Technological Forecasting & Social Change 204 (2024) 123403

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

# The impact of intelligent decision-support systems on humans' ethical decision-making: A systematic literature review and an integrated framework

Franziska Poszler [a,*], Benjamin Lange [b]

[a] Peter Löscher Chair of Business Ethics, Institute for Ethics in Artificial Intelligence, Technical University of Munich (TUM), Arcisstraße 21, 80333 München, Germany
[b] Professorship for Ethics in AI, Ludwig-Maximilian University of Munich (LMU), Geschwister-Scholl-Platz 1, 80539 München, Germany

ARTICLE INFO

*Keywords:*
Moral psychology
Ethical decision-making
Moral enhancement
Intelligent decision-support systems
AI ethics

ABSTRACT

With the rise and public accessibility of AI-enabled decision-support systems, individuals outsource increasingly more of their decisions, even those that carry ethical dimensions. Considering this trend, scholars have highlighted that uncritical deference to these systems would be problematic and consequently called for investigations of the impact of pertinent technology on humans' ethical decision-making. To this end, this article conducts a systematic review of existing scholarship and derives an integrated framework that demonstrates how intelligent decision-support systems (IDSSs) shape humans' ethical decision-making. In particular, we identify resulting consequences on an individual level (i.e., deliberation enhancement, motivation enhancement, autonomy enhancement and action enhancement) and on a societal level (i.e., moral deskilling, restricted moral progress and moral responsibility gaps). We carve out two distinct methods/operation types (i.e., process-oriented and outcome-oriented navigation) that decision-support systems can deploy and postulate that these determine to what extent the previously stated consequences materialize. Overall, this study holds important theoretical and practical implications by establishing clarity in the conceptions, underlying mechanisms and (directions of) influences that can be expected when using particular IDSSs for ethical decisions.

## 1. Introduction

That technological applications (for example, obstetric ultrasound or traditional decision-support systems) can shape human ethical actions, interpretations and decisions is a well-known, discussed phenomenon (e.g., Meredith and Arnott, 2003; Verbeek, 2008). However, rapid advancements in artificial intelligence (AI) have sparked the power and 'intelligence' of decision-support systems (Phillips-Wren, 2012) and enabled humans to directly and completely outsource ever more of their decision-making (Erler and Müller, 2021), even those that carry ethical dimensions (Giubilini and Savulescu, 2018). Pertinent applications of such intelligent decision-support systems (IDSSs) are, for example, systems that lead users through their independent moral deliberation process as a sort of supportive instructor (Lara and Deckers, 2020). Similarly, more invasive decision-support systems called 'artificial moral advisors' that provide concrete moral advice to their users are put forward (Savulescu and Maslen, 2015). "[I]f AI ethical advisors become widespread in the future […] it may even become an everyday activity to seek advice from such artificial moral experts" (Rodríguez-López and Rueda, 2023; p.7). It could be argued that now – with the launch of ChatGPT, which can offer moral advice to anyone upon request (Krügel et al., 2023) – the phenomenon of humans outsourcing ethical decision-making to IDSSs has already become a reality. Since ethical decision-making is a fundamental ability for humans (that, among others, plays a crucial role in shaping morale), it is imperative to closely monitor this trend and its resulting societal ramifications (Eisikovits and Feldman, 2022). Understanding (the sources and methods of) IDSSs' influence on human ethical decision-making empowers us to anticipate and proactively shape their impact toward the direction that we desire, embracing the ethos of human-centered design (Auernhammer, 2020).

Correspondingly, research has started to investigate the impact of IDSSs on humans' ethical decision-making, but is divided to what extent such systems exclusively hold positive implications. On a positive note, it is argued that such systems may allow individuals to make more informed, value-aligned and rational decisions by providing them with information on morally relevant factors in a decision situation. In

* Corresponding author.
*E-mail address:* franziska.poszler@tum.de (F. Poszler).

https://doi.org/10.1016/j.techfore.2024.123403
Received 15 September 2022; Received in revised form 3 February 2024; Accepted 8 April 2024

contrast, when relying on particular systems, individuals may (learn to) submit to decisions that are inconsistent with their own values but rather in line with external (biased) values (Vallor, 2015). Given the prevailing ambiguity and range of indicated influences of IDSSs on humans' ethical decision-making, consolidating existing findings seems timely to foster knowledge and keep pace with the advancement of the field. After all, it is essential to anticipate in what way such systems will affect humans' ethical decision-making, how they precisely accomplish this and assess whether we should refrain from developing/adopting particular types of IDSSs (Volkman and Gabriels, 2023). As existing literature is scattered and does not provide a uniform understanding of the human-technology interaction in this context, a comprehensive synthesis and analysis of pertinent publications is needed. Therefore, in this study, we conducted a systematic literature review to answer the following questions:

- ***Research question 1***: In what way can intelligent decision-support systems hinder or facilitate humans' ethical decision-making?
- ***Research question 2***: By what mechanisms do intelligent decision-support systems hinder or facilitate humans' ethical decision-making?

