentation of certain regions (Jobin et al., 2019) should thus be balanced in further research. The study Huang (2019) shows that enacted laws can have an impact on human moral judgments of AI. These findings should be reviewed as they could have important implications for stakeholders in the development of AI for international markets.

Topic	Future Research Questions
Ethical and Moral Implications of AI	Is there a difference of ethical and moral implications within AI based programs and products between private use and corporate use? Which ethical and moral factors influence the use in each case?
	What are the different ethical and moral implications that need to be considered during the private use of AI-based physical and digital products (e.g., smart watches, smart speakers, smartphones, etc.)?
	To what extent do ethical concerns and ethical justifications influence the use of AI by users on the one hand and companies on the other?
Approaches of an Ethical Handling of AI	To what extent is there a need for complementary sector-specific or application-specific approaches for an ethical approach to AI? Do the possibilities for implementing existing approaches differ with regard to organizational sizes and forms?
	To what extent do AI-developing companies or AI-using organizations currently take ethics guidelines and principles into account? What criteria do they use to select them?
	What approaches are needed to help private users deal with the ethical and moral problems of using AI?
Demand for Governance and Regulations	How should the developed governance models of ethical and moral AI be evaluated for their practicality and effectiveness?
	In what areas is AI already regulated due to ethical and moral concerns? Are these regulations fulfilling their intended purposes? How do they affect the development and application of AI in the regulated areas?

AI Programming Requirements	To what extent are some approaches to programming ethics in AI better suited to particular application contexts than others?
	Which of the existing approaches to programming ethics in AI are preferred by users? Are there differences here in terms of application contexts?
	How should the approaches to programming ethics in AI that have been developed in recent years be evaluated for their practical feasibility and effectiveness? To what extent are such approaches already being used?
Human Moral Judgement and Moral Responsibility	How do different stakeholders define the term "moral responsibility" in the discourse around AI?
	How does the perception of AI as a moral bearer of responsibility influence the demand and application of corresponding products and services?
	How do marketers adapt their communications strategies in the context of ethical and moral AI?
Moral Status of AI	What is the practical significance of recognizing that some users ascribe moral status to AI, while other users do not ascribe morality to AI? How will the moral status of AI change in the future?
Importance of Law and Politics	What influence do regional policy debates and regional legal frameworks on ethics and morality in the context of AI show on the development and use of AI on different regions of the world?
	How do internationally operating companies deal with the legislation of different regions on the ethical development and application of AI?

Table 4. Future Research Questions.

Summary

In summary, the literature review illustrates that discussions from academia about ethics and morality of humans throughout AI have increased significantly in recent years. The entire discourse is characterized by different research approaches and arguments, as well as by mutual criticism within the scientific community and of external stakeholders. This can be attributed to the complexity of the topic. Thus, an important finding of this work is that research in recent years is not numerically tangible, nor can they be easily represented in tables. Nevertheless, the systematic literature analysis allows us to show how many research articles can be assign to which research area. Thus, articles dealing with ethical and moral implications and the approaches of an ethical handling of AI could be identified most often throughout the review. In particular, implications in the application areas of businesses, in the health sector and the public sector indicate a lot of different articles, while significantly fewer papers deal with implications of, e.g., physical or digital products. Additionally, the majority of the articles cover principles and further approaches to integrate ethical aspects in AI but do not include recommendations. In the context of governance and regulations, most of the literature lay the focus on regulations. While literature on AI programming requirements explores many heterogeneous approaches, many papers could be identified that deal with the moral status of AI and especially try to predict how it will be in the future. However, it should be noted, that many articles cover several subject areas and not only one. Therefore, it was considered which topic or which topics were discussed centrally in the respective paper.

Since the lack of verification of existing results results in a high need for future research, our systematization of the literature not only provides an overview and added value for academics, but also for practitioners. For example, it shows overall that close collaboration between science and policy is necessary to enable governance and regulation for an ethical approach to AI. Furthermore, it has become clear that organizations need to understand their responsibilities in this context, of which our work provides a holistic overview.

Although we provided several contributions to the current state of literature and future research questions there are some limitations to be considered. As this research focuses more on a holistic approach of ethics and morality in the context of AI since it investigates the current state of literature through different databases and journals in general, it would be interesting to take a deeper look in solely branches which are mentioned here such as the marketing sector or the healthcare sector within the information system research. Moreover, the review concentrates on articles with a two-year impact factor over 0.5 available in selected scholarly databases. Other publican forms like e.g., books were excluded. Furthermore, some papers in the IS field consider ethics and morality in the context of AI as a sub-issue (Dwivedi, Yogesh K. et al., 2021; Wu, 2020; Duan et al., 2019). Since they were not identified based on the Boolean factors of our literature analysis, future research could provide more in-depth overviews of the different research areas with extended search algorithms.

References

Abràmoff, M. D., D. Tobey and D. S. Char (2020). "Lessons Learned About Autonomous AI: Finding a Safe, Efficacious, and Ethical Path Through the Development Process" *American Journal of Ophthalmology* 214, 134–142.

Abrams, M., J. Abrams, P. Cullen and L. Goldstein (2019). "Artificial Intelligence, Ethics, and Enhanced Data Stewardship" *IEEE Security & Privacy* 17 (2), 17–30.

Adamson, G., J. C. Havens and R. Chatila (2019). "Designing a Value-Driven Future for Ethical Autonomous and Intelligent Systems" *Proceedings of the IEEE* 107 (3), 518–525.

Aizenberg, E. and J. van den Hoven (2020). "Designing for human rights in AI" *Big Data & Society* 7 (2), 205395172094956.

Almeida, P. G. R. de, C. D. dos Santos and J. S. Farias (2021). "Artificial Intelligence Regulation: a framework for governance" *Ethics and Information Technology*, 1–21.

Analytics, C. (2018). *InCites indicators handbook.* Philadelphia, PA, USA: Clarivate Analytics.

Anshari, M., M. N. Almunawar, M. Masri and M. Hrdy (2021). "Financial Technology with AI-Enabled and Ethical Challenges" *Society* 58 (3), 189–195.

Antao, D. (2019). "Technological Advancement, Artificial Intelligence and the Relevance of Classroom Teachers" *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)* 6 (2), 353–361.

Anthis, J. R. and E. Paez (2021). "Moral circle expansion: A promising strategy to impact the far future" *Futures* 130, 102756.

