Ethics and Information Technology (2025) 27:17
https://doi.org/10.1007/s10676-025-09826-5

ORIGINAL PAPER

Navigating the Nexus of ethical standards and moral values

Teresa Hammerschmidt[1]

Abstract
This study examines how ethical standards established by stakeholders such as developers and policymakers provide top-down guidance aligned with deontological ethics or utilitarian goals. It also highlights a complementary bottom-up approach, rooted in virtue ethics, in which individuals engage in ethical deliberations shaped by their moral values. Both approaches have limitations, and, at times, ethical standards can clash with moral values, thus blurring lines of responsibilities. Deontological principles may offer a structured framework, but often lack adaptability to diverse cultural contexts; bottom-up approaches foster intrinsic moral intentions, but universal applicability may be challenging, thus raising moral dilemmas. Through a theoretical literature review, this study explains how different ontological and normative ethical perceptions lead to moral dilemmas in various AI application scenarios (e.g., algorithmically managed platforms, crime detection systems, medical AI assistants). It addresses top-down and bottom-up approaches that may help account for moral dilemmas ethically. The study discusses the balance between top-down regulatory frameworks and bottom-up community-driven ethics to navigate the complex ethical landscape of AI applications, whose increasing capabilities alter expectations of AI's agency and morality. This study calls for holistic and multi-objective ethical frameworks that incorporate diverse normative ethical perspectives and recognizes context-specific ontologies throughout the AI lifecycle. It emphasizes a nuanced and context-specific combination of top-down standards (e.g., regulatory oversight, clear guidelines) and bottom-up fostering of moral values (e.g., by improving ethical knowledge). This tailored and ongoing reflection of ethical standards and moral values accounts for an ethical development, deployment, and utilization of AI technologies.

Keywords Literature review · AI ethics · AI governance · Normative ethics

Introduction

Artificial intelligence (AI) technologies have advanced from simple rule-based systems into complex models that inform, predict, and guide decisions (Berente et al., 2019; Dwivedi et al., 2021). These changes in AI capabilities have shifted the traditional power dynamics between humans and machines, moving AI from a subordinated role to one of autonomy (Giermindl et al., 2022; Maedche et al., 2019). With AI increasingly integrated into society and business, it is essential to consider ethical standards such as fairness, accountability, and transparency to ensure that AI is designed, deployed, and used responsibly (Aslan et al., 2022; Mirbabaie et al., 2022). While fairness, accountability, and transparency are often framed as values, Winfield (2019) suggests it is useful to approach them through the lens of "ethical standards," as these standards encompass shared norms and formalized principles of core ethical values developed through consensus among various stakeholders—developers, organizational managers, governmental representatives, market parties, and so on—"to, at best remove, hopefully reduce, or at very least highlight the potential for unethical impacts of consequences" of using AI (Winfield, 2019, p. 46). Even though the diversity of moral values among stakeholders poses challenges to establishing clear ethical standards (e.g., Badea & Artus, 2022; Telkamp & Anderson, 2022), these standards provide guidance within specific communities to help determine ethically responsible behaviors in contexts involving AI (Siau & Wang, 2020). This top-down approach to ethics, closely related to deontological ethics, relies on principles rooted

✉ Teresa Hammerschmidt
teresa.heyder@uni-bamberg.de

[1] Information Systems and Social Networks, University of Bamberg, An der Weberei 5, 96047 Bamberg, Germany

Published online: 24 March 2025

in duty, norm compliance, and obligation (Millar, 2016; Swanepoel, 2021; Unver, 2023).

Conversely, a complementary bottom-up approach allows individuals to engage in ethical deliberation informed by personal experiences and societal influences (Roberts & Montoya, 2022). This bottom-up approach draws on virtue ethics, where an individual's moral convictions are shaped by cultural, religious, societal, and philosophical opinions and are cultivated through intrinsic virtues such as justice, courage, temperance, and prudence (Graff, 2024; Neubert & Montañez, 2020), enabling individuals to either conform with or diverge from established standards (Bogosian, 2017).

Both the top-down implementation of ethical standards (deontology) and bottom-up acquisition of moral values (virtue ethics) have inherent limitations. While bottom-up learning risks unintentionally generalizing examples in undesired ways, top-down standards, because they depend on stakeholder consensus within specific communities, may fail to address complex situations (Constantinescu et al., 2021; Graff, 2024; Meijer et al., 2023). Additionally, the unsupervised learning capacities of AI result in AI increasingly adapts based on user data and feedback, reflecting user values and virtues, besides adhering to given ethical principles by design (Etzioni & Etzioni, 2017; Greene et al., 2023; Peters et al., 2020). This user feedback-based ethical sense-making contrasts with traditional, top-down ethical standards, underscoring the need for flexible ethical frameworks that account for diverse moral inputs (Siau & Wang, 2020).

The IEEE 7000 standards initiative, which brought together experts from various cultures and professional backgrounds, emphasizes the importance of integrating local ethical traditions when establishing AI ethics standards, reflecting a variety of normative ethical approaches, including deontology, virtue ethics, and utilitarianism (Spiekermann, 2021). The alignment of ethical standards and individual moral values are increasingly seen as essential for mitigating moral dilemmas in AI use (e.g., Bankins & Formosa, 2023; Buhmann & Fieseler, 2023; Hagendorff, 2020; Rodgers et al., 2023). Absent alignment, ethical standards may not be socially accepted and may give rise to conflicts in value (Buruk et al., 2020; Córdova & Vicari, 2022). Ethical standards can serve as guiding principles for determining responsible behavior and making subsequent judgments, but they also may clash with individuals' behavior with respect to moral blame for wrongdoing (McDonald & Pan, 2020). Some scholars argue that the primary object should not revolve solely around alignment, as this approach imposes limitations by setting a minimum threshold for acceptable ethical standards, potentially lagging behind higher moral values of individuals (Constantinescu et al., 2021; Meijer et al., 2023).

Hence, there is a need to increase our understanding of the interplay between ethical standards and moral values within human-AI interaction by developing ethical frameworks that adhere to several normative ethical theories (Wickson & Forsberg, 2015). While there has been some valuable research on different normative ethical theories in the context of human-AI interaction (e.g., Heyder et al., 2023; Serafimova, 2020; Telkamp & Anderson, 2022), further investigations are required for a deeper analysis of how ethical standards can either strengthen or impair moral values and vice versa, and how this may lead to moral dilemmas in organizations. Particularly, sociotechnical systems (STS) can be a useful framework as it views organizations as integrated systems where social entities (like people, roles, and culture) and technical elements (like tools, processes, and technologies) interact closely and shape one another, sch as moral values shape ethical standards and vice versa. Based on a systematic literature review, this work aims to answer the following research question is: *How can normative ethical theories help explain the relationship between ethical standards and moral values in human interaction with AI technologies*?[1]

This study offers a comprehensive analysis of how ethical standards can either bolster or undermine moral values within organizations and, conversely, how these dynamics can impact the ethical design, deployment and outcomes of AI technologies. In doing so, this work explains how and why moral dilemmas may arise when humans interact with AI technologies according to their different ontological and normative ethical assumptions and further examines the underlying top-down and bottom-up approaches that may help account for those conflicts.

Integrating normative ethics in sociotechnical systems: the interplay of moral values and ethical standards

This section describes how different normative ethical approaches shape human-AI interactions in organizations as STS. The sociotechnical perspective highlights interdependencies among human users, AI technologies, and broader organizational entities (e.g., stakeholders, societal, norms,

[1] This study uses the term "AI technologies" more broadly to refer to any technological artifact with AI whose action may have normatively relevant consequences. This includes specific types of AI technologies, such as autonomous vehicles. Although not all findings about one technology apply to others, this study aggregates findings regarding various AI technologies being applied in different domains and ethical contexts, as an overly focused discussion could inadvertently omit crucial aspects of responsible AI usage and outcomes.

and regulation). Building upon Leonardi's (2012) conceptualization of STS, the interaction between AI and human users can be understood as *sociomaterial entanglement* of the *material entity* (AI technology) and the *social entity* (human user). The *social subsystem* consists of additional organizational stakeholders, processes, norms, and values that further shape human-AI interactions, influencing perceptions of ethical or unethical behavior (Cecez-Kecmanovic et al., 2014). The three main ethical approaches, deontology, virtue ethics, and utilitarianism, may explain different assumptions regarding ethical or non-ethical behavior (Bankins & Formosa, 2023; Bilal et al., 2021). The following sections describe how these ethical approaches manifest within STS.

How ontological frameworks impact normative ethical perceptions

Ontology, the philosophical study of being, provides frameworks to understand the fundamental nature, properties, and relationships of entities within an STS, including *material* and *social entities*, their agency, their actions, and the outcomes of the *sociomaterial entanglement* (Orlikowski, 2010). The environment—with its actors, objects, and artifacts (i.e., *social subsystem*)—creates different realities regarding ethical AI design, deployment, and usage and, thus, determining the ontology (Cunneen et al., 2019). This, in turn, influences our understanding of what ethical standards and moral values might be deemed appropriate.

Relational and substantial ontologies are typical of the ontological frameworks that have been discussed within the sociomateriality and STS literature (Cecez-Kecmanovic et al., 2014;). Relational ontology posits that entities do not have intrinsic properties independent of their relations with other entities (Barad, 2003; Latour, 1992), whereas substantial ontology posits that entities have an independent and intrinsic existence that is characterized by stable properties (Faulkner & Runde, 2010; Mutch, 2013). In the context of human-AI interaction, relational ontology suggests that the nature and behavior of both humans and AI systems are shaped by their dynamics, and that their roles are co-constructed through the *sociomaterial entanglement* (Heyder et al., 2023). In comparison, substantial ontology would treat AI technologies and humans as distinct entities with inherent characteristics that do not change (Heyder et al., 2023).