By addressing these questions, main contributions of this study are identified underlying sources (i.e., human creators and technological system) and types of operations (i.e., process-oriented navigation and outcome-oriented navigation) that can emerge from IDSSs (thereby, providing answers to research question 2). Furthermore, the findings point to resulting outcomes for humans' ethical decision-making on an individual level (i.e., deliberation enhancement, motivation enhancement, autonomy enhancement and action enhancement) and on a societal level (i.e., moral deskilling, restricted moral progress and moral responsibility gaps) (thereby, providing answers to research question 1). Overall, this article aims to generate theoretical and practical insights for a range of stakeholders, including the users/broader public affected by the IDSSs' ethical decisions, software designers and other practitioners shaping these technologies as well as policymakers and academics concerned with IDSSs' role for humans' ethical decision-making. For example, the broader public can engage in more informed and reflected use of IDSSs, while developers can look out for the here identified negative outcomes and mitigate these by adopting the here sketched technological features during the design phase.

This article is structured as follows. Section 1 provides an introduction highlighting the relevance, need and actuality of the discussed topic. Section 2 introduces the theoretical background on ethical decision-making and the link between IDSSs and moral enhancement. Section 3 states the methodology used in this article (i.e., a systematic literature review). Section 4 depicts key results of the literature review, which constitute identified theoretical underpinnings, sources and methods of influence that emerge from IDSSs as well as resulting outcomes for humans' ethical decision-making on an individual and a societal level. Section 5 illustrates a holistic framework that integrates all findings by postulating interrelations and highlights key theoretical and practical implications. Lastly, in Section 6, a short conclusion will be drawn.

## 2. Theoretical background

This section provides background information on the ethical decision-making process (see 2.1. Ethical decision-making) and the link between IDSSs and moral enhancement (see 2.2. Intelligent decision-support systems and moral enhancement) to introduce key terms and clarify the scope of this article.

### 2.1. Ethical decision-making

Throughout (professional) life, individuals are confronted with dilemmas or end-of-life situations (Inthorn et al., 2015; Kvalnes, 2015), which require them to engage in ethical decision-making. Ethical decision-making can be defined as "a process by which individuals use their moral base to determine whether a certain issue is right or wrong" (Carlson et al., 2009; p.536). In the past, this process has been elucidated by relying on rationalist theories of ethical decision-making and theories emphasizing intuition. Concerning the former, Rest (1986) proposed a model for individual ethical decision-making and behavior that consists of four steps: moral awareness, moral reasoning, moral intent and moral behavior. First, individuals need to recognize an ethical issue at hand (moral awareness) (Rest, 1986), which, among others, depends on the moral intensity of a situation (Jones, 1991). Afterward, individuals engage in moral reasoning by extracting, weighing and integrating morally relevant information and drawing from coexisting standards and ethical principles (Bandura, 1991). Individuals then establish moral intent by resolving to place moral concerns ahead of others and by discovering moral priorities (e.g., Rest, 1986; Campbell, 2017). As a result of this, individuals perform moral behavior, meaning they decide/act according to their moral intent (Rest, 1986). The act of "choosing in accord with [one's] own moral convictions or principles" has been associated with an individual's (moral) autonomy (Wallach and Allen, 2009).

Over time, studies stressed the importance of intuition, heuristics, and automatic processes as additional sources underlying ethical decision-making (Cushman et al., 2006; Haidt, 2001; Kahneman, 2003; Schwartz, 2016). In particular, it is postulated that intuition and affects antecede the moral reasoning process (Zollo et al., 2017), so that human judgment is primarily based on intuition instead of deliberation (Krügel et al., 2023). Despite individuals' reasoning capabilities or intuition, they still sometimes make logical errors or overlook normative features in particular decision situations (Higgins et al., 1984; O'Neill et al., 2022), which may lead to the moral lag problem, in which humans are "not as moral as [they] could or should be" (Klincewicz, 2016; p.172). Thus, "it is perhaps not surprising that researchers have attempted to improve the quality of decisions" through the assistance of technology (Phillips-Wren, 2012; p.1).

### 2.2. Intelligent decision-support systems and moral enhancement

Decision support systems (DSS) refer to a broad category of computer systems that inform and "assist decision makers to utilize data, models and knowledge to solve semi-structured, ill-structured, or unstructured problems" (Phillips-Wren, 2013; p.5). Traditional approaches to these systems include, for example, rule-based expert systems that simply reflect and communicate the knowledge of experts in a specific subject matter to its users (e.g., "if you perceive this, then do that") (Gonzalez Fabre et al., 2021; p.294). In the twenty-first century, intelligent decision-support systems (IDSS) that utilize AI techniques emerged (Stefan and Carutasu, 2020). These advances in AI have sparked the power, sophistication and autonomy of these decision-support systems so that they can assist humans in many more areas and (ethical) decision scenarios (Phillips-Wren, 2012). Pertinent applications span from automated weapons that help soldiers determine whether a certain target shall be hit (Vallor, 2015), clinical decision-support systems that help healthcare professionals distribute scarce medical resources (Erler and Müller, 2021) to artificial moral advisors that provide concrete moral advice to help users with their personal matters (Savulescu and Maslen, 2015). These developments give rise to the phenomenon of moral enhancement through AI (Lara, 2021), which in the past was mainly addressed in association with biomedical interventions (Savulescu and Maslen, 2015). Moral enhancement entails interventions that aim to improve an individual's moral capacities, ultimately leading to moral improvement (e.g., better motives, increased understanding of what is right and higher frequency of right actions) (DeGrazia, 2014) and thus, is closely linked to the process and outcome of ethical decision-making. However, that IDSSs purely hold positive implications for individuals' ethical decision-making is viewed critically.