Antle, A. N. and A. Kitson (2021). "1,2,3,4 tell me how to grow more: A position paper on children, design ethics and biowearables" *International Journal of Child-Computer Interaction* 30, 100328.

Arnold, T. and M. Scheutz (2018). "The "big red button" is too late: an alternative model for the ethical evaluation of AI systems" *Ethics and Information Technology* 20 (1), 59–69.

Arvan, M. (2018). "Mental time-travel, semantic flexibility, and A.I. ethics" *AI & SOCIETY*, 1–20.

Asaro, P. M. (2019). "AI Ethics in Predictive Policing: From Models of Threat to an Ethics of Care" *IEEE Technology and Society Magazine* 38 (2), 40–53.

Autili, M., D. Di Ruscio, P. Inverardi, P. Pelliccione and M. Tivoli (2019). "A Software Exoskeleton to Protect and Support Citizen's Ethics and Privacy in the Digital World" *IEEE Access* 7, 62011–62021.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon and I. Rahwan (2018). "The Moral Machine experiment" *Nature* 563 (7729), 59–64.

Bærøe, K., A. Miyata-Sturm and E. Henden (2020). "How to achieve trustworthy artificial intelligence for health" *Bulletin of the World Health Organization* 98 (4), 257–262.

Banks, J. (2020). "Good Robots, Bad Robots: Morally Valenced Behavior Effects on Perceived Mind, Morality, and Trust" *International Journal of Social Robotics*, 1–18.

Bartneck, C., C. Lütge, A. Wagner and S. Welsh (2019). *Ethik in KI und Robotik.* München: Carl Hanser Verlag GmbH Co KG.

Bauer, W. A. (2020). "Virtuous vs. utilitarian artificial moral agents" *AI & SOCIETY* 35 (1), 263–271.

Baum, S. D. (2020). "Social choice ethics in artificial intelligence" *AI & SOCIETY* 35 (1), 165–176.

Belk, R. (2021). "Ethical issues in service robotics and artificial intelligence" *The Service Industries Journal* 41 (13-14), 860–876.

Benzmüller, C., X. Parent and L. van der Torre (2020). "Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support" *Artificial Intelligence* 287, 103348.

Bogosian, K. (2017). "Implementation of Moral Uncertainty in Intelligent Machines" *Minds and Machines* 27 (4), 591–608.

Borau, S., T. Otterbring, S. Laporte and S. Fosso Wamba (2021). "The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI" *Psychology & Marketing* 38 (7), 1052–1068.

Borenstein, J., J. R. Herkert and K. W. Miller (2019). "Self-Driving Cars and Engineering Ethics: The Need for a System Level Analysis" *Science and Engineering Ethics* 25 (2), 383–398.

Brady, A. P. and E. Neri (2020). "Artificial Intelligence in Radiology-Ethical Considerations" *Diagnostics (Basel, Switzerland)* 10 (4), 231.
Brendel, A. B., M. Mirbabaie, T.-B. Lembcke and L. Hofeditz (2021). "Ethical Management of Artificial Intelligence" *Sustainability* 13 (4), 1974.
Briner, R. B. and D. Denyer (2012). *Systematic Review and Evidence Synthesis as a Practice and Scholarship Tool:* Oxford University Press.
Brożek, B. and B. Janik (2019). "Can artificial intelligences be moral agents?" *New Ideas in Psychology* 54, 101–106.
Bryson, J. J. (2018). "Patiency is not a virtue: the design of intelligent systems and systems of ethics" *Ethics and Information Technology* 20 (1), 15–26.
Burton, E., J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei and T. Walsh (2017). "Ethical Considerations in Artificial Intelligence Courses" *AI Magazine* 38 (2), 22–34.
Burton, S., I. Habli, T. Lawton, J. McDermid, P. Morgan and Z. Porter (2020). "Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective" *Artificial Intelligence* 279, 103201.
Buruk, B., P. E. Ekmekci and B. Arda (2020). "A critical perspective on guidelines for responsible and trustworthy artificial intelligence" *Medicine, Health Care and Philosophy* 23 (3), 387–399.
Calvo, P. (2020). "The ethics of Smart City (EoSC): moral implications of hyperconnectivity, algorithmization and the datafication of urban digital society" *Ethics and Information Technology* 22 (2), 141–149.
Campbell, C., S. Sands, C. Ferraro, Tsao, Hsiu-Yuan (Jody) and A. Mavrommatis (2020). "From data to action: How marketers can leverage AI" *Business Horizons* 63 (2), 227–243.
Carillo, R. M. (2020). "Artificial intelligence: From ethics to law" *Telecommunications Policy* 44 (6), 101937.
Cath, C., S. Wachter, B. Mittelstadt, M. Taddeo and L. Floridi (2018). "Artificial Intelligence and the 'Good Society': the US, EU, and UK approach" *Science and Engineering Ethics* 24 (2), 505–528.
Cervantes, J.-A., S. López, L.-F. Rodríguez, S. Cervantes, F. Cervantes and F. Ramos (2020a). "Artificial Moral Agents: A Survey of the Current Status" *Science and Engineering Ethics* 26 (2), 501–532.
Cervantes, S., S. López and J.-A. Cervantes (2020b). "Toward ethical cognitive architectures for the development of artificial moral agents" *Cognitive systems research* 64, 117–125.
Chounta, I.-A., E. Bardone, A. Raudsep and M. Pedaste (2021). "Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education" *International Journal of Artificial Intelligence in Education*, 1–31.
Cichocki, A. and A. P. Kuleshov (2021). "Future Trends for Human-AI Collaboration: A Comprehensive Taxonomy of AI/AGI Using Multiple Intelligences and Learning Styles" *Computational Intelligence and Neuroscience* 2021, 1–21.
Clarke, R. (2019). "Principles and business processes for responsible AI" *Computer Law & Security Review* 35 (4), 410–422.
Coman, A. and D. W. Aha (2018). "AI Rebel Agents" *AI Magazine* 39 (3), 16–26.
Cunneen, M., M. Mullins and F. Murphy (2019a). "Autonomous Vehicles and Embedded Artificial Intelligence: The Challenges of Framing Machine Driving Decisions" *Applied Artificial Intelligence* 33 (8), 706–731.
Cunneen, M., M. Mullins, F. Murphy and S. Gaines (2019b). "Artificial driving intelligence and moral agency: Examining the decision ontology of unavoidable road traffic accidents through the prism of the trolley dilemma" *Applied Artificial Intelligence* 33 (3), 267–293.
Cunneen, M., M. Mullins, F. Murphy, D. Shannon, I. Furxhi and C. Ryan (2020). "Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics" *Cybernetics and Systems* 51 (1), 59–80.
Cvik, Š. (2021). "Categorization and challenges of utilitarianisms in the context of artificial intelligence" *AI & SOCIETY*, 1–7.