Ontological frameworks lead to varying ethical priorities among stakeholders within an STS, such as managers, developers, politicians, and users. They may have conflicting views regarding necessary ethical standards given their heterogeneous underlying ontological frameworks. These different ontological frameworks can lead to varied preferences for ethically designed AI technologies (Kieslich et al., 2022), thereby shaping perceptions about accountability as a societal normative obligation for dealing with unintended outcomes of human-AI interaction (Horneber & Laumer, 2023; Novelli et al., 2023). Cunneen et al. (2019) argue that the analysis of ontology must precede ethical analysis. In Sect. 4, this study outlines the different ontological frameworks that guide normative ethical perceptions in the reviewed literature.

How moral values shape agency within the entanglement

Within STS, human agency (*social entity*), characterized by intentions and behavioral maxims, governs how individuals respond to technologies (Leonardi, 2012). This agency is shaped by people's ethical ideology, moral values, and responsibility perceptions, which are grounded in virtue ethics (Constantinescu et al., 2021). Consequently, people can decide to adhere to their moral values (virtue ethics), pursue their personal goals (utilitarianism), or comply with regulatory frameworks and ethical standards (deontology) (Heyder et al., 2023; Swanepoel, 2021). Human agency, thereby, reflects the interplay between internal moral values and external normative influences.

AI, in contrast, lacks intrinsic moral considerations. The agency of the AI technology (*material entity*) can be described as "the way objects act when humans provoke it" (Leonardi, 2012, p. 37), influenced primarily by given design principles and pre-defined parameters of AI models (Shneiderman, 2020). While some scholars argue that this can be mainly linked to deontology (e.g., Heyder et al., 2023; Hooker & Kim, 2018), Yu et al. (2021) contend that the adoption of AI's agency is based on hedonic principles (utilitarianism). This implies that AI is programmed to maximize outputs, resulting in people delegating more tasks to AI technologies with significant social consequences. Others argue that AI can, to a limited extent, reflect human moral behavior in line with people's virtue ethics, extending beyond ethical standards; this results from AI technologies' abilities to learn from human data inputs (e.g., Etzioni & Etzioni, 2017; Moor, 2006). Even though AI can process data that involves moral considerations of users, it does not possess an understanding of moral values similar to human agency.

Thus, people have the free will to decide according to their moral values as moral agents (Coeckelbergh, 2020), but AI lacks moral agency—"a computationally correct answer is not a moral outcome" (Puri, 2020, p. 52). In Sect. 4, this study analyses the impact of moral values on AI and human agency in the reviewed literature.

How ethical standards from the social subsystem guide entities

With respect to deontology, ethical standards are established by various stakeholders—such as developers, managers, and policymakers—within what Leonardi (2012) terms the *social subsystem*. Ethical standards serve as design guidelines for AI technologies to perform certain tasks within legal frameworks; they can also help with people's perceptions of what is "right" or "wrong," thus guiding their decisions and leading to more ethical AI use (Behdadi & Munthe, 2020; Champagne & Tonkens, 2023).

Guiding human users is required because moral agency alone does not ensure ethical behavior (Reynolds & Ceranic, 2007). People may misuse AI for personal gain or encounter moral dilemmas due to conflicting values (Lee, 2018; Mayer et al., 2020). In such cases, the presence of moral agency alone is insufficient to prevent unethical behavior, and ethical standards become essential. Robust ethical standards are required to ensure a responsible *sociomaterial entanglement*. For human users, ethical standards can cultivate moral awareness and promote ethical behavior. For AI technologies, ethical standards involve testing for biases, ensuring transparency, and establishing mechanisms to respect privacy and attribute accountability to those responsible for AI-driven actions. Some researchers, therefore, refer to an artificial moral agency that differs from traditional human agency and that is further limited and guided by human moral agency (e.g., Behdadi & Munthe, 2020; Tigard, 2021). In Sect. 4, this study analyses the impact of ethical standards on AI and human agency in the reviewed literature.

How Top-down and Bottom-up approaches interlock

A holistic approach to ethical human-AI interaction requires combining top-down establishment of ethical standards and norms (e.g., laws and governance guidelines) with bottom-up integration of societal values reflecting diverse moral communities.

AI adhering strictly to pre-defined ethical standards (deontology) may be limited by the ethical understanding of individual developers (virtue ethics of a single community) and could cause more harm than if an AI were to learn ethical reasoning from the collective input of diverse human users with varying moral perceptions (collective virtue ethics). Top-down ethical standards, like regulations or utilitarian principles for AI design, provide a necessary starting point but establish limited legal boundaries (Hagendorff, 2020; Hawkins & Mittelstadt, 2023). Incorporating Aristotle's concept of dianoetic virtues (i.e., learnable intellectual and moral excellence) and practical wisdom into AI can help to consider specific situational and circumstantial elements that influence moral behavior (i.e., the context of action) (Constantinescu et al., 2021). This method allows AI technologies to evolve and adapt based on real-world inputs and societal feedback, ensuring that the ethical reasoning embedded within these systems reflects a broader spectrum of moral values and perspectives (Cook, 2023; Etzioni & Etzioni, 2017; Saariluoma & Leikas, 2020).

Integrating top-down and bottom-up approaches can lead to human-AI interactions that are more ethical than if developers were to embed single principles in AI design (Cook, 2023; Saariluoma & Leikas, 2020). Such dual approaches enable a balance between standard ethical guidelines and adaptable, context-sensitive AI behavior that reflects diverse moral perspectives within the *sociomaterial entanglement*. This, in turn, may influence future AI regulations, thereby impacting the *social subsystem* with its top-down, deontological approaches. It is crucial to recognize different moral communities with unique ethical identities—not only for guiding human users but also for guiding AI design (Bryson, 2018; Champagne & Tonkens, 2023; Horváth, 2022).

In Sect. 4, this study shows how top-down regulatory frameworks are combined with bottom-up moral values in the literature to handle complex moral dilemmas.

Method

This study uses a theoretical literature to identify, describe, and reconfigure relevant findings into a novel framework (Paré et al., 2015). The initial search, conducted in October 2023 and updated in March 2024, followed a systematic procedure recommended by Templier and Paré (2018).

Literature Search

This study assessed five databases that index scientific publications relevant to our review to ensure broad coverage of literature: ACM Digital Library, AIS Electronic Library, Business Source Complete, Science Direct, and Web of Science. The search query for title, abstracts, and keywords—if available—was: ("human-AI") AND ("ethic*" OR "moral*") AND ("standards" OR "principles" OR "norms" OR "guidelines" OR "values" OR "conflict" or "dilemma"). The final search yielded 2,161 overall hits and 2,090 unique hits after the elimination of 71 duplicates.

Literature selection and format screening

This study conducted a format screening based on format exclusion criteria as outlined in Appendix I. Whenever possible, these criteria were applied using database filter

options. As a result, 653 articles were removed during format screening, leaving 1,437 articles.

Quality assessment

Quality was assessed based on articles having been published in outlets within the Harzing Ranking (Harzing, 2005; Mingers & Harzing, 2007) and the VHB JOURQUAL-3 Ranking (VHB, 2015). Given the focus on the two domains ethics and AI, this study further included articles published in AI- or ethics-related outlets, even if they were not part of the VHB JOURQUAL-3 Ranking (see Appendix II). Consequently, 1,009 articles were removed, leaving 428 articles.

Relevance screening and data extraction

Data were extracted only from relevant articles. The relevance screening was conducted through a full-text assessment based on three criteria (see Appendix III). That eliminated 402 articles, leaving 26 articles in the final sample (see Appendix IV). The final sample was imported into the software MAXQDA for qualitative analysis.

Data analysis

This study developed a coding scheme for qualitative analysis, following Mayring (2014). Deductive coding following a coding scheme with definitions for codes, coding rules, and anchored examples was used to structure the content of relevant articles along the constructs of the theoretical background (Mayring, 2014, pp. 95–98). This was used to code ontological frameworks, normative ethical theories, and top-down and bottom-up approaches linked to ethical standards and moral values. In addition, inductive coding was used to code agency expectations and possible pathways to balance bottom-up and top-down approaches for an ethical AI design, deployment, and utilization. A snapshot of coding is given in Appendix V.

Results

The analysis reveals that different expectations about embedding moral values in AI stems from distinct ontological frameworks. Section 4.1 examines these frameworks leading to different moral agency expectations. Section 4.2 analyzes the reasons behind the occurrence of moral dilemmas. Section 4.3 delineates seven real-world examples with context-specific analyses. Section 4.4 presents pathways for balancing top-down approaches with bottom-up approaches.

Ontological frameworks form moral agency expectations

The reviewed literature reveals that different ontological frameworks shape perceptions of ethics in AI.

AI agency is determined by STS entities

Within the analyzed literature, it becomes evident that AI's agency is described primarily through its level of abilities (e.g., cognitive abilities) and performance (e.g., accuracy). For instance, Fernandes et al. (2020) and Ibáñez and Olmeda (2022) define agency via the performance level of AI technologies compared to humans, noting that people can make errors due to tiredness and that AI can analyze larger quantities of data in less time. However, they note that AI lacks emotional intelligence, as it lacks the personal and cultural experiences essential for adapting to unforeseen situations. Thus, while human agency is based on inherent values (substantial ontology), AI's agency is determined externally by design guidelines, data, and functions (relational ontology) (Cunneen et al., 2019).