Scholars have started to categorize AI systems to better understand the relationship between IDSSs and ethical decision-making. These distinctions differentiate between decision-support systems that lead users through their independent moral deliberation, systems that provide moral advice based on the user's values or systems that supplant entirely the user's ethical decision-making (Lara and Deckers, 2020; Volkman and Gabriels, 2023). Similarly, Liu et al. (2022) distinguished between 'narrow AI moral enhancement', where technological systems help users make ethical decisions and 'broad AI moral enhancement', where systems aim to improve users' moral character. While these articles highlight a few benefits and drawbacks of particular systems, a systematic analysis of why and how IDSSs shape humans' ethical decision-making does not yet exist to the authors' knowledge. Therefore, this article aims to synthesize knowledge on why (i.e., due to which sources and through which features exactly) different operations of IDSSs lead to specific kinds of enhancements throughout the entire ethical decision-making process of their users and what long-term repercussions can be expected on a societal level.

## 3. Research method

The underlying methodology of this article is a systematic literature review of publications focusing on IDSSs that aim to assist or take over ethical decision-making. In particular, publications that investigate the impact of such systems on humans' ethical decision-making were examined. Considering the prevailing ambiguity and range of assumed influences, a review of this sort can serve as a useful methodology that analyzes past research to facilitate theory development and research agendas in the hopes of advancing knowledge and preparing for the future (Webster and Watson, 2002). To do so, this article conducts a three-stage iterative process (adapted from Theurer et al., 2018 and Poszler et al., 2023), consisting of the following steps: identification of the relevant literature (see 3.1. Identification of relevant literature), structural and in-depth content analysis of the literature (see 3.2. Structural & content-based analysis of literature) and integration of the literature (see 3.3. Integration & synthesis of literature). In addition, this section elaborates on the validity and reliability of the utilized methodology (see 3.4. Validity & reliability of methodology).

### 3.1. Identification of relevant literature

In line with the scope of the study and based on a small sample of key publications in the research field (e.g., Lara, 2021; Klincewicz, 2016; Savulescu and Maslen, 2015; Verbeek, 2006), previously stated keywords were identified as relevant search terms. These included, for example, "ethical decision making", "moral enhancement" or "technological mediation" in combination with terms such as "artificial intelligence", "algorithm", "decision-support system" or "moral advisor" (see Appendix A – Overview of the literature search process for the complete list of keywords, the utilized search string and obtained number of hits). The literature search was conducted via the databases *Scopus*, *Web of Science* (resp. Science Citation Index Expanded, Social Sciences Citation Index, Conference Proceedings Citation Index - Science, Conference Proceedings Citation Index - Social Science & Humanities) and *IEEE Xplore* to ensure coverage of relevant and up-to-date publications.

The search was not limited to a specific date and all publications written in English until the 31st of December 2023 were included. To enhance the review and provide an intentionally broad view of the topic, in addition to peer-reviewed academic journal publications, the search included edited books and book chapters. Due to the novelty of the subject matter, conference proceedings and practitioner-oriented articles were also considered. Content-wise, publications were considered *in scope* if they investigate the influence of intelligent decision-support systems on individuals' ethical decision-making processes. Only those publications that discussed IDSSs (i.e., systems that function based on AI techniques) and that aim to assist or fully automate humans' decision-making were included. In contrast, *exclusion criteria* included content-related factors, for example, if a particular publication did not concretely refer to technological decision-support systems (e.g., Henslee et al., 2022) or did not address the influence on ethical decision-making (e.g., Kempt et al., 2023). Similarly, we excluded articles that focused on bioenhancement (e.g., Danaher, 2019) or that addressed how to technically integrate ethical decision-making into systems (e.g., Badea, 2022). Adapted from the PRISMA flowchart for study selection (Page et al., 2021), the complete list of exclusion criteria and the entire search funnel are itemized in Fig. 1.

As illustrated in Fig. 1, the literature search resulted in a total of 1.731 hits, 1.211 hits without duplicates, respectively. After the title, abstract and full paper analysis, 34 journal articles, book chapters and conference proceedings were considered relevant to the topic at hand. In addition to the search in the three stated databases, forward- and backward searches were conducted by reviewing the reference lists of the initially identified publications as well as utilizing Google Scholar's "cited by" function. These forward- and backward searches revealed another 11 relevant contributions. Therefore, a total of 45 publications were identified as the baseline for this review and thus, subjected to further analysis.