D'Acquisto, G. (2020). "On conflicts between ethical and logical principles in artificial intelligence" *AI & SOCIETY* 35 (4), 895–900.
Danaher, J. (2019). "The rise of the robots and the crisis of moral patiency" *AI & SOCIETY* 34 (1), 129–136.
Du, S. (2021). "Reimagining the Future of Technology: "The Social Dilemma" Review" *Journal of Business Ethics*, 1–3.
Du, S. and C. Xie (2021). "Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities" *Journal of Business Research* 129, 961–974.
Duan, Y., J. S. Edwards and Y. K. Dwivedi (2019). "Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agend" *International Journal of Information Management* 48, 63–71.
Dubljević, V. (2020). "Toward Implementing the ADC Model of Moral Judgment in Autonomous Vehicles" *Science and Engineering Ethics* 26 (5), 2461–2472.
Dwivedi, Yogesh K. et al. (2021). "Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy." *International Journal of Information Management* 57, 101994.
EU Commission (2020). "White Paper on Artificial Intelligence - A European approach to excellence and trust" *Brussels*, 1–26.
Fabre, R. G., J. Camacho Ibáñez and P. Tejedor Escobar (2021). "Moral control and ownership in AI systems" *AI & SOCIETY* 36 (1), 289–303.
Farisco, M., K. Evers and A. Salles (2020). "Towards Establishing Criteria for the Ethical Analysis of Artificial Intelligence" *Science and Engineering Ethics* 26 (5), 2413–2425.
Felzmann, H., E. F. Villaronga, C. Lutz and A. Tamò-Larrieux (2019). "Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns" *Big Data & Society* 6 (1), 205395171986054.
Fenech, M. E. and O. Buston (2020). "AI in Cardiac Imaging: A UK-Based Perspective on Addressing the Ethical, Social, and Political Challenges" *Frontiers in cardiovascular medicine* 7, 54.
Feng, X.-Q. and B.-H. Pan (2021). "The evolution of patent system: invention created by artificial intelligence" *Procedia Computer Science* 183, 245–253.
Fernandes, P. M., F. C. Santos and M. Lopes (2020). "Norms for beneficial A.I.: A computational analysis of the societal value alignment problem" *AI Communications* 33 (3-6), 155–171.
Fiske, A., P. Henningsen and A. Buyx (2019). "Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy" *Journal of Medical Internet Research* 21 (5), 1–12.
Floridi, L. (2018). "Soft ethics, the governance of the digital and the General Data Protection Regulation" *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 376 (2133).
Floridi, L., J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke and E. Vayena (2018). "AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations" *Minds and Machines* 28 (4), 689–707.
Floridi, L., J. Cowls, T. C. King and M. Taddeo (2021). "How to Design AI for Social Good: Seven Essential Factors". In *Ethics, Governance, and Policies in Artificial Intelligence*, pp. 125–151: Springer, Cham. URL: https://link.springer.com/chapter/10.1007/978-3-030-81907-1_9.
Formosa, P. and M. Ryan (2021). "Making moral machines: why we need artificial moral agents" *AI & SOCIETY* 36 (3), 839–851.
Frank, D.-A., P. Chrysochou, P. Mitkidis and D. Ariely (2019). "Human decision-making biases in the moral dilemmas of autonomous vehicles" *Scientific Reports* 9 (1), 13080.
Franzke, A. S., I. Muis and M. T. Schäfer (2021). "Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands" *Ethics and Information Technology*, 1–17.

Fritz, O., C. Weber, A. König and J. Wolf (2019). "Ethische Aspekte der Künstlichen Intelligenz" *KCT Schriftenreihe der FOM* (1).
Fröding, B. and M. Peterson (2020). "Friendly AI" *Ethics and Information Technology*, 1–8.
Fukuda-Parr, S. and E. Gibbons (2021). "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines" *Global Policy* 12 (S6), 32–44.
Gabriel, I. (2020). "Artificial Intelligence, Values, and Alignment" *Minds and Machines* 30 (3), 411–437.
Gamez, P., D. B. Shank, C. Arnold and M. North (2020). "Artificial virtue: the machine question and perceptions of moral character in artificial moral agents" *AI & SOCIETY* 35 (4), 795–809.
Geis, J. R., A. P. Brady, C. C. Wu, J. Spencer, E. Ranschaert, J. L. Jaremko, S. G. Langer, A. B. Kitts, J. Birch, W. F. Shields, R. van Hoven Genderen, E. Kotter, J. W. Gichoya, T. S. Cook, M. B. Morgan, an Tang, N. M. Safdar and M. Kohli (2019). "Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement" *Canadian Association of Radiologists journal = Journal l'Association canadienne des radiologistes* 70 (4), 329–334.
Ghotbi, N., M. T. Ho and P. Mantello (2021). "Attitude of college students towards ethical issues of artificial intelligence in an international university in Japan" *AI & SOCIETY*, 1–8.
Gibbons, E. D. (2021). "Toward a More Equal World: The Human Rights Approach to Extending the Benefits of Artificial Intelligence" *IEEE Technology and Society Magazine* 40 (1), 25–30.
Gibert, M. and D. Martin (2021). "In search of the moral status of AI: why sentience is a strong argument" *AI & SOCIETY*, 1–12.
Gill, T. (2020). "Blame It on the Self-Driving Car: How Autonomous Vehicles Can Alter Consumer Morality" *SSRN Electronic Journal*.
Gordon, J.-S. (2020). "Building Moral Robots: Ethical Pitfalls and Challenges" *Science and Engineering Ethics* 26 (1), 141–157.
Gordon, J.-S. (2021). "Artificial moral and legal personhood" *AI & SOCIETY* 36 (2), 457–471.
Gruson, D., T. Helleputte, P. Rousseau and D. Gruson (2019). "Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation" *Clinical Biochemistry* 69, 1–7.
Gupta, M., C. M. Parra and D. Dennehy (2021). "Questioning Racial and Gender Bias in AI-based Recommendations: Do Espoused National Cultural Values Matter?" *Information Systems Frontiers*, 1–17.
Haas, J. (2020). "Moral Gridworlds: A Theoretical Proposal for Modeling Artificial Moral Cognition" *Minds and Machines* 30 (2), 219–246.
Haenlein, M. and A. Kaplan (2019). "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence" *California Management Review* 61 (4), 5–14.
Hagendorff, T. (2020). "The ethics of AI ethics: An evaluation of guidelines" *Minds and Machines* 30 (1), 99–120.
Hamilton, R. H. and H. K. Davison (2021). "Legal and Ethical Challenges for HR in Machine Learning" *Employee Responsibilities and Rights Journal*, 1–21.
Héder, M. (2020). "The epistemic opacity of autonomous systems and the ethical consequences" *AI & SOCIETY*, 1–9.
Hermann, E. (2021a). "Artificial intelligence and mass personalization of communication content—An ethical and literacy perspective" *New Media & Society*, 146144482110227.
Hermann, E. (2021b). "Leveraging Artificial Intelligence in Marketing for Social Good-An Ethical Perspective" *Journal of Business Ethics*, 1–19.
Herzog, C. (2021). "Three Risks That Caution Against a Premature Implementation of Artificial Moral Agents for Practical and Economical Use" *Science and Engineering Ethics* 27 (1), 3.
Ho, C. W. L., J. Ali and K. Caals (2020). "Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance" *Bulletin of the World Health Organization* 98 (4), 263–269.
Hohenstein, J. and M. Jung (2020). "AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust" *Computers in Human Behavior* 106, 106190.