Human agency affected by AI

As Endacott and Leonardi (2022, p. 465) illustrate, "AI [communication technologies] that are configured to operate autonomously not only make decisions about communication, but act on them. For example, they do not just determine whom to invite to a particular meeting, when to invite them, how long the meeting should be, and through what communication channel the meeting should occur, but they can also proceed to invite the attendees. When AI [communication technologies] act autonomously the principal loses a certain degree of control over the communication environment." Hence, advances in AI technologies impact the sociomaterial entanglement and, consequently, human agency, which can be linked to relational ontology when considering how human agency is affected by AI.

Misleading perceptions and errors within normative spaces

Misinterpretations of AI's agency, stemming from differing ontological views, can lead to normative errors. Badea and Artus (2022) identify two types of mistakes: *instrumental* (fact-based errors due to flawed data or processes) and *intention* mistakes (ethical errors tied to developers' assumptions). Hence, the agency of AI is fundamentally shaped by the distinct nature of humans as social beings (relational ontology), forming their perceptions of humans compared to AI's unique characteristics and capabilities. Zhang et al. (2021) illustrate how participants perceive AI

as inherently utilitarian, believing AI maximizes collective welfare, which illustrates *intention* mistakes. Similarly, Giroux et al. (2022) show that anthropomorphizing AI may lead people to attribute moral agency to it mistakenly, resulting in *intention* mistakes. In contrast, *instrumental* mistakes in AI arise when biased datasets lead to skewed outcomes, causing moral concerns despite perceptions that AI operates neutrally (Hong et al., 2020; Ratoff, 2021). People fail to differentiate between AI and human agency, leading to normative errors.

Different ontological frameworks determine AI's morality and legal status

While people can act according to common-sense morality—either through learning from environmental interactions or their own pro-moral motivation (substantial ontology)—AI's abilities to acquire pro-moral motivations may be limited to its programming to adapt to users' responses (relational ontology) (Ratoff, 2021). Ratoff (2021) argues that the continuous cognitive process of hypotheses generation, prediction, and error correction may enable AI technologies to align with human values. Thereby, AI replicates, to some extent, ethical reasoning via predictive processing models, potentially granting it some form of moral agency, which may endow these technologies with a legal status (Dahiyat, 2021). This would represent a substantial ontological framework, which raises the critical question posed by Dahiyat (2021, p. 67): "At what point or degree of autonomy and sophistication would an agent acquire personality?" The legal status of AI technologies and responsibility for their actions and decisions may depend on the degree of agency attributed to them as material entities in contrast to human ones (substantial vs. relational ontology).

Why moral dilemmas arise

The review identifies several reasons for moral dilemmas arising from the mismatch between ethical standards and moral values in AI design, deployment, and usage.

Moral dilemmas linked to perceptions of AI's agency

Users' perceptions of AI's moral agency can reduce personal accountability because of lower moral intensity on an ethical issue when AI systems perform tasks autonomously (Ebrahimi & Hassanein, 2021; Telkamp & Anderson, 2022; Zhang et al., 2021). According to Dahiyat (2021, p. 60), "intelligent software agents are capable of independent action rather than merely following instructions. […]. This is why difficulties arise in deciding who should be responsible for the actions and mistakes such agents make.". The author argues that 'one-size-fits-all' regulation that does not recognize differences in AI's agency and users' corresponding perceptions can lead to a gap between legal standards (deontology) and responsibility practices (virtue ethics) (Dahiyat, 2021).

Moral dilemmas linked to different moral foundations

Ethical frameworks often emphasize principles like accountability and transparency, but Telkamp and Anderson (2022) argue that different stakeholders may prioritize different moral foundations, resulting in diverse judgments about the same AI-driven decision, wherefore AI systems often involve trade-offs between competing moral foundations (Telkamp & Anderson, 2022). Badea and Artus (2022, p. 3) argue that "[i]t is in principle impossible to specify any rule in such a way that it cannot be misinterpreted and what we mean by a rule can never be unambiguously represented." They build upon Wittgenstein's (1918) *Tractatus logico-philosophicus*[2] and describe this issue as the "interpretation problem," where abstract language fails to capture the complexity of real-world ethical concerns. These challenges demonstrate a gap between standard ethical frameworks and the diverse moral values of stakeholders.

Moral dilemmas linked to different moral maxims

Another potential source of moral dilemmas lies in striking a balance between strong regulation of AI, which may reduce users' motivation to adopt it, and weak regulation that fosters AI innovations that in the long term might maximize overall good but may comprise data privacy and risks misuse by malicious entities (Fernandes et al., 2020; Ratoff, 2021). Fernandes et al. (2020) argue that strong regulation could mandate AI technologies to adhere to utilitarian principles (utilitarianism), potentially conflicting with users' interests (virtue ethics). Conversely, a lack of regulation could spur innovation but also invade privacy issues, prompting non-AI adopters to demand stricter regulation (deontology). Thus, finding ethical standards that balance user preferences (related to their moral values) is a challenge.

Moral dilemmas linked to the absence of holistic guidance for different stakeholders that appear throughout AI's lifecycle

Another reason for moral dilemmas may stem from the absence of clear operational guidelines for the different roles of stakeholders being part of the AI lifecycle. For instance, machine learning (ML) developers focus on data

[2] Translated by David Pears and Brian McGuinness. London: Routledge, 1921.

preparation and model training, with efforts centered on developing de-biasing solutions (Gerdes, 2022). However, their exposure to AI ethics beyond the technology's design phase is limited. In comparison, Ibáñez and Olmeda (2022) found that managers rather focusing on norm-compliance and guidelines for employees, while governmental representatives focus on addressing unintended consequences for society. End-users may rather be aware of potential discrimination when using AI. Integrating ethics into AI technologies for all stakeholders throughout AI's lifecycle requires interdisciplinary approaches that consider relevant ethical standards before, during, and after implementation, including appropriate ethical measures (Gerdes, 2022).

Context-specific moral dilemmas: Understanding the conflicts' ontology

The review reveals that AI applications frequently generate moral dilemmas, often unnoticed in daily life, beyond well-known cases like the classical trolley problem of autonomous driving (Shank & Gott, 2020; Zhang et al., 2021). Table 1 provides an overview of context-specific moral dilemmas, showing how differing ontological views on AI and human agency shape perceptions and ethical conflicts.

Managing ethical Human-AI interaction

The review outlines three pathways for integrating top-down standards with bottom-up values:

Clarifying misperceptions of AI's (moral) agency

There is a need to address misleading human perceptions of AI's moral agency through training and standardized AI contracts. Despite AI technologies lacking moral agency in the way humans possess it, people often ascribe some level of morality to these machines (Ratoff, 2021). Zhang et al. (2021) emphasize the importance of aligning AI's role with human ethical expectations to prevent misleading perceptions and ensure responsible integration into society. Dahiyat (2021) further suggests standardized AI contracts to ensure that AI acts according to different people's will to increase user information on legal statuses and responsibilities when interacting with these systems. Ibáñez and Olmeda (2022) recommend organizations to consider the entire lifecycle of AI technologies when providing ethical training for several stakeholders. Also, Wang et al. (2020) propose robust data governance frameworks to increase compliance with existing regulations.

Incorporating virtue ethics into the AI design and implementation

Badea and Artus (2022) argue that virtue ethics should be embedded into AI technologies using game theory logic to interpret given standards through external moral values. Similar to games, rules as ethical standards can establish a normative space of possibilities in which players interpret rules differently and make various decisions to achieve their individual moral objectives, which creates relational moral values such as "being trusted" (virtue ethics)., extending mere principles of trustworthiness principles (deontology). Gerdes (2022) suggests another approach involving user-centered and participatory design methods to integrate domain-specific values into AI design. This approach builds on Friedman et al.'s (2017) idea of value-sensitive design, which aims to involve interdisciplinary stakeholders at various levels (conceptual, empirical, technical) of AI design to translate moral values (virtue ethics) into technically realizable design requirements. La Fors et al. (2019) further support this by advocating for a holistic approach that considers the entire lifecycle of AI technologies and involves diverse stakeholders.

Introducing regulatory oversights and multi-objective accountability

Ratoff et al. (2021) and Telkamp and Anderson (2022) highlight the need for regulatory frameworks to ensure the ethical development and use of AI technologies. Both articles argue that there is no 'one-size-fits-all' approach, as the perceptions of AI ethics vary across different contexts. Given the value-laden nature of AI, it is crucial to integrate multiple moral perspectives to create systems perceived as ethical across various cultural settings and application scenarios (Telkamp & Anderson, 2022). For instance, Ebina and Kinjo (2021) demonstrate that different countries favor different normative ethical approaches. The authors integrate three different types of social welfare functions—Bentham's, Nash's, and Rawl's type—into an automatic control model for optimal driving behavior within the context of autonomous vehicles to determine inequality aversion. Their findings indicate that while Bentham's type, which aligns with utilitarianism by saving as many people as possible, is preferred in the United States, Canada, Australia, and some Western European countries, African and Asian countries favor settings where the autonomous vehicle adheres to given rules, reflecting a deontological approach. Multi-objective approaches that share common defining characteristics of different ethical frameworks (e.g., utilitarian and deontological ones) may help to align AI decision-making standards more closely with human values, enabling better