### 3.2. Structural & content-based analysis of literature

The *structural analysis* assesses and summarizes formal factors of the identified literature such as the year, type (i.e., book chapter, conference proceeding or journal article) and discipline of publication (i.e., Arts and Humanities, Computer Science, Engineering, Ethics & Philosophy, Multidisciplinary or Social Sciences) as well as their underlying utilized methodology (i.e., conceptual/theory, empirical: qualitative or empirical: quantitative). These analyses are summarized and visualized as background information in Appendix B – Descriptive/structural analysis of literature.

For the *content-based analysis* of the literature, all 45 articles were read in their entirety and coded in light of this article's research topic, which is the influence of IDSSs on humans' ethical decision-making. In more detail, an inductive coding and clustering process was conducted in an iterative and circular manner. Inductive coding is a bottom-up method in which no predetermined codes are implied, but all codes are developed from the datasets themselves, enabling the emergence of themes right from the raw textual data (Fereday and Muir-Cochrane, 2006). In line with Gioia et al. (2013), this coding process was structured into three steps: identification of 1st-order codes, organization of 1st-order codes into 2nd-order themes and distillation of 2nd-order themes into aggregated dimensions. The 1st-order analysis aims to develop codes that adhere faithfully to the stated text. As the research progresses, 2nd-order coding seeks to establish similarities and differences among the many codes to arrange them accordingly. Lastly, 2nd-oder themes are refined even further by developing aggregate dimensions, allowing the building of a *data structure* (Fig. 2) to visualize the conducted process from raw data to final dimensions (Gioia et al., 2013).

This entire content-based coding process was performed manually via the MAXQDA software. To ensure the completeness of codes, in addition to manual coding, the 'automatic text search and coding' feature was utilized to identify overlooked text sections that contained the previously derived codes.

### 3.3. Integration & synthesis of literature

Five broad themes were identified for categorizing the respective literature, namely: theoretical underpinnings, sources of influence, methods of influence as well as subsequent direct and secondary outcomes (as illustrated in Fig. 2). 'Theoretical underpinnings' constitute theories and concepts that were mentioned in past literature as