Hong, J.-W., S. Choi and D. Williams (2020). "Sexist AI: An Experiment Integrating CASA and ELM" *International Journal of Human–Computer Interaction* 36 (20), 1928–1941.
Huang, B. I. (2019). "Law´s Halo and the Moral Machine" *Columbia Law Review* 119, 1811–1828.
Huang, M.-H. and R. T. Rust (2021). "A strategic framework for artificial intelligence in marketing" *Journal of the Academy of Marketing Science* 49 (1), 30–50.
Hynek, N. and A. Solovyeva (2021). "Operations of power in autonomous weapon systems: ethical conditions and socio-political prospects" *AI & SOCIETY* 36 (1), 79–99.
Iphofen, R. and M. Kritikos (2021). "Regulating artificial intelligence and robotics: ethics by design in a digital society" *Contemporary Social Science* 16 (2), 170–184.
Janssen, M., P. Brous, E. Estevez, L. S. Barbosa and T. Janowski (2020). "Data governance: Organizing data for trustworthy Artificial Intelligence" *Government Information Quarterly* 37 (3), 101493.
Jobin, A., M. Ienca and E. Vayena (2019). "The global landscape of AI ethics guidelines" *Nature Machine Intelligence* 1 (9), 389–399.
Joerin, A., M. Rauws, R. Fulmer and V. Black (2020). "Ethical Artificial Intelligence for Digital Health Organizations" *Cureus* 12 (3), e7202.
Jotterand, F. and C. Bosco (2020). "Keeping the "Human in the Loop" in the Age of Artificial Intelligence : Accompanying Commentary for "Correcting the Brain?" by Rainey and Erden" *Science and Engineering Ethics* 26 (5), 2455–2460.
Kane, T. B. (2019). "Artificial Intelligence in Politics: Establishing Ethics" *IEEE Technology and Society Magazine* 38 (1), 72–80.
Kaplan, A. and M. Haenlein (2019). "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence" *Business Horizons* 62 (1), 15–25.
Kelley, D. and K. Atreides (2020). "AGI Protocol for the Ethical Treatment of Artificial General Intelligence Systems" *Procedia Computer Science* 169, 501–506.
Kelley, D. and M. Twyman (2020). "Biasing in an Independent Core Observer Model Artificial General Intelligence Cognitive Architecture" *Procedia Computer Science* 169, 535–541.
Kerr, A., M. Barry and J. D. Kelleher (2020). "Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance" *Big Data & Society* 7 (1), 205395172091593.
Khargonekar, P. P. and M. Sampath (2020). "A Framework for Ethics in Cyber-Physical-Human Systems" *IFAC-PapersOnLine* 53 (2), 17008–17015.
Khrais, L. T. (2020). "Role of Artificial Intelligence in Shaping Consumer Demand in E-Commerce" *Future Internet* 12 (12), 226.
King, O. C. (2020). "Presumptuous aim attribution, conformity, and the ethics of artificial social cognition" *Ethics and Information Technology* 22 (1), 25–37.
Köbis, N., J.-F. Bonnefon and I. Rahwan (2021). "Bad machines corrupt good morals" *Nature Human Behaviour* 5 (6), 679–685.
Kochupillai, M., C. Lütge and F. Poszler (2020). "Programming Away Human Rights and Responsibilities? "The Moral Machine Experiment" and the Need for a More "Humane" AV Future" *NanoEthics* 14 (3), 285–299.
Kretzschmar, K., H. Tyroll, G. Pavarini, A. Manzini and I. Singh (2019). "Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support" *Biomedical Informatics Insights* 11, 1178222619829083.
Kriebitz, A. and C. Lütge (2020). "Artificial Intelligence and Human Rights: A Business Ethical Assessment" *Business and Human Rights Journal* 5 (1), 84–104.
Laakasuo, M., J. Palomäki and N. Köbis (2021). "Moral Uncanny Valley: A Robot's Appearance Moderates How its Decisions are Judged" *International Journal of Social Robotics* 13 (7), 1679–1688.
Landon-Murray, M., E. Mujkic and B. Nussbaum (2019). "Disinformation in Contemporary U.S. Foreign Policy: Impacts and Ethics in an Era of Fake News, Social Media, and Artificial Intelligence" *Public Integrity* 21 (5), 512–522.