Table 1 Overview of context-dependent potential moral dilemmas

Ontological and normative ethical assumptions	Potential moral dilemmas and conflict description	Top-down and bottom-up approaches for ethical human-AI interaction
Recruitment and selection systems (e.g., decision-support tools, management advice)		
Material entity: AI searches for candidates, screens and prioritizes applicants, automates first-round interviews, and recommends top candidates by predicting cultural fit (Fritts & Cabrera, 2021). This increases the agency of AI in recruitment but may dehumanize the process, treating applicants as a mere means to an end in a Kantian sense (Fritts & Cabrera, 2021). As Fritts and Cabrera (2021, p. 797) argue, using "artificial values" (substantial ontology) in hiring could be morally objectionable in itself. *Social entity*: The "machine heuristics" reflects employees' perception that AI is more neutral than humans, even though they may still react to AI's moral violations (Hong et al., 2020). Recruiters' perceived distance from the applicants affected by their decisions can reduce moral intensity (relational ontology), potentially harming employee-employer relationships (Ebrahimi & Hassanein, 2021; Fritts & Cabrera, 2021).	*Should AI predictions be used to make more neutral application decisions even if it means that artificial values decide upon the life of candidates?* Here, the balance lies between neutrality (*deontology*, making objective decisions free from human bias) and preserving humanity (*virtue ethics*, empathy and contextual judgment by human recruiters). This conflict is rooted in substantial ontology, as it concerns the substitution of human values with artificial ones.	*Top-down approaches*: • Limit AI's role in decisions that impact other people's lives (Fritts & Cabrera, 2021) • Guide employees to consider all stakeholders (Fritts & Cabrera, 2021) • Increase moral recognition to decision subjects and reducing acceptance of discriminatory AI recommendations (Ebrahimi & Hassanein, 2021) *Bottom-up approaches*: • Develop AI systems aligned with societal values (Hong et al., 2020) • Address value conflicts on a case-by-case basis (Fritts & Cabrera, 2021)
***Conversational agents*(e.g., Siri or Alexa)**		
Material entity: Despite AI's programmed principles, its gamified interactions can alter discourse, sometimes in unintended ways, representing relational ontology (Fritts & Cabrera, 2021). For instance, Microsoft's Tay.AI, a Twitter bot that quickly learned and replicated Nazi ideology, led to a disturbing interaction (Shank & Gott, 2020). Thus, AI agents "have the capability to make consequential decisions about communication on their own without necessarily needing input from a principal user and, in so doing, may profoundly shape how meaning is interactionally constructed" (Endacott & Leonardi, 2022, p. 464). Following Endacott and Leonardi (2022), AI agents can make autonomous decisions without direct user input (substantial ontology). *Social entity*: Respondents in the study of Shank and Gott (2020) described their smartphone's AI apps as independent, attributing some degree of agency to the technologies (substantial ontology). When a certain app failed to meet expectations, respondents reclaimed their own agency to make more careful selections (substantial ontology).	*Should AI prioritize user convenience or ensure thorough consent to prevent inappropriate content exposure or the violation of personal rights?* Here, the balance lies between the greatest happiness principles (*utilitarianism*, maximizing user satisfaction) and the protection of personal rights and specific groups of people (*deontology*, protecting personal rights, especially for vulnerable groups). For instance, one participant in the study of Shank and Gott (2020, p. 1641) described the following situation: "Two children were in the room, ages 3 and 6. At one point, Alexa misunderstood one of our questions, and began reading an erotic story. I'm not sure what emotion the kids felt, but they were clearly disturbed and confused." Here, prioritizing user rights protection is crucial, though, in other contexts, users may prefer optimized convenience, highlighting a nuanced challenge.	*Top-down approaches*: • Raise awareness of everyday situations in which moral dilemmas occur beyond the prominent examples (Shank & Gott, 2020) *Bottom-up approaches*: • Increase diplomacy with respect to problematic AI outputs; apologies to customers for misleading or unsatisfying AI outputs can help ensure those see the organization as competent (Endacott & Leonardi, 2022) • Increase knowledge of AI's operating processes to aid understanding of what leads to certain outcomes and thus maintain agency based on knowledge and explanations (Endacott & Leonardi, 2022)
***Crime detection algorithms*(e.g., algorithms that build risk scores for prisoners; AI assistants that predict crime hotspots)**		

Table 1 (continued)

Ontological and normative ethical assumptions	Potential moral dilemmas and conflict description	Top-down and bottom-up approaches for ethical human-AI interaction
Material entity: Following Martin (2019, p. 838), "algorithms are not neutral but value-laden in that they (1) create moral consequences, (2) reinforce or undercut ethical principles, or (3) enable or diminish stakeholder rights and dignity." Crime detection algorithms act autonomously, and their implications affect individuals' lives (Martin, 2019). Thus, their implications are related to humans (relational ontology). *Social entity*: Organizational representatives remain responsible, as they are the ones designing, deploying, and using AI technologies that are value-laden in nature (Martin, 2019). For instance, the company that is developing crime detection algorithms remains responsible to some extent, and the institutions that use such algorithms for public safety goals, such as the police, are also responsible (relational ontology).	*Should public institutions (e.g., policy) deploy an AI algorithm to predict crime if it risks reinforcing disproportionately targeting certain groups of people*? Here, the balance lies between maximizing public safety (*utilitarianism*, maximizing overall safety) and the risk of unfair treating certain groups (*virtue ethics*, disproportionately targeting certain demographic groups). This conflict pertains to relational ontology, as it is addresses dilemmas that arise when AI outputs affect individuals' lives and considers how institutions deploying AI should respond to these challenges.	*Top-down approaches*: • Implement regulatory oversight to ensure that ethical implications of algorithms meet societal standards (Martin, 2019) • Hold institutions accountable for the ethical implications of their algorithms, regardless of the technologies' transparency or complexity (Martin, 2019) *Bottom-up approaches*: • Engage with the community affected by AI technologies to learn their concerns (Martin, 2019) • Allow stakeholders to understand AI's processes and predictions by communicating transparently the underlying mechanisms that lead to certain predictions (Martin, 2019)
***Medical AI assistants*(e.g., brain tumor detection and segmentation tool, medical treatment adviser)**		
Material entity: When AI assistants undertake diagnostic tasks, prescribe medicines, or make other medical treatment decisions, the performance of these tasks depends mainly on these systems (substantial ontology during the operational processes) (Lebovitz et al., 2021; Yokoi et al., 2021). AI technologies can work solely with what Lebovitz et al. (2021) call the "know-what aspects of knowledge" that limit the system's outputs and usefulness, particularly in highly sensitive domains such as healthcare. *Social entity*: Healthcare is a knowledge-intensive and highly sensitive domain. The expertise of physicians and medical staff is developed over years as they accrue rich know-how knowledge through practice (substantial ontology during the operational processes) (Lebovitz et al., 2021). Thus, it is important to assess the outputs of medical AI assistants based on human experts' tacit knowledge when the underlying situation is uncertain (Lebovitz et al., 2021).	*Should a medical AI assistant be able to prioritize treatment for patients with higher survival probabilities over those with severe conditions but lower survival probabilities?* Here, the balance lies between beneficence (*utilitarianism*, maximizing the lives saved and improving overall health outcomes through the allocation of medical resources where they are most likely to result in successful treatments) and equality expectations (*virtue* ethics, respecting the intrinsic value of each individual and avoid discrimination based on health status). This conflict can be grounded in substantial ontology because of the distinct and independent aspects of knowledge that differentiate medical AI assistants (know-what) from medical staff (know-how).	*Top-down approaches*: • Foster societal debate on AI's benefits and risks in the context of medical treatment to enhance trust in the use of AI to support healthcare tasks and raise awareness of its limitations (Yokoi et al., 2021) *Bottom-up approaches*: • Do not rely on a single "ground truth" diagnosis, as the complexity and nuances of real-world medical practices require multifaceted know-how (Lebovitz et al., 2021) • Do not rely on AI's outputs alone, as this impedes learning processes; instead, recognize the importance of human experts' know-how in contexts where uncertainty and nuanced judgment are essential for making critical decisions (Lebovitz et al., 2021) • Enhance understanding of the needs of patients, physicians and other medical staff, and the way the AI can and cannot address those needs (Yokoi et al., 2021)
***Online shopping assistants*(e.g., AI-based self-services in retail, online brand community detectors)**		

Table 1 (continued)