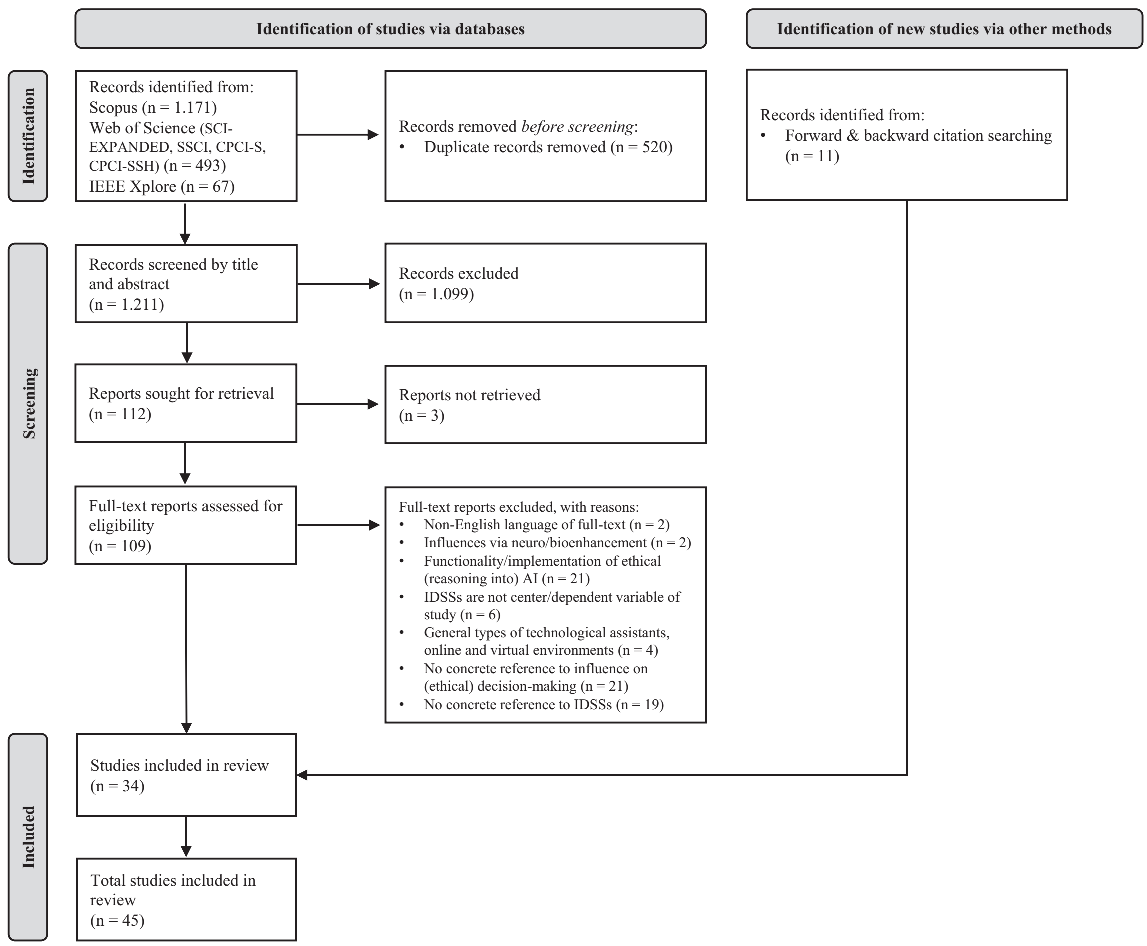


**Fig. 1.** Search funnel of the systematic literature review, displaying the consulted databases, screening process (e.g., exclusion criteria) and the final number of included publications (adapted from Page et al., 2021).

underlying assumptions or as a baseline for explaining particular interrelations between IDSSs and humans' ethical decision-making. The 'sources of influence' refer to stated origins and corresponding underlying characteristics from which IDSSs derive their capacity to influence human ethical decision-making in the first place. 'Methods of influence' comprises references in the literature that address the nature, operation and utilized features by which IDSSs guide users through the decision-making process. The dimension of 'direct outcomes' covers the immediate impacts that IDSSs are reported to have on ethical decision-making on an individual level (i.e., for the user in a specific decision situation). Lastly, 'secondary outcomes' refer to indirect and long-term repercussions for ethical decision-making on a societal level. What each dimension constitutes in more detail is summarized in the findings section (see 4. Findings).

To synthesize the literature, the next step involved the transformation of the static data structure into a *dynamic, grounded theory model* since "[t]he key question for us as model builders is how to account for not only all the major emergent concepts, themes, and dimensions, but also for their dynamic interrelationships" (Gioia et al., 2013; p.22). For example, by visualizing boxes with connecting arrows, such a synthesis aims to holistically conceptualize relations between all emergent themes to explain the phenomenon or question of interest (i.e., the influence of IDSSs on humans' ethical decision-making). The derived model permits the chance to gain theoretical insights that would not have transpired simply by examining the data structure (Gioia et al., 2013). The generated grounded model of this article is illustrated in Fig. 3 in the discussion section.

### *3.4. Validity & reliability of methodology*

Validity refers to the "'appropriateness' of the tools, processes, and data" utilized during research, while reliability involves the replicability of the research process and corresponding results (Leung, 2015; p.325). *Validity measures* consulted in this article include, for example, the reliance on and adoption of established methods for conducting systematic literature reviews such as the ones put forward by Gioia et al. (2013) or Webster and Watson (2002). In addition, existing literature reviews such as the ones published by Theurer et al. (2018) or Corley and Gioia (2004) were drawn on as orientation to crosscheck the legitimacy of utilized tools, data analysis and document writing. Furthermore, the list of keywords that underlie the database search was agreed upon among the co-authors to avoid selection bias, i.e., overlooking terms that are relevant to the topic at hand. Similarly, in line with Leung (2015), triangulation among researchers was conducted by repeatedly consulting generated codes and the derived model among the co-authors as well as with fellow researchers during research colloquia. Furthermore, a preliminary version of this article was presented and discussed with the science community at the international conference "2023 Forum on Philosophy, Engineering & Technology". *Reliability* was warranted by comprehensively documenting the literature search process, which included the disclosure of utilized keywords and databases (see Appendix A – Overview of the literature search process), exclusion criteria (see Fig. 1) as well as the referencing of consulted methodologies. This allows other researchers to replicate or update this study in the future (Brocke et al., 2009) so that similar results to the ones sketched in the following sections can be achieved.