Lara, F. and J. Deckers (2020). "Artificial Intelligence as a Socratic Assistant for Moral Enhancement" *Neuroethics* 13 (3), 275–287.
Larson, D. B., D. C. Magnus, M. P. Lungren, N. H. Shah and C. P. Langlotz (2020). "Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework" *Radiology* 295 (3), 675–682.
Larsson, S. (2020). "On the Governance of Artificial Intelligence through Ethics Guidelines" *Asian Journal of Law and Society* 7 (3), 437–451.
Lauwaert, L. (2021). "Artificial intelligence and responsibility" *AI & SOCIETY* 36 (3), 1001–1009.
Lepri, B., N. Oliver and A. Pentland (2021). "Ethical machines: The human-centric use of artificial intelligence" *iScience* 24 (3), 102249.
Li, O. (2021). "Problems with "Friendly AI"" *Ethics and Information Technology*, 1–8.
Libai, B., Y. Bart, S. Gensler, C. F. Hofacker, A. Kaplan, K. Kötterheinrich and E. B. Kroll (2020). "Brave New World? On AI and the Management of Customer Relationships" *Journal of Interactive Marketing* 51, 44–56.
Liu, H.-W., C.-F. Lin and Y.-J. Chen (2019). "Beyond State v Loomis: artificial intelligence, government algorithmization and accountability" *International Journal of Law and Information Technology* 27 (2), 122–141.
Liu, H.-Y. and M. M. Maas (2021). "'Solving for X?' Towards a problem-finding framework to ground long-term governance strategies for artificial intelligence" *Futures* 126, 102672.
Livingston, S. and M. Risse (2019). "The Future Impact of Artificial Intelligence on Humans and Human Rights" *Ethics & International Affairs* 33 (02), 141–158.
Loureiro, S. M. C., J. Guerreiro and I. Tussyadiah (2021). "Artificial intelligence in business: State of the art and future research agenda" *Journal of Business Research* 129, 911–926.
Lu, H., Y. Li, M. Chen, H. Kim and S. Serikawa (2018). "Brain Intelligence: Go beyond Artificial Intelligence" *Mobile Networks and Applications* 23 (2), 368–375.
Luk, J. W., L. D. Pruitt, D. J. Smolenski, J. Tucker, D. E. Workman and B. E. Belsher (2021). "From everyday life predictions to suicide prevention: Clinical and ethical considerations in suicide predictive analytic tools" *Journal of Clinical Psychology*.
Mabaso, B. A. (2021). "Computationally rational agents can be moral agents" *Ethics and Information Technology* 23 (2), 137–145.
Macnish, K. and F. I. Ana (2019). "Case Study - Customer Relation Management, Smart Information Systems and Ethics" *The ORBIT Journal* 2 (2), 1–24.
Magrani, E. (2019). "New perspectives on ethics and the laws of artificial intelligence" *Internet Policy Review* 8 (3).
Mantelero, A. (2018). "AI and Big Data: A blueprint for a human rights, social and ethical impact assessment" *Computer Law & Security Review* 34 (4), 754–772.
Mantelero, A. and S. Esposito (2021). "An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems" *Computer Law & Security Review* 41, 1–35.
Mantovani, E., I. Böröcz and P. de Hert (2020). "Conducting research with school children and data in line with "ethical principles" lawyers at work in the ethics management of the H2020 mathisis project" *Computer Law & Security Review* 38, 105451.
Mark, R. (2019a). "Ethics of Public Use of AI and Big Data" *The ORBIT Journal* 2 (2), 1–33.
Mark, R. (2019b). "Ethics of Using AI and Big Data in Agriculture: The Case of a Large Agriculture Multinational" *The ORBIT Journal* 2 (2), 1–27.
Mark, R. (2020). "In AI We Trust: Ethics, Artificial Intelligence, and Reliability" *Science and Engineering Ethics* 26 (5), 2749–2767.
Mark, R. and G. Anya (2019). "Ethics of Using Smart City AI and Big Data: The Case of Four Large European Cities" *The ORBIT Journal* 2 (2), 1–36.

Mark, R. and B. C. Stahl (2021). "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications" *Journal of Information, Communication and Ethics in Society* 19 (1), 61–86.
Martin, D. (2017). "Who Should Decide How Machines Make Morally Laden Decisions?" *Science and Engineering Ethics* 23 (4), 951–967.
Martin, K. (2019). "Ethical Implications and Accountability of Algorithms" *Journal of Business Ethics* 160 (4), 835–850.
Martinho, A., M. Kroesen and C. Chorus (2021). "Computer Says I Don't Know: An Empirical Approach to Capture Moral Uncertainty in Artificial Intelligence" *Minds and Machines* 31, 215–237.
McAleenan, P. (2020). "Moral responsibility and action in the use of artificial intelligence in construction" *Proceedings of the Institution of Civil Engineers - Management, Procurement and Law* 173 (4), 166–174.
McDougall, R. J. (2019). "Computer knows best? The need for value-flexibility in medical AI" *Journal of medical ethics* 45 (3), 156–160.
McKernan, L. C., E. W. Clayton and C. G. Walsh (2018). "Protecting Life While Preserving Liberty: Ethical Recommendations for Suicide Prevention With Artificial Intelligence" *Frontiers in psychiatry* 9, 650.
McManus, R. M. and A. M. Rutchick (2019). "Autonomous Vehicles and the Attribution of Moral Responsibility" *Social Psychological and Personality Science* 10 (3), 345–352.
McStay, A. (2020). "Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy" *Big Data & Society* 7 (1), 205395172090438.
McStay, A. and G. Rosner (2021). "Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies" *Big Data & Society* 8 (1), 205395172199487.
Miernicki, M. and I. Ng (2021). "Artificial intelligence and moral rights" *AI & SOCIETY* 36 (1), 319–329.
Miguel, I. de, B. Sanz and G. Lazcoz (2020). "Machine learning in the EU health care context: exploring the ethical, legal and social issues" *Information, Communication & Society* 23 (8), 1139–1153.
Milano, S., M. Taddeo and L. Floridi (2020). "Recommender systems and their ethical challenges" *AI & SOCIETY* 35 (4), 957–967.
Milano, S., M. Taddeo and L. Floridi (2021). "Ethical aspects of multi-stakeholder recommendation systems" *The Information Society* 37 (1), 35–45.
Miles, M. B. and A. M. Hubermann (1994). *Qualitative data analysis: An expanded sourcebook:* sage.
Miller, S. (2019). "Machine Learning, Ethics and Law" *Australasian Journal of Information Systems* 23.
Milossi, M., E. Alexandropoulou-Egyptiadou and K. E. Psannis (2021). "AI Ethics: Algorithmic Determinism or Self-Determination? The GPDR Approach" *IEEE Access* 9, 58455–58466.
Misselhorn, C. (2018). "Artificial Morality. Concepts, Issues and Challenges" *Society* 55 (2), 161–169.
Mittelstadt, B. (2019). "Principles alone cannot guarantee ethical AI" *Nature Machine Intelligence* 1 (11), 501–507.
Mökander, J. and M. Axente (2021). "Ethics-based auditing of automated decision-making systems: intervention points and policy implications" *AI & SOCIETY*, 1–19.
Mökander, J. and L. Floridi (2021). "Ethics-Based Auditing to Develop Trustworthy AI" *Minds and Machines* 31 (2), 323–327.
Montemayor, C., J. Halpern and A. Fairweather (2021). "In principle obstacles for empathic AI: why we can't replace human empathy in healthcare" *AI & SOCIETY*, 1–7.
Mörch, C.-M., A. Gupta and B. L. Mishara (2020). "Canada protocol: An ethical checklist for the use of artificial Intelligence in suicide prevention and mental health" *Artificial intelligence in medicine* 108, 101934.
Morley, J., A. Elhalal, F. Garcia, L. Kinsey, J. Mökander and L. Floridi (2021). "Ethics as a Service: A Pragmatic Operationalisation of AI Ethics" *Minds and Machines* 31 (2), 239–256.
Morley, J., C. C. V. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo and L. Floridi (2020). "The ethics of AI in health care: A mapping review" *Social science & medicine (1982)* 260, 113172.