Ontological and normative ethical assumptions	Potential moral dilemmas and conflict description	Top-down and bottom-up approaches for ethical human-AI interaction
Material entity: Human-like characteristics of AI shopping assistants increase customers' intention to engage in moral behavior (i.e., moral intention, such as reporting errors) (Giroux et al., 2022). This illustrates relational ontology rooted in anthropomorphic AI design and their underlying nudges (Wu et al., 2020). *Social entity*: Moral behavior and customer feelings of guilt decrease when they consider the AI assistant as less human-like (Giroux et al., 2022). When AI assistants aim to influence certain behaviors, punishment messages for misbehaving customers are effective in the short term but often reduce customers' future purchases (Wu et al., 2020). This illustrates how human behavior is linked to the nudges of online shopping assistants (relational ontology).	*Should people be encouraged to view AI shopping assistants as more human-like to promote more sustainable shopping behavior, even if this could lead to misplaced guilt and responsibility?* Here, the balance lies between influencing moral behavior positively (*virtue ethics*, nudging customers to make more sustainable purchasing decisions) and transparency (*deontology*, emphasizing transparency about the capabilities and limitations of AI shopping assistants to avoid unrealistic expectations). This conflict can be grounded in relational ontology due to the impact of the anthropomorphic design of AI shopping assistants on human moral behavior.	*Top-down approaches*: • Avoid "bad character animations" when designing human-like mental capacities (Giroux et al., 2022) • Consider the impact of anthropomorphism of AI technologies on customer morality (Giroux et al., 2022) *Bottom-up approaches*: • Trigger moral values and foster moral identity among communities with AI technologies' mechanisms to enhance ethical behavior rather than implementing penalizing mechanisms (Wu et al., 2020) • Use anthropomorphism to enhance customer moral actions, but be aware of the "uncanny valley" phenomenon, where increasingly the human-like characteristics of technologies can elicit feelings of discomfort when they reach a certain point of near-realism (Giroux et al., 2022)
***Algorithmically managed platforms*(e.g., algorithmic managed work, such as Uber)**		
Material entity: Algorithmically managed platforms aim to influence user behavior through diverse nudging mechanisms based on collecting vast amounts of data from the user's environment (e.g., from several devices, their GPS navigation, usage behavior, and sensor data) (Liu et al., 2021). Thus, there is user dependency on the AI's outcomes within the platform's environment, which can be linked to relational ontology. *Social entity*: To succeed within algorithmically managed platforms, users must engage with an artificial value set (Fritts & Cabrera, 2021). Ratings, real-time monitoring, and incentives may reduce the possibility of engaging in moral hazard (Liu et al., 2021). However, mechanisms such as surge pricing in ride-sharing apps can also incentivize drivers to behave opportunistically (Liu et al., 2021). This also can be linked to relational ontology.	*Should algorithmically managed platforms use nudges (e.g., surge pricing) and constant real-time monitoring to reduce users' moral hazards, despite the potential for opportunistic behavior?* Here, the balance lies between influencing moral behavior positively (*virtue ethics*, employing nudges and real-time monitoring transparently to encourage users to behave more ethically) and opportunistic behavior (*utilitarianism*, exploiting loopholes in algorithms to maximize personal gain). This conflict is rooted in relational ontology as the algorithm's mechanism is directly linked to certain human behavior.	*Top-down approaches*: • Implement regulatory restrictions to avoid fraud (Liu et al., 2021) • Fully inform users on the platform's mechanisms and provide concrete data policies as a platform provider (Liu et al., 2021) *Bottom-up approaches*: • Avoid biases that could unjustly penalize certain users. There is a risk of reinforcing existing inequalities or creating new ones through biased data and algorithmic decisions (Liu et al., 2021) • Consider the potential for ML techniques to detect opportunistic behavior (Liu et al., 2021)
***Autonomous vehicles*(e.g., autonomous driving, autonomous aircraft)**		

Table 1 (continued)

Ontological and normative ethical assumptions	Potential moral dilemmas and conflict description	Top-down and bottom-up approaches for ethical human-AI interaction
Material entity: Autonomous vehicles use pre-crash algorithms to make decisions about collision avoidance, enhancing safety, and reducing damage (Cunneen et al., 2019). This predetermined, calculated approach minimizes human error or biased decisions, even though the AI lacks moral reasoning (Cunneen et al., 2019). When faced with unavoidable accidents, autonomous systems require pre-set moral decision-making frameworks independent of the driver (substantial ontology) (Cunneen et al., 2019). *Social entity*: In unavoidable traffic incidents, human drivers have only a moment to make instinctual, sometimes morally complex decisions, leading to moral shortcuts (substantial ontology) (Cunneen et al., 2019). Rhim et al. (2020) introduce three moral-reasoning types of drivers: moral altruists who prioritize the safety of all parties (utilitarianisms), moral non-determinists who make decisions based on the specific context (virtue ethics), and moral deontologist who base their decisions on traffic rules (deontology). As automation reduces manual control, operators may become ethically disengaged, shifting accountability onto vehicle systems, creating "moral buffers" that distance humans from their actions (Holford, 2022).	*Should responsibility for autonomous transportation be distributed among all entities to increase safety despite the risk of ethical disengagement and potential failures in unprecedented emergencies?* Here, the balance lies between the maxim of minimizing overall harm (*utilitarianism*, prioritizing actions that minimize harm by distributing responsibility among all entities involved to collectively enhance safety and reliability of autonomous vehicles) and potential disagreement over values and spread responsibilities for unforeseen outcomes (*virtue ethics*, distributing responsibilities might lead to ethical disengagement, where entities may deflect responsibility). This dilemma can be linked to relational ontology since it relates responsibility for AI operations and outcomes to the diverse stakeholders involved.	*Top-down approaches*: • Empower operators with control and information to influence autonomous systems as needed (Holford, 2022) • Be aware that different countries prefer different normative structures (Ebina & Kinjo, 2021) • Establish culture-specific guidelines for autonomous vehicle practitioners (Rhim et al., 2020) *Bottom-up approaches*: • Ensure active human engagement to maintain awareness of vehicle status and operations (Holford, 2022) • Incorporate social welfare functions to address complex moral decisions, like the trolley problem, in unavoidable situations (Ebina & Kinjo, 2021)

management of multiple conflicting objectives (Vamplew et al., 2018). Such regulatory frameworks would identify moral thresholds so that AI developers and organizations are compelled to focus on creating AI that makes decisions that are not only logical but also morally understandable and do not deviate from human moral standards (Hong et al., 2020; Ratoff, 2021).

Discussion

This section provides theoretical and practical contributions.

Theoretical contributions

This work structures the findings of the reviewed literature along the entities of an STS, following the conceptualization of Leonardi (2012), illustrating its appropriateness in explaining the interplay between ethical standards and moral values.

Beginning with the **positive influence of moral values** (arising from the *social subsystem*) **on ethical standards** (guiding the *sociomaterial entanglement*): Certain work practices within the *social subsystem*, such as specific data preparation rules developed by AI designers based on their moral values, can be perceived as ethical by others, resulting in their adoption as best practices and thus guiding *sociomaterial entanglement* and transforming into broader ethical standards (e.g., Gerdes, 2022). This carries the risk, though, of unintentionally generalizing best practice examples in undesired ways (Constantinescu et al., 2021). Furthermore, attempts to integrate moral values of various stakeholders from the *social subsystem* into the design of AI technologies through user-centered and participatory methods ensure that ethical standards are better aligned with societal values (e.g., Badea & Artus, 2022; Gerdes, 2022; La Fors et al., 2019). However, embedded moral values of organizational representatives can make it challenging to develop universally applicable normative structures, resulting in value-laden ethical principles (Badea & Artus, 2022; Martin, 2019; Ratoff, 2021). This becomes evident when the ethical standards derived from one specific *social subsystem*, with its unique normative structure, are used to guide the *sociomaterial entanglement* between humans and AI that are embedded within another normative structure.

Continuing with the **positive influence of ethical standards** (in AI technologies as *material entities*) **on moral values** (of human users as *social entities*): High ethical standards in AI technologies can prevent the dehumanization of individuals, ensuring— by recognizing and respecting human dignity and the worth of every person (Fritts & Cabrera, 2021)—that people are not treated merely as

a means to an end, reflecting Kantian ethics (Kant, 1991). Ethical standards from the *social subsystem* can positively influence the moral values of users, as regulatory or organizational guidelines can enhance moral recognition among employees, fostering a sense of responsibility and ethical conduct (Ebrahimi & Hassanein, 2021). Neglect of ethical standards by AI technologies can, paradoxically, also foster moral values of users by motivating them to provide explanations and apologies to those affected by wrong AI decisions, thus reinforcing ethical reflection and moral accountability (Endacott & Leonardi, 2022). Failures to meet ethical standards in AI design can further exacerbate negative moral outcomes.

Conversely, there can also be a **negative influence of ethical standards** (from the social *subsystem*) **on moral values** (of human users as *social entities*): Ethical standards from the *social subsystem* can also adversely affect user moral values as ethical frameworks often fail to accommodate fundamental disagreements about what constitutes ethical behavior (Telkamp & Anderson, 2022). Given that ethical standards are built through consensus, this approach imposes limitations by setting a minimal threshold for acceptable ethical standards, potentially lower than the higher moral values of the individuals involved (Constantinescu et al., 2021; Meijer et al., 2023). In addition, AI regulations that are too strict and accompanied by harsh judgments from the *social subsystem* can prompt the development of coping mechanisms or a search for alternatives (Wu et al., 2020), such as loopholes, ultimately resulting in a lack of willingness to take responsibility (Shank & Gott, 2020). Thus, ethical standards may not capture the diverse and complex nature of moral values from different stakeholders, potentially clash with individuals' moral blame behavior about wrongdoing (McDonald & Pan, 2020). Thus, while top-down integration of ethical standards through the *social subsystem* may work well within one *sociomaterial entanglement*, it may not be adequate within another entanglement.

When examining the **negative influence of ethical standards** (in AI technologies as *material entities*) **on moral values** (of human users as *social entities*), it becomes apparent that too much focus on certain ethical standards can have unintended consequences. For instance, excessive transparency in the mechanisms behind AI algorithms may provoke opportunistic behavior among users if algorithms are linked to financial benefits (Liu et al., 2021). Moreover, AI technologies that adhere strictly to principles without considering contextual nuances may undermine the expertise and know-how of human users, which are crucial in uncertain situations (Fritts & Cabrera, 2021; Lebovitz et al., 2021). Therefore, while AI technologies designed with rigid principles may be beneficial for some human users within one *sociomaterial entanglement*, that rigidity could lead to immoral behavior when the entanglement is within another context.

Practical contributions

The findings offer three practical contributions.

Nuanced integration of both approaches

Developing hybrid approaches that effectively balance top-down and bottom-up measures is essential to bridge gaps between community-driven ethics and regulatory oversights. Measures should be complemented by educational programs and user training - otherwise top-down regulations risk alienating users (Endacott & Leonardi, 2022). Complex or opaque AI contracts may discourage engagement, leading users to ignore them entirely, thereby exacerbating misunderstandings about human-like AI systems.