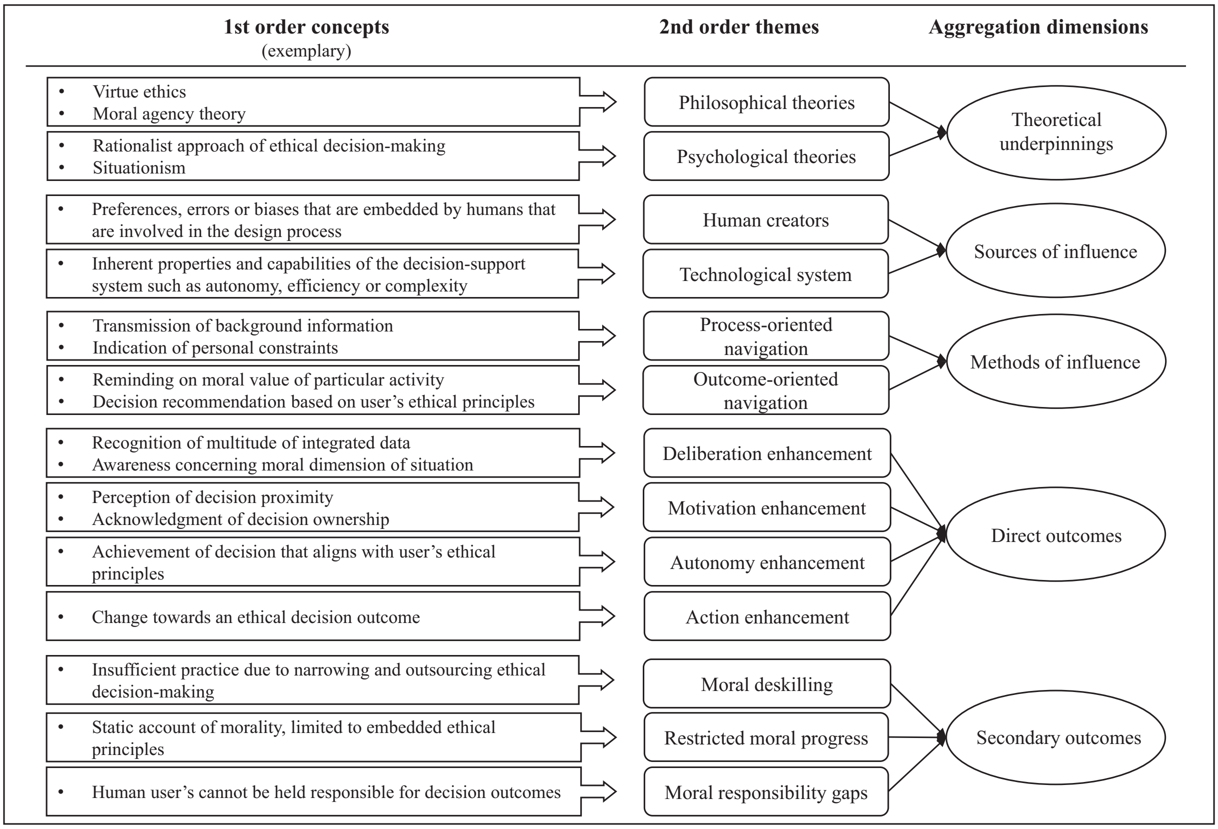


**Fig. 2.** Data structure illustrating the identified 1st-order concepts and 2nd-order themes that underlie the five aggregated dimensions (reproduced from Corley and Gioia, 2004).

## 4. Findings

This section summarizes the analyzed literature by elaborating on the five identified categories of themes. In particular, mentioned theories and concepts that were utilized as underlying assumptions or as a baseline for explaining distinct influences (theoretical underpinnings) (see 4.1. Theoretical underpinnings), stated sources of influence (see 4.2. Types of sources) and different methods through which IDSSs can exert influence (see 4.3. Methods of influence) are outlined, thereby offering answers to research question 2. Furthermore, postulated outcomes for humans' ethical decision-making on an individual level (see 4.4. Direct outcomes and on a societal level (see 4.5. Secondary outcomes) are disclosed, thereby answering research question 1.

### *4.1. Theoretical underpinnings*

The key theoretical underpinnings that scholars stated prevalently in the identified literature can be clustered into two categories: *philosophical theories* and *psychological theories* (as illustrated in Table 1). These theories and concepts can serve as a reference to examine how the identified sources and methods of influence relate to the direct and secondary outcomes (see the takeaways put forward in 5.1. Proposed interrelations).

#### *4.1.1. Philosophical theories*

Concerning philosophical theories, scholars have made remarks to the *technological mediation theory,* which suggests that technological artifacts and individuals reciprocally act upon each other (e.g., De Boer and Kudina, 2021). Furthermore, *virtue ethics* was emphasized by referring to the necessity of repeating moral practice in order to cultivate moral virtues: "[T]he virtue of honesty, for example, can only be acquired through repeated practice of truth-telling" (Vallor, 2013; p. 476). Eventually, through habituation, individuals will develop practical wisdom that directs desires intelligently and help rapidly adapt to each situation's unique moral demands (Vallor, 2013). Moreover, Rawls' (1971) concept of *reflective equilibrium* has been mentioned in the context of IDSSs. It emphasizes a method or process of deliberative mutual adjustments among one's considered judgments or intuitions and specific general moral principles, ultimately aiming at a state of coherence among these (e.g., Klincewicz, 2019). Other scholars have referred to the *ideal observer theory*, initially introduced by Firth (1952), to argue that IDSS can become versions of ideal observers by exhibiting characteristics such as being "omnipercipient (i.e., it is capable of visualizing, imagining, and using all the information simultaneously), [...] disinterred [or] dispassionate" (Giubilini and Savulescu, 2018; p.171). Lastly, the *moral agency theory* has been highlighted in the articles, elaborating the conditions to be ascribed moral agency and responsibility, such as an actor's intention, control over the outcomes of an event, knowledge about the outcome of their actions as well as the existence of action choices (e.g., van de Voort et al., 2015).