Mosakas, K. (2021). "On the moral status of social robots: considering the consciousness criterion" *AI & SOCIETY* 36 (2), 429–443.
Mowbray, M. (2021). "Moral Status for Malware! The Difficulty of Defining Advanced Artificial Intelligence" *Cambridge Quarterly of Healthcare Ethics* 30 (3), 517–528.
Murdock, G. (2018). "Media Materialties: For A Moral Economy of Machines" *Journal of Communication* 68 (2), 359–368.
Nallur, V. (2020). "Landscape of Machine Implemented Ethics" *Science and Engineering Ethics* 26 (5), 2381–2399.
Nath, R. and V. Sahu (2020). "The problem of machine ethics in artificial intelligence" *AI & SOCIETY* 35 (1), 103–111.
Nemitz, P. (2018). "Constitutional democracy and technology in the age of artificial intelligence" *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 376 (2133).
Nigam, A., R. Pasricha, T. Singh and P. Churi (2021). "A Systematic Review on AI-based Proctoring Systems: Past, Present and Future" *Education and Information Technologies* 26 (5), 1–25.
Noothigattu, R., D. Bouneffouf, N. Mattei, R. Chandra, P. Madan, K. R. Varshney, M. Campbell, M. Singh and F. Rossi (2019). "Teaching AI agents ethical values using reinforcement learning and policy orchestration" *IBM Journal of Research and Development* 63 (4/5), 2:1-2:9.
Novak, T. P. (2020). "A Generalized Framework for Moral Dilemmas Involving Autonomous Vehicles: A Commentary on Gill" *Journal of Consumer Research* 47 (2), 292–300.
ÓhÉigeartaigh, S. S., J. Whittlestone, Y. Liu, Y. Zeng and Z. Liu (2020). "Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance" *Philosophy & Technology* 33 (4), 571–593.
Orr, W. and J. L. Davis (2020). "Attributions of ethical responsibility by Artificial Intelligence practitioners" *Information, Communication & Society* 23 (5), 719–735.
Osório, A. and A. Pinto (2019). "Information, uncertainty and the manipulability of artificial intelligence autonomous vehicles systems" *International Journal of Human-Computer Studies* 130, 40–46.
Pagallo, U. (2017). "When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law" *Minds and Machines* 27 (4), 625–638.
Park, S. H., Y.-H. Kim, J. Y. Lee, S. Yoo and C. J. Kim (2019). "Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review" *Science Editing* 6 (2), 91–98.
Peters, D., K. Vold, D. Robinson and R. A. Calvo (2020). "Responsible AI—Two Frameworks for Ethical Design Practice" *IEEE Transactions on Technology and Society* 1 (1), 34–47.
Pillai, V. S. and K. J. M. Matus (2021). "Towards a responsible integration of artificial intelligence technology in the construction sector" *Science and Public Policy* 47 (5), 689–704.
Pinka, R. (2021). "Synthetic Deliberation: Can Emulated Imagination Enhance Machine Ethics?" *Minds and Machines* 31 (1), 121–136.
Podschwadek, F. (2017). "Do androids dream of normative endorsement? On the fallibility of artificial moral agents" *Artificial Intelligence and Law* 25 (3), 325–339.
Popa, E. O., M. van Hilten, E. Oosterkamp and M.-J. Bogaardt (2021). "The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks" *Life Sciences, Society and Policy* 17 (1), 6.
Puntoni, S., R. Walker Reczek, M. Giesler and S. Botti (2021). "Consumers and Artificial Intelligence: An Experimental Perspective" *Journal of Marketing* 85 (1), 131–151.
Raab, C. D. (2020). "Information privacy, impact assessment, and the place of ethics" *Computer Law & Security Review* 37, 105404.
Radu, R. (2021). "Steering the governance of artificial intelligence: national strategies in perspective" *Policy and Society* 40 (2), 178–193.
Reddy, S., S. Allan, S. Coghlan and P. Cooper (2020). "A governance model for the application of AI in health care" *Journal of the American Medical Informatics Association : JAMIA* 27 (3), 491–497.
Reiss, M. J. (2021). "The use of AI in education: Practicalities and ethical considerations" *London Review of Education* 19 (1).