Cultural and industrial contexts in AI applications matter

Balancing both approaches further requires careful consideration of the specific ethical demands inherent to different AI application contexts and cultural environments (Fritts & Cabrera, 2021). The industry of application and the prevailing cultural norms significantly influence ethical requirements, necessitating that AI practitioners address potential conflicts between different cultural factors when implementing AI technologies (Rhim et al., 2020). These conflicts might arise between regulatory restrictions and bottom-up approaches that alignment with the societal values of particular cultural communities (Fritts & Cabrera, 2021; Martin, 2019), wherefore recognizing ethical pluralism is crucial (Dahiyat, 2021). There are instances when localized ethical considerations are required, creating a tension between universal standards and community-driven ethics—such as when comparing the use of AI in healthcare decisions to its use in military applications.

Tailored steps for different stakeholders throughout the AI lifecycle

Ethical AI design, deployment, and utilization necessitate considering diverse stakeholders' perspectives on ethical AI practices at different stages of the AI lifecycle. *Developers* should integrate virtue ethics into AI models to align with core societal values. Participatory design methods and interdisciplinary stakeholder involvement are vital bottom-up strategies to translate moral values into technically realizable design requirements (Badea & Artus, 2022; Gerdes, 2022). *Managers* play a pivotal role in aligning AI deployment objectives with corporate social responsibility

(CSR) practices (Wang et al., 2020). Given the immaturity of "responsible AI practices," the establishment of AI ethics advisory boards as a top-down measure is recommended (Wang et al., 2020, p. 4966). *End-users* should actively engage with AI providers by voicing concerns and sharing experiences, informing bottom-up refinements to AI systems. *Governments* must implement top-down measures to educate users about their rights.

Conclusion

This study provides a comprehensive analysis of how ethical standards can either bolster or undermine moral values within organizations, and how these dynamics influence ethical AI design, deployment, and utilization. By examining the balance of top-down and bottom-up approaches, the review highlights the need for tailored, context-specific ethical frameworks.

Appendix I

Table I.1 gives an overview of the format screening.

Table I.2 Format screening

Exclusion criteria	Exclusion description	Eliminated articles
Non-scientific and non-peer-reviewed articles ($n=437$)	Gray literature, book chapters, book reviews, dissertations, tutorials, bachelor's or master's theses, workshops, posters, letter of editors, special issues, symposiums, talks, magazine, and articles of non-scientific conferences. Articles not being peer-reviewed.	ACM Digital Library: 123 articles; AIS Electronic Library: 147 articles; Business Source Complete: 36 articles; Science Direct: 25 articles; Web of Science: 106 articles
Incomplete articles ($n=60$)	Short papers or research-in-progress papers, extended abstracts (articles with fewer than ten pages). Articles without full-text availability.	ACM Digital Library: 2 articles; AIS Electronic Library: 17 articles; Business Source Complete: 7 articles; Science Direct: 3 articles; Web of Science: 31 articles
Non-English articles ($n=7$)	Articles not completely written in English are excluded.	ACM Digital Library: 1 article; AIS Electronic Library: 0 articles; Business Source Complete: 4 articles; Science Direct: 1 article; Web of Science: 1 article
Stand-alone literature reviews ($n=149$)	Standalone literature review papers following a systematic approach are excluded.	ACM Digital Library: 8 articles; AIS Electronic Library: 19 articles; Business Source Complete: 34 articles; Science Direct: 25 articles; Web of Science: 63 articles

Total: 653 articles

Appendix II

Even though not being part of the VHB JOURQUAL-3 Ranking, articles that have been published in the academic journals or conferences that either refer to "Artificial Intelligence" or "Ethics" in a technological context are included. These are:

- Artificial Intelligence and Law
- AI Communication
- AI Practitioners
- Computers and Education: Artificial Intelligence
- Engineering Application of Artificial Intelligence
- Journal of Experimental & Theoretical Artificial Intelligence#
- Artificial Intelligence in Medicine
- Ethics and Information Technology
- Science and Engineering Ethics
- AI & Society

Appendix III

Table III.1 gives an overview of the relevance screening.

Table III.3 Relevance Screening

Exclusion criteria	Exclusion description	Eliminated articles
No considerations on both standards and values ($n=177$)	Articles that did not contribute to the discussion on the interplay between ethical standards and moral values (e.g., by analyzing moral dilemmas)—by either focusing solely on ethical principles (e.g., AI regulation) or solely on moral perceptions (e.g., human virtues)—were excluded. According to the manuscript, moral values represent personal beliefs about what is "right" or "wrong," shaped by an individual's cultural, societal, or philosophical influences, and rooted in virtues like justice, courage, and prudence. These values guide personal decision-making and actions. In comparison, ethical standards are formalized principles or shared norms developed through consensus among stakeholders (e.g., developers, managers, policymakers) to ensure responsible behavior, especially in contexts like AI. These standards provide guidance on what is ethically acceptable, often aligning with deontological ethics, which focuses on duty, norms, and compliance with regulations.	ACM Digital Library: 13 articles; AIS Electronic Library: 1 article; Business Source Complete: 96 articles; Science Direct: 34 articles; Web of Science: 34 articles
No AI technologies ($n=164$)	Articles that did not mention AI technologies following the definition provided at the beginning of the Introduction as technologies that have evolved from simple rule-based systems to complex models capable of informing, self-learning, predicting, and guiding decisions and actions in various contexts, both professional and personal.	ACM Digital Library: 16 articles; AIS Electronic Library: 1 article; Business Source Complete: 70 articles; Science Direct: 17 articles; Web of Science: 60 articles
No human-AI interaction focus ($n=61$)	Articles that did not consider the interaction between humans and AI following the idea of the sociomaterial entanglement of Leonardi (2012)— by either focusing solely on the technological artifact or solely on the management of human workers—were excluded.	ACM Digital Library: 3 articles; AIS Electronic Library: 0 articles; Business Source Complete: 52 articles; Science Direct: 2 articles; Web of Science: 4 articles

Total: 402 articles

Appendix IV

Table IV.1 gives an overview of the final sample.

Table IV.4 Final Sample

Authors (Year)	Title	Outlet
Badea and Artus (2022)	Morality, Machines, and the Interpretation Problem: A Value-based, Wittgensteinian Approach to Building Moral Agents	Artificial Intelligence
Cunneen et al. (2019)	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
Dahiyat (2021)	Law and Software Agents: Are They "Agents" by the Way?	Artificial Intelligence and Law
Ebina and Kinjo (2021)	Approaching the Social Dilemma of Autonomous Vehicles with a General Social Welfare Function	Engineering Applications of Artificial Intelligence
Ebrahimi and Hassanein (2021)	Decisional Guidance for Detecting Discriminatory Data Analytics Recommendations	Information & Management
Endacott and Leonardi (2022)	Artificial Intelligence and Impression Management: Consequences of Autonomous Conversational Agents Communicating on One's Behalf.	Human Communication Research
Fernandes et al. (2020)	Norms for Beneficial AI: A Computational Analysis of the Societal Value Alignment Problem	AI Communication
Fritts and Cabrera (2021)	AI Recruitment Algorithms and the Dehumanization Problem	Ethics and Information Technology
Gerdes (2022)	A Participatory Data-centric Approach to AI Ethics by Design	Applied Artificial Intelligence
Giroux et al. (2022)	Artificial Intelligence and Declined Guilt: Retailing Morality Comparison Between Human and AI	Journal of Business Ethics

Table IV.4 (continued)

Authors (Year)	Title	Outlet
Holford (2022)	An Ethical Inquiry of the Effect of Cockpit Automation on the Responsibilities of Airline Pilots: Dissonance or Meaningful Control?	Journal of Business Ethics
Hong et al. (2020)	Sexist AI: An Experiment Integrating CASA and ELM	International Journal of Human-Computer Interaction
Ibáñez and Olmeda (2022)	Operationalising AI Ethics: How Are Companies Bridging the Gap between Practice and Principles? An Exploratory Study	AI & Society
Lebovitz et al. (2021)	Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What	MIS Quarterly
La Fors et al. (2019)	Reassessing Values for Emerging Big Data Technologies: Integrating Design-Based and Application-Based Approaches	Ethics and Information Technology
Liu et al. (2021)	Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis	Management Science
Martin (2019)	Ethical Implications and Accountability of Algorithms	Journal of Business Ethics
Ratoff (2021)	Can the Predictive Processing Model of the Mind Ameliorate the Value-alignment Problem?	Ethics and Information Technology
Rhim et al. (2020)	Human Moral Reasoning Types in Autonomous Vehicle Moral Dilemma: A Cross-cultural Comparison of Korea and Canada	Computers in Human Behavior
Shank and Gott (2020)	Exposed by AIs! People Personally Witness Artificial Intelligence Exposing Personal Information and Exposing People to Undesirable Content	International Journal of Human-Computer Interaction
Telkamp and Anderson (2022)	The Implications of Diverse Human Moral Foundations for Assessing the Ethicality of Artificial Intelligence	Journal of Business Ethics
Vamplew et al. (2018)	Human-aligned Artificial Intelligence is a Multiobjective Problem	Ethics and Information Technology
Wang et al. (2020)	Toward an Understanding of Responsible Artificial Intelligence Practices	Hawaii International Conference on System Sciences
Wu et al. (2020)	FairPlay: Detecting and Deterring Online Customer Misbehavior	Information System Research
Yokoi et al. (2021)	Artificial Intelligence Is Trusted Less than a Doctor in Medical Treatment Decisions: Influence of Perceived Care and Value Similarity	International Journal of Human-Computer Interaction
Zhang et al. (2021)	Artificial Intelligence and Moral Dilemmas: Perception of Ethical Decision-Making in AI	Journal of Experimental Social Psychology

Total: 26 articles

Appendix V

Table V.1 provides a snapshot of the coding scheme.