#### *4.1.2. Psychological theories*

In the category of psychological theories, scholars pointed to *rationalist decision-making approaches* (such as Rest's (1986) four component model), which characterizes ethical decision-making as a conscious, deliberative, reason-based, intentional, and individually controlled process (e.g., Ebrahimi and Hassanein, 2021). On the other hand, *non-rationalist approaches* stress emotions, empathy, will and intuition, which lack conscious reasoning, also play an essential role in decision-making (e.g., Giubilini and Savulescu, 2018). One specific example of such irrationality constitutes the *automation bias*, which is the tendency of humans to accept judgments of automated systems as final and superior to their own (e.g., French and Lindsay, 2022). Additionally, by referring to the theory of *situationism*, scholars have argued situational factors

**Table 1**
Overview of identified theoretical underpinnings, illustrating key philosophical and psychological theories/concepts consulted to explain interrelations.

| Category | Theories/concepts | Applied by (alphabetical order) | Particularly relevant for deriving takeaway: |
|---|---|---|---|
| Philosophical theories | • Technological mediation theory | De Boer and Kudina (2021); Susser (2019); van de Voort et al. (2015) | 1 (Dual sources of influence) |
| | • Virtue ethics | Ach and Beck (2023); Cappuccio et al. (2021); Eisikovits and Feldman (2022); Frank (2020); French and Lindsay (2022); Klincewicz (2016); Klincewicz (2019); Lara and Deckers (2020); Liu et al. (2022); Mathieson (2007); Vallor (2013); Vallor (2015); Wong (2019) | 4 (Moral deskilling)<br>9 (Moral progress) |
| | • Reflective equilibrium | Bang et al. (2023); Giubilini and Savulescu (2018); Klincewicz (2016); Klincewicz (2019); Lara and Deckers (2020); Liu et al. (2022); Lara (2021) | 6 (Autonomy enhancement) |
| | • Ideal observer theory | Giubilini and Savulescu (2018); Lara (2021); Liu et al. (2022); Volkman and Gabriels (2023) | 2 (Two methods of influence);<br>3 (Deliberation enhancement) |
| | • Moral agency theory | Boddington (2021); De Cremer and Narayanan (2023); Lara (2021); van de Voort et al. (2015) | 7 (Moral responsibility gap) |
| Psychological theories | • Rationalist ethical decision-making, e.g.:<br>○ Cognitive elaboration model of ethical decision making<br>○ Rest's (1986) model of ethical decision-making | Ebrahimi and Hassanein (2021); Lara and Deckers (2020); Ogunbiyi et al. (2021) | 3 (Deliberation enhancement) |
| | • Non-rationalist ethical decision-making, e.g.:<br>○ Emotionism<br>○ Weakness of the will<br>○ Sensemaking intuition model<br>○ Automation bias | Ach and Beck (2023); Biggar (2023); De Cremer and Narayanan (2023); French and Lindsay (2022); Giubilini and Savulescu (2018); Inthorn et al. (2015); Klincewicz (2019); Krügel et al. (2023); Lara and Deckers (2020); Ogunbiyi et al. (2021); O'Neill et al. (2022); Renic and Schwarz (2023); Savulescu and Maslen (2015); Schwarz (2018); Schwarz (2023); Seville and Field (2000); Shaikh (2020); Vallor (2013); Volkman and Gabriels (2023) | 3 (Deliberation enhancement);<br>5 (Motivation enhancement) |
| | • Situationism, e.g.:<br>○ Construal level theory<br>○ Moral intensity theory<br>○ Motivated cognition | Ebrahimi and Hassanein (2021); Frank (2020); Liu et al. (2022); Ogunbiyi et al. (2021); Straßmann et al., 2020; Wong, 2019 | 5 (Motivation enhancement)<br>6 (Action enhancement) |

such as the proximity or *moral intensity* of an issue are critical determinants for human decision-making and behavior (e.g., Frank, 2020). The *motivated cognition* thesis describes "the tendency to selectively accept or attack incoming information as a function of ideology or worldview compatibility" and in the context of IDSSs, stresses users are more likely to act on IDSSs' recommendations if they appear to align with their own moral intuitions (Liu et al., 2022; p.440).

### 4.2. Types of sources

Sources of influence within IDSSs identified in the respective literature are twofold: they either emerge from the human creators and/or the technological system itself (as illustrated in Table 2). In this sense, the creator refers to the humans (e.g., programmers, data engineers, users) involved in the IDSS development. The technological system here refers to the particular IDSS and its inherent properties and capabilities.

#### 4.2.1. Human creators

The influences that can be ascribed to the creator are errors, biases and preferences that they implement during the development process. For example, De Boer and Kudina (2021) state, "many anticipated issues are conditioned by the affordances and implementation choices embedded within ML [machine learning] systems" (p.257). Similarly, if users can determine the moral criteria to which the system should comply, their potentially biased preferences will be embedded (Klincewicz, 2016). Concerning implicit errors and biases, programmers may *select biased training data* that insufficiently represents individuals from a specific social or cultural background (Cappuccio et al., 2021) or they *misclassify and label particular data* based on biased context information (Ebrahimi and Hassanein, 2021). Concerning the embedding of explicit preferences, developers or organizations with an agenda or partisan interests may design technological features that nudge users' choices and serve companies' *economic imperatives* (O'Neill et al., 2022).