Rességuier, A. and R. Rodrigues (2020). "AI ethics should not remain toothless! A call to bring back the teeth of ethics" *Big Data & Society* 7 (2), 205395172094254.
Rhim, J., J.-H. Lee, M. Chen and A. Lim (2021). "A Deeper Look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-Making Framework to Explain Moral Pluralism" *Frontiers in Robotics and AI* 8, 632394.
Robbins, S. (2019). "A Misdirected Principle with a Catch: Explicability for AI" *Minds and Machines* 29 (4), 495–514.
Robbins, S. (2020). "AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines" *AI & SOCIETY* 35 (2), 391–400.
Roberts, H., J. Cowls, J. Morley, M. Taddeo, V. Wang and L. Floridi (2021). "The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation" *AI & SOCIETY* 36 (1), 59–77.
Rochel, J. and F. Evéquoz (2021). "Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics" *AI & SOCIETY* 36 (2), 609–622.
Roff, H. M. (2019). "Artificial Intelligence: Power to the People" *Ethics & International Affairs* 33 (02), 127–140.
Sætra, H. S. (2020). "A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government" *Technology in society* 62, 101283.
Salem, F. A., U. F. Moreno and F. Lamnabhi-Lagarrigue (2020). "A control approach to address ethical issues on social (robotic) networks" *IFAC-PapersOnLine* 53 (2), 17016–17022.
Sand, M., J. M. Durán and K. R. Jongsma (2021). "Responsibility beyond design: Physicians' requirements for ethical medical AI" *Bioethics*.
Schanke, S., G. Burtch and G. Ray (2021). "Estimating the Impact of "Humanizing" Customer Service Chatbots" *Information Systems Research* 32 (3), 736–751.
Scheutz, M. (2017). "The Case for Explicit Ethical Agents" *AI Magazine* 38 (4), 57–64.
Schönberger, D. (2019). "Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications" *International Journal of Law and Information Technology* 27 (2), 171–203.
Schrader, D. E. and D. Ghosh (2018). "Proactively Protecting Against the Singularity: Ethical Decision Making in AI" *IEEE Security & Privacy* 16 (3), 56–63.
Sebastián, M. Á. and F. Rudy-Hiller (2021). "First-person representations and responsible agency in AI" *Synthese*, 1–19.
Seeber, I., E. Bittner, R. O. Briggs, G.-J. de Vreede, T. de Vreede, D. Druckenmiller, R. Maier, A. B. Merz, S. Oeste-Reiß, N. Randrup, G. Schwabe and M. Söllner (2018). "Machines as Teammates: A Collaboration Research Agenda" *Proceedings of the 51st Hawaii International Conference on System Sciences*.
Segun, S. T. (2021a). "Critically engaging the ethics of AI for a global audience" *Ethics and Information Technology* 23 (2), 99–105.
Segun, S. T. (2021b). "From machine ethics to computational ethics" *AI & SOCIETY* 36 (1), 263–276.
Sekiguchi, K. and K. Hori (2020). "Organic and dynamic tool for use with knowledge base of AI ethics for promoting engineers' practice of ethical AI design" *AI & SOCIETY* 35 (1), 51–71.
SFR-IA Group and CERF (2018). "Artificial intelligence and medical imaging 2018: French Radiology Community white paper" *Diagnostic and Interventional Imaging* 99 (11), 727–742.
Shank, D. B. and A. DeSanti (2018). "Attributions of morality and mind to artificial intelligence after real-world moral violations" *Computers in Human Behavior* 86, 401–411.
Shevlin, H. (2021). "How Could We Know When a Robot was a Moral Patient?" *Cambridge Quarterly of Healthcare Ethics* 30 (3), 459–471.
Shneiderman, B. (2020). "Bridging the Gap Between Ethics and Practice" *ACM Transactions on Interactive Intelligent Systems* 10 (4), 1–31.
Simon, H. A. (1995). "Artificial intelligence: an empirical science" *Artificial Intelligence* 77 (1), 95–127.
Simshaw, D. (2018). "Ethical issues in robo-lawyering: The need for guidance on developing and using artificial intelligence in the practice of law" *Hastings Law Journal* 70 (1), 173–214.

Smith, M. and S. Miller (2021). "The ethical application of biometric facial recognition technology" *AI & SOCIETY*, 1–9.
Smith, N. and D. Vickers (2021). "Statistically responsible artificial intelligences" *Ethics and Information Technology*, 1–11.
Sood, K. (2018). "The Ultimate Black Box: The Thorny Issue of Programming Moral Standards in Machines [Industry View]" *IEEE Technology and Society Magazine* 37 (2), 27–29.
Sparrow, R. (2021). "Why machines cannot be moral" *AI & SOCIETY* 36 (3), 685–693.
Stahl, B. C., A. Andreou, P. Brey, T. Hatzakis, A. Kirichenko, K. Macnish, S. Laulhé Shaelou, A. Patel, M. Ryan and D. Wright (2021a). "Artificial intelligence for human flourishing – Beyond principles for machine learning" *Journal of Business Research* 124, 374–388.
Stahl, B. C., Antoniou Josephina, M. Ryan, K. Macnish and T. Jiya (2021b). "Organisational responses to the ethical issues of artificial intelligence" *AI & SOCIETY*, 1–15.
Steels, L. and R. Lopez de Mantaras (2018). "The Barcelona declaration for the proper development and usage of artificial intelligence in Europe" *AI Communications* 31 (6), 485–494.
Stewart, C., S. K. Y. Wong and Sung Joseph JY (2021). "Mapping ethico-legal principles for the use of artificial intelligence in gastroenterology" *Journal of Gastroenterology and Hepatology* 36 (5), 1143–1148.
Stix, C. (2021). "Actionable Principles for Artificial Intelligence Policy: Three Pathways" *Science and Engineering Ethics* 27 (1), 1–17.
Stokes, F. and Amitabha Palmer (2020). "Artificial Intelligence and Robotics in Nursing: Ethics of Caring as a Guide to Dividing Tasks Between AI and Humans" *Nursing Philosophy* 21 (4), e12306.
Swanepoel, D. (2021). "The possibility of deliberate norm-adherence in AI" *Ethics and Information Technology* 23 (2), 157–163.
Swarte, T. de, O. Boufous and P. Escalle (2019). "Artificial intelligence, ethics and human values: the cases of military drones and companion robots" *Artificial Life and Robotics* 24 (3), 291–296.
Szocik, K. and R. Abylkasymova (2021). "Ethical Issues in Police Robots. The Case of Crowd Control Robots in a Pandemic" *Journal of Applied Security Research*, 1–16.
Tajalli, P. (2021). "AI ethics and the banality of evil" *Ethics and Information Technology*, 1–8.
Tigard, D. W. (2021). "Artificial Moral Responsibility: How We Can and Cannot Hold Machines Responsible" *Cambridge Quarterly of Healthcare Ethics* 30 (3), 435–447.
Timmers, P. (2019). "Ethics of AI and Cybersecurity When Sovereignty is at Stake" *Minds and Machines* 29 (4), 635–645.
Togni, G. de, S. Erikainen, S. Chan and S. Cunningham-Burley (2021). "What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care" *Social science & medicine (1982)* 277, 113874.
Tolmeijer, S., M. Kneer, C. Sarasua, M. Christen and A. Bernstein (2020). "Implementations in Machine Ethics: A Survey" *ACM Computing Surveys (CSUR)* 53 (6), 1–38.
Tomalin, M., B. Byrne, S. Concannon, D. Saunders and S. Ullmann (2021). "The practical ethics of bias reduction in machine translation: why domain adaptation is better than data debiasing" *Ethics and Information Technology*, 1–15.
Truby, J. (2020). "Governing artificial intelligence to benefit the UN sustainable development goals" *Sustainable Development* 28 (4), 946–959.
Truby, J. and R. Brown (2021). "Human digital thought clones: the Holy Grail of artificial intelligence for big data" *Information & Communications Technology Law* 30 (2), 140–168.
Umbrello, S. (2020). "Imaginative value sensitive design: Using moral imagination theory to inform responsible technology design" *Science and Engineering Ethics* 26 (2), 575–595.
Umbrello, S., M. Capasso, M. Balistreri, A. Pirni and F. Merenda (2021). "Value Sensitive Design to Achieve the UN SDGs with AI: A Case of Elderly Care Robots" *Minds and Machines*, 1–25.
Vakkuri, V., K.-K. Kemell, J. Kultanen and P. Abrahamsson (2020). "The current state of industrial practice in artificial intelligence ethics" *IEEE Software* 37 (4), 50–57.