Table V.5 Snapshot of coding

Code	Coding Rule	Example
Relational ontology (*deductive code*)	Relational ontology posits that entities do not have intrinsic properties independent of their relations with other entities. In the context of human-AI interaction, relational ontology suggests that the nature and behavior of both humans and AI systems are shaped by their dynamics and that their roles are co-constructed through the sociomaterial entanglement.	"In fact, intelligent software agents are capable of independent action rather than merely following instructions. They further exhibit high levels of mobility, intelligence, and autonomy according to which their actions are not always completely anticipated, intended, or known by their users." (Dahiyat, 2021, p. 60)
Substantial ontology (*deductive code*)	Substantial ontology posits that entities have an independent and intrinsic existence that is characterized by stable properties. In the context of human-AI interaction, substantial ontology would treat AI technologies and humans as distinct entities with inherent characteristics and attributes that do not change based on their interactions.	"Even though people are now relying more on the decisions made by AI, it does not mean that the decisions get accepted unconditionally." (Hong et al., 2020, p. 7)
Top-down approach (*inductive code*)	Top-down approaches refer to the top-down establishment of collective ethical standards and norms accounting for several layers of AI governance guidelines, policies, and laws.	"…, AI ethics multi-disciplinary advisory board can be established to provide advice and guidance to the Board of Directors." (Wang et al., 2020, p. 4968)
Bottom-up approach (*inductive code*)	Bottom-up approaches refer to the bottom-up integration of societal values accounting for diverse moral communities and their perceptions on how to use AI ethically.	"Consequently, a participatory data-centric approach to AI Ethics by Design can engage domain experts during the system design process and, furthermore, help raise organizational awareness of the challenges related to data-driven knowledge generation." (Gerdes, 2022, p. 770)

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Not applicable.

Declarations

Declarations of interest None.

References

Aslan, A., Greve, M., & Lembcke, T. B. (2022). Let´s Do Our Bit: How Information Systems Research Can Contribute to Ethical Artificial Intelligence. *Proceedings of the Americas Conference on Information Systems (AMCIS)*. https://aisel.aisnet.org/amcis2022/sig_odis/sig_odis/10

Badea, C., & Artus, G. (2022). Morality, Machines, and the Interpretation Problem: A Value-based, Wittgensteinian Approach to Building Moral Agents. In M. Bramer & F. Stahl (Eds.), *Artificial Intelligence XXXIX* (Vol. 13652, pp. 124–137). Springer International Publishing. https://doi.org/10.1007/978-3-031-21441-7_9

Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, *185*(4), 725–740. https://doi.org/10.1007/s10551-023-05339-7

Barad, K. (2003). Posthumanist performativity: Toward an Understanding of how matter comes to matter. *Signs: Journal of Women in Culture and Society*, *28*(3), 801–831. https://doi.org/10.1086/345321

Behdadi, D., & Munthe, C. (2020). A normative approach to artificial moral agency. *Minds and Machines*, *30*(2), 195–218. https://doi.org/10.1007/s11023-020-09525-8

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2019). MANAGING ARTIFICIAL INTELLIGENCE. *MIS Quarterly*, *45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bilal, A., Wingreen, S., Sharma, R., & Jahanbin, P. (2021). Trust Development in Artificial Intelligence-based Emerging Technologies: Rise of Technomoral Virtues and Data Ethics. *Proceedings of the Australasian Conference on Information Systems (ACIS)*.

Bogosian, K. (2017). Implementation of moral uncertainty in intelligent machines. *Minds and Machines*, *27*(4), 591–608. https://doi.org/10.1007/s11023-017-9448-z

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. https://doi.org/10.1007/s10676-018-9448-6

Buhmann, A., & Fieseler, C. (2023). Deep learning Meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. *Business Ethics Quarterly*, *33*(1), 146–179. https://doi.org/10.1017/beq.2021.42

Buruk, B., Ekmekci, P. E., & Arda, B. (2020). A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Medicine Health Care and Philosophy*, *23*(3), 387–399. https://doi.org/10.1007/s11019-020-09948-1

Cecez-Kecmanovic, D., Galliers, R. D., Henfridsson, O., Newell, S., & Vidgen, R. (2014). The sociomateriality of information systems. *MIS Quarterly*, *38*(3), 809–830.

Champagne, M., & Tonkens, R. (2023). A comparative defense of Self-initiated prospective moral answerability for autonomous robot harm. *Science and Engineering Ethics*, *29*(4), 27. https://doi.org/10.1007/s11948-023-00449-x

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8

Constantinescu, M., Voinea, C., Uszkai, R., & Vică, C. (2021). Understanding responsibility in responsible AI. Dianoetic virtues and the hard problem of context. *Ethics and Information Technology*, *23*(4), 803–814. https://doi.org/10.1007/s10676-021-09616-9

Cook, T. (2023). Robust artificial moral agents and metanormativity. *Proceedings of the 2023 AAAI/ACM Conference on AI Ethics and Society*, *162*, 169. https://doi.org/10.1145/3600211.3604703

Córdova, P. R., & Vicari, R. M. (2022). Practical Ethical Issues for Artificial Intelligence in Education. In *Technology and Innovation in Learning, Teaching and Education: Third International Conference, TECH-EDU 2022, Lisbon, Portugal, August 31–September 2, 2022, Revised Selected Papers* (Vol. 1720, pp. 437–453). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-22918-3

Cunneen, M., Mullins, M., Murphy, F., & Gaines, S. (2019). 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. https://doi.org/10.1080/08839514.2018.1560124

Dahiyat, E. A. R. (2021). Law and software agents: Are they agents by the way? *Artificial Intelligence and Law*, *29*(1), 59–86. https://doi.org/10.1007/s10506-020-09265-1

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Ebina, T., & Kinjo, K. (2021). Approaching the social dilemma of autonomous vehicles with a general social welfare function. *Engineering Applications of Artificial Intelligence*, *104*, 104390. https://doi.org/10.1016/j.engappai.2021.104390

Ebrahimi, S., & Hassanein, K. (2021). Decisional guidance for detecting discriminatory data analytics recommendations. *Information & Management*, *58*(7), 103520. https://doi.org/10.1016/j.im.2021.103520

Endacott, C. G., & Leonardi, P. M. (2022). Artificial intelligence and impression management: Consequences of autonomous conversational agents communicating on one's behalf. *Human Communication Research*, *48*(3), 462–490. https://doi.org/10.1093/hcr/hqac009

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics*, *21*(4), 403–418. https://doi.org/10.1007/s10892-017-9252-2

Faulkner, P., & Runde, J. (2010). The social, the material, and the ontology of Non-Material technological objects. *European Group for Organizational Studies (EGOS) Colloquium*, *Gothenburg*(985), 4–8.

Fernandes, P., Santos, F. C., & Lopes, M. (2020). Norms for beneficial A.I.: A computational analysis of the societal value alignment problem. *AI Communications*, *33*(3–6), 155–171. https://doi.org/10.3233/AIC-201502

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends®in Human–Computer Interaction*, *11*(2), 63–125. https://doi.org/10.1561/1100000015

Fritts, M., & Cabrera, F. (2021). AI recruitment algorithms and the dehumanization problem. *Ethics and Information Technology*, *23*(4), 791–801. https://doi.org/10.1007/s10676-021-09615-w

Gerdes, A. (2022). A participatory data-centric approach to AI ethics by design. *Applied Artificial Intelligence*, *36*(1), 2009222. https://doi.org/10.1080/08839514.2021.2009222

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410–435. https://doi.org/10.1080/0960085X.2021.1927213

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, *178*(4), 1027–1041. https://doi.org/10.1007/s10551-022-05056-7

Graff, J. (2024). Moral sensitivity and the limits of artificial moral agents. *Ethics and Information Technology*, *26*(1), 13. https://doi.org/10.1007/s10676-024-09755-9

Greene, T., School, C. B., Shmueli, G., University, N. T. H., Ray, S., & National Tsing Hua University. (2023). Taking the person seriously: Ethically aware IS research in the era of reinforcement Learning-Based personalization. *Journal of the Association for Information Systems*, *24*(6), 1527–1561. https://doi.org/10.17705/1jais.00800

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Harzing, A. W. (2005). *Journal Quality List*. http://www.harzing.com/

Hawkins, W., & Mittelstadt, B. (2023). The ethical ambiguity of AI data enrichment: Measuring gaps in research ethics norms and practices. *2023 ACM Conference on Fairness Accountability and Transparency*, 261–270. https://doi.org/10.1145/3593013.3593995

Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, *32*(3), 101772. https://doi.org/10.1016/j.jsis.2023.101772

Holford, W. D. (2022). An ethical inquiry of the effect of cockpit automation on the responsibilities of airline pilots: Dissonance or meaningful control?? *Journal of Business Ethics*, *176*(1), 141–157. https://doi.org/10.1007/s10551-020-04640-z

Hong, J. W., Choi, S., & Williams, D. (2020). Sexist AI: An experiment integrating CASA and ELM. *International Journal of Human–Computer Interaction*, *36*(20), 1928–1941. https://doi.org/10.1080/10447318.2020.1801226

Hooker, J. N., & Kim, T. W. N. (2018). Toward Non-Intuition-Based Machine and Artificial Intelligence Ethics: A Deontological Approach Based on Modal Logic. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 130–136. https://doi.org/10.1145/3278721.3278753

Horneber, D., & Laumer, S. (2023). Algorithmic accountability. *Business & Information Systems Engineering*, *65*(6), 723–730. https://doi.org/10.1007/s12599-023-00817-8

Horváth, I. (2022). AI in interpreting: Ethical considerations. *Across Languages and Cultures*, *23*, 1–13. https://doi.org/10.1556/084.2022.00108

Ibáñez, J. C., & Olmeda, M. V. (2022). Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI & SOCIETY*, *37*(4), 1663–1687. https://doi.org/10.1007/s00146-021-01267-0

Kant, I. (1991). *The metaphysics of morals*. Cambridge University Press.