Vallejos, E. P., R. Wortham and E. Miakinkov (2017). "When AI goes to war: youth opinion, fictional reality and autonomous weapons" *The ORBIT Journal* 1 (1), 1–20.
Vamplew, P., R. Dazeley, C. Foale, S. Firmin and J. Mummery (2018). "Human-aligned artificial intelligence is a multiobjective problem" *Ethics and Information Technology* 20 (1), 27–40.
van Berkel, N., B. Tag, J. Goncalves and S. Hosio (2020). "Human-centred artificial intelligence: a contextual morality perspective" *Behaviour & Information Technology*, 1–17.
van de Poel, I. (2020). "Embedding values in artificial intelligence (AI) systems" *Minds and Machines* 30 (3), 385–409.
van der Waa, J., S. Verdult, K. van den Bosch, J. van Diggelen, T. Haije, B. van der Stigchel and I. Cocu (2021). "Moral decision making in human-agent teams: Human control and the role of explanations" *Frontiers in Robotics and AI* 8, 1–20.
Verbeek, P.-P. (2008). "Morality in design: Design ethics and the morality of technological artifacts" *Philosophy and Design*, 91–103.
Vesnic-Alujevic, L., S. Nascimento and A. Pólvora (2020). "Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks" *Telecommunications Policy* 44 (6), 101961.
Vidgen, R., G. Hindle and I. Randolph (2020). "Exploring the ethical implications of business analytics with a business ethics canvas" *European Journal of Operational Research* 281 (3), 491–501.
Vrščaj, D., S. Nyholm and G. P. J. Verbong (2020). "Is tomorrow's car appealing today? Ethical issues and user attitudes beyond automation" *AI & SOCIETY* 35 (4), 1033–1046.
Wakunuma, K., T. Jiya and S. Aliyu (2020). "Socio-ethical implications of using AI in accelerating SDG3 in Least Developed Countries" *Journal of Responsible Technology* 4, 100006.
Wang, P. (2019). "On Defining Artificial Intelligence" *Journal of Artificial General Intelligence* 10 (2), 1–37.
Wangmo, T., M. Lipps, R. W. Kressig and M. Ienca (2019). "Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders" *BMC Medical Ethics* 20 (1), 1–11.
Wareham, C. S. (2021). "Artificial intelligence and African conceptions of personhood" *Ethics and Information Technology* 23 (2), 127–136.
Wasilow, S. and J. B. Thorpe (2019). "Artificial Intelligence, Robotics, Ethics, and the Military: A Canadian Perspective" *AI Magazine* 40 (1), 37–48.
Weber-Guskar, E. (2021). "How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners" *Ethics and Information Technology*, 1–10.
Whittlestone, J., R. Nyrup, A. Alexandrova and S. Cave (2019). "The Role and Limits of Principles in AI Ethics". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. Ed. by V. Conitzer. New York,NY,United States: Association for Computing Machinery, pp. 195–200.
Winfield, A. F., K. Michael, J. Pitt and V. Evers (2019). "Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue]" *Proceedings of the IEEE* 107 (3), 509–517.
Winfield, A. F. T. and M. Jirotka (2018). "Ethical governance is essential to building trust in robotics and artificial intelligence systems" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2133), 1–13.
Wirtz, B. W. and W. M. Müller (2019). "An integrated artificial intelligence framework for public management" *Public Management Review* 21 (7), 1076–1100.
Wolff, A., J. Gomez-Pilar, T. Nakao and G. Northoff (2019). "Interindividual neural differences in moral decision-making are mediated by alpha power and delta/theta phase coherence" *Scientific Reports* 9 (1), 1–13.
Wu, F., C. Lu, M. Zhu, H. Chen, J. Zhu, K. Yu, L. Li, M. Li, Q. Chen, X. Li, X. Cao, Z. Wang, Z. Zha, Y. Zhuang and Y. Pan (2020a). "Towards a new generation of artificial intelligence in China" *Nature Machine Intelligence* 2 (6), 312–316.
Wu, W., T. Huang and K. Gong (2020b). "Ethical Principles and Governance Technology Development of AI in China" *Engineering* 6 (3), 302–309.

Wu, Y. (2020). "Robust learning-enabled intelligence for the internet of things: A survey from the perspectives of noisy data and adversarial examples" *IEEE Internet of Things Journal* 8 (12), 9568–9579.

Yew, G. C. K. (2021). "Trust in and Ethical Design of Carebots: The Case for Ethics of Care" *International Journal of Social Robotics* 13 (4), 629–645.