Kieslich, K., Keller, B., & Starke, C. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data & Society*, *9*(1), 205395172210929. https://doi.org/10.1177/20539517221092956

La Fors, K., Custers, B., & Keymolen, E. (2019). Reassessing values for emerging big data technologies: Integrating design-based and application-based approaches. *Ethics and Information Technology*, *21*(3), 209–226. https://doi.org/10.1007/s10676-019-09503-4

Latour, B. (1992). Where are the missing masses?? the sociology of a few mundane artifacts. *Shaping Technology/Building Society: Studies in Sociotechnical Change*, *1*, 225–258.

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI ground truth really true?? The dangers of training and evaluating AI tools based on experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 205395171875668. https://doi.org/10.1177/2053951718756684

Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How Are They Different? Do We Need Them? In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.), *Materiality and Organizing* (pp. 24–48). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199664054.003.0002

Liu, M., Brynjolfsson, E., & Dowlatabadi, J. (2021). Do digital platforms reduce moral hazard?? The case of Uber and taxis. *Management Science*, *67*(8), 4665–4685. https://doi.org/10.1287/mnsc.2020.3721

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mayer, A. S., Strich, F., University of Passau (Germany), Fiedler, M., & University of Passau (Germany). (2020). &. Unintended Consequences of Introducing AI Systems for Decision Making. *MIS Quarterly Executive*, 239–257. https://doi.org/10.17705/2msqe.00036

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. *Klagenfurt*.

McDonald, N., & Pan, S. (2020). Intersectional AI: A study of how information science students think about ethics and their impact. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–19. https://doi.org/10.1145/3415218

Meijer, A., Wiarda, M., Doorn, N., & Van De Kaa, G. (2023). Towards responsible standardisation: Investigating the importance of responsible innovation for standards development. *Technology Analysis & Strategic Management*, 1–15. https://doi.org/10.1080/09537325.2023.2225108

Millar, J. (2016). An ethics evaluation tool for automating ethical Decision-Making in robots and Self-Driving cars. *Applied Artificial Intelligence*, *30*(8), 787–809. https://doi.org/10.1080/08839514.2016.1229919

Mingers, J., & Harzing, A. W. (2007). Ranking journals in business and management: A statistical analysis of the Harzing data set. *European Journal of Information Systems*, *16*, 303–316.

Mirbabaie, M., Brendel, A. B., Faculty of Business and Economics, Technische Universität Dresden, Hofeditz, L., & Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen. (2022). &. Ethics and AI in Information Systems Research. *Communications of the Association for Information Systems*, *50*(1), 726–753. https://doi.org/10.17705/1CAIS.05034

Moor, J. H. (2006). *The nature, importance, and difficulty of machine ethics*. IEEE INTELLIGENT SYSTEMS.

Mosakas, K. (2021). On the moral status of social robots: Considering the consciousness criterion. *AI & SOCIETY*, *36*(2), 429–443. https://doi.org/10.1007/s00146-020-01002-1

Mutch, A. (2013). Sociomateriality—Taking the wrong turning? *Information and Organization*, *23*(1), 28–40. https://doi.org/10.1016/j.infoandorg.2013.02.001

Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. *Business Horizons*, *63*(2), 195–204. https://doi.org/10.1016/j.bushor.2019.11.001

Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: What it is and how it works. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01635-y

Orlikowski, W. J. (2010). The sociomateriality of organisational life: Considering technology in management research. *Cambridge Journal of Economics*, *34*(1), 125–141. https://doi.org/10.1093/cje/bep058

Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), 34–47. https://doi.org/10.1109/TTS.2020.2974991

Puri, A. (2020). Moral imitation: Can algorithm really be ethical?? *Rutgers Law Record*, *48*(1), 47–57.

Ratoff, W. (2021). Can the predictive processing model of the Mind ameliorate the value-alignment problem? *Ethics and Information Technology*, *23*(4), 739–750. https://doi.org/10.1007/s10676-021-09611-0

Reynolds, S. J., & Ceranic, T. L. (2007). The effects of moral judgment and moral identity on moral behavior: An empirical examination of the moral individual. *Journal of Applied Psychology*, *92*(6), 1610–1624. https://doi.org/10.1037/0021-9010.92.6.1610

Rhim, J., Lee, G., & Lee, J. H. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, *102*, 39–56. https://doi.org/10.1016/j.chb.2019.08.010

Roberts, J. S., & Montoya, L. N. (2022). Contextualizing artificially intelligent morality: A Meta-Ethnography of Top-Down, Bottom-Up, and hybrid models for theoretical and applied ethics in artificial intelligence. *ArXiv*. ArXiv:2204.07612. http://arxiv.org/abs/2204.07612

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2023). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, *33*(1), 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Saariluoma, P., & Leikas, J. (2020). Designing Ethical AI in the Shadow of Hume's Guillotine. In T. Ahram, W. Karwowski, A. Vergnano, F. Leali, & R. Taiar (Eds.), *Intelligent Human Systems Integration 2020* (Vol. 1131, pp. 594–599). Springer International Publishing. https://doi.org/10.1007/978-3-030-39512-4_92

Serafimova, S. (2020). Whose morality? Which rationality? Challenging artificial intelligence as a remedy for the lack of moral enhancement. *Humanities and Social Sciences Communications*, *7*(1), 119. https://doi.org/10.1057/s41599-020-00614-8

Shank, D. B., & Gott, A. (2020). Exposed by AIs! people personally witness artificial intelligence exposing personal information and exposing people to undesirable content. *International Journal of Human–Computer Interaction*, *36*(17), 1636–1645. https://doi.org/10.1080/10447318.2020.1768674

Shneiderman, B. (2020). Human-Centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 109–124. https://doi.org/10.17705/1thci.00131

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management*, *31*(2), 74–87. https://doi.org/10.4018/JDM.2020040105

Spiekermann, S. (2021). What to expect from IEEE 7000: The first standard for Building ethical systems. *IEEE Technology and Society Magazine*, *40*(3), 99–100. https://doi.org/10.1109/MTS.2021.3104386

Swanepoel, D. (2021). The possibility of deliberate norm-adherence in AI. *Ethics and Information Technology*, *23*(2), 157–163. https://doi.org/10.1007/s10676-020-09535-1

Telkamp, J. B., & Anderson, M. H. (2022). The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal of Business Ethics*, *178*(4), 961–976. https://doi.org/10.1007/s10551-022-05057-6

Templier, M., & Paré, G. (2018). Transparency in literature reviews: An assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems*, *27*(5), 503–550. https://doi.org/10.1080/0960085X.2017.1398880

Tigard, D. W. (2021). Artificial moral responsibility: How we can and cannot hold machines responsible. *Cambridge Quarterly of Healthcare Ethics*, *30*(3), 435–447. https://doi.org/10.1017/S0963180120000985

Unver, M. B. (2023). Reebuilding 'ethics' to Govern AI: How to re-set the boundaries for the legal sector? *International Conference on Artificial Intelligence and Law (ICAIL)*. https://doi.org/10.1145/3594536.3595156

Vamplew, P., Dazeley, R., Foale, C., Firmin, S., & Mummery, J. (2018). Human-aligned artificial intelligence is a multiobjective problem. *Ethics and Information Technology*, *20*(1), 27–40. https://doi.org/10.1007/s10676-017-9440-6

VHB (2015). *VHB-JOURQUAL3 Ranking*. Verband der Hochschullehrerinnen und Hochschullehrer für Betriebswirtschaft e.V. (VHB). https://vhbonline.org/service/vhb-jourqual/vhb-jourqual-3/gesamtliste

Wang, Y., Xiong, M., & Olya, H. (2020). Toward an Understanding of responsible artificial intelligence practices. *Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2020.610

Wickson, F., & Forsberg, E. M. (2015). Standardising responsibility?? The significance of interstitial spaces. *Science and Engineering Ethics*, *21*(5), 1159–1180. https://doi.org/10.1007/s11948-014-9602-4

Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, *2*(2), 46–48.

Wittgenstein, L. (1918). *Tractatus logico-philosophicus*.

Wu, W., Huang, T., & Gong, K. (2020). Ethical principles and governance technology development of AI in China. *Engineering*, *6*(3), 302–309. https://doi.org/10.1016/j.eng.2019.12.015

Yokoi, R., Eguchi, Y., Fujita, T., & Nakayachi, K. (2021). Artificial intelligence is trusted less than a Doctor in medical treatment decisions: Influence of perceived care and value similarity.

International Journal of Human–Computer Interaction, *37*(10), 981–990. https://doi.org/10.1080/10447318.2020.1861763

Yu, B., Vahidov, R., & Kersten, G. E. (2021). Acceptance of technological agency: Beyond the perception of utilitarian value. *Information & Management*, *58*(7), 103503. https://doi.org/10.1016/j.im.2021.103503

Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., & Dafoe, A. (2021). Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers. *Journal of Artificial Intelligence Research*, *71*. https://doi.org/10.1613/jair.1.12895