**Table 2**
Overview of identified sources from which IDSSs derive their capacity to influence human ethical decision-making in the first place. These sources (and their corresponding expressions within IDSSs) are twofold in that they either emerge from the human creators and/or the technological system itself.

| Types of sources | Expressions | Exemplary research (alphabetical order) |
|---|---|---|
| Human creators | • Errors/biases/preferences implemented by the humans involved in the system's development, e.g.:<br>○ Subjective selection of training data<br>○ Subjective data labeling<br>○ Embedding of economic imperatives | Benzinger et al. (2023); Cappuccio et al. (2021); De Boer and Kudina (2021); De Cremer and Narayanan (2023); Ebrahimi and Hassanein (2021); Liu et al. (2022); O'Neill et al. (2022); Schwarz (2023) |
| Technological system | • Inherent properties and capabilities of IDSSs, e.g.:<br>○ Autonomy<br>○ Efficiency<br>○ Reliability<br>○ Value neutrality<br>○ Flexibility<br>○ Immersiveness<br>○ Complexity/opaqueness | Ach and Beck (2023); Bang et al. (2023); Berber (2023); Biller-Andorno et al. (2022); De Cremer and Narayanan (2023); French and Lindsay (2022); Krügel et al. (2023); Lara (2021); Lara and Deckers (2020); Liu et al. (2022); Ogunbiyi et al. (2021); Renic and Schwarz (2023); Rodríguez-López and Rueda (2023); Susser (2019); Schwarz (2023); Vallor (2015); van de Voort et al. (2015); Wong (2019) |

#### 4.2.2. Technological system

The influences that emerge from the system relate to the inherent properties and capabilities of IDSSs themselves. Academics in the literature have highlighted the following properties of corresponding systems: autonomy, efficiency, reliability, value neutrality, flexibility, immersiveness and complexity/opaqueness.

*Autonomy* here refers to technology's capability to conduct "autonomous decisions and action without direct human instruction and interference" (Wong, 2019; p.58). It is assumed that a system's autonomy can even increase to the extent that humans are entirely 'out of the loop' (Renic and Schwarz, 2023) so that they cannot control and configure its calculations and operation anymore (van de Voort et al., 2015). It is argued that "there are even workarounds to get ChatGPT to break the rules it is supposed to follow" (Krügel et al., 2023; p.4). Concerning e*fficiency*, scholars have indicated the superiority of technological systems to humans when it comes to the speed of accessing, integrating and analyzing complex data in an aggregated manner (Rodríguez-López and Rueda, 2023). For example, "[a]n artificially intelligent miniaturized drone might be able to record conversations in a multitude of language and dialects, and decode more quickly than a human whether the conversation pattern in a targeted vehicle indicated a friendly or hostile presence" (Vallor, 2015; p.116). Moreover, *reliability* was highlighted in past literature by addressing to what extent systems produce consistent output (when confronted with the same input). While some scholars claimed that artificial moral advisors are consistent in their moral judgment and suggestions (e.g., Giubilini and Savulescu, 2018), other studies disagreed. For example, Bang et al. (2023) found out that, compared to human experts, natural language processing systems provide only reasonably coherent suggestions or answers to ethical quandary questions. Similarly, another study discovered that ChatGPT provides inconsistent and contradictory answers when asked for advice on the same moral issue (Krügel et al., 2023).

Also, the applicability of characterizing IDSSs as *value-neutral* has been discussed ambiguously in past research. On the one hand, systems are seen as disinterested or dispassionate (i.e., not prioritizing/favoring any particular values) as well as unaffected by factors such as emotions, stress, personal biases or fear of legal consequences (e.g., Biller-Andorno et al., 2022). On the other hand, scholars argue that technological systems – as creations of biased human beings – can be considered mirrors that reflect human biases (De Cremer and Narayanan, 2023) and are prone to value-ladenness (French and Lindsay, 2022). Furthermore, academics ascribe *flexibility* to IDSSs as they can learn from and about users' past decisions and cognitive idiosyncrasies to correspondingly adapt by, for example, tailoring user-specific choice architectures (Susser, 2019). Other scholars stress the limited flexibility of systems as their knowledge of how to make 'appropriate' decisions is confined to situations they have been trained on but may not translate to new situations (De Cremer and Narayanan, 2023). Concerning *immersiveness,* scholars have highlighted technologies' capability of offering an environment that allows realistic, active and engaged experiences for users (Lara, 2021). This can be achieved by, for example, incorporating virtual reality (Lara and Deckers, 2020). Lastly, the feature of *complexity/opaqueness* has been mentioned by academics, meaning the limited transparency and comprehensibility regarding the operation and logic by which a system determines its decisions and outcomes (Ogunbiyi et al., 2021).

### 4.3. Methods of influence

The types of methods are specific ways in which IDSSs operate and guide their users throughout the ethical decision-making process. The two types identified in respective literature can be construed as 'process-oriented navigation' or 'outcome-oriented navigation'. The two types and their corresponding operation features are elaborated on in the following section as well as summarized in Table 3.

IDSSs that adopt *process-oriented navigation* follow a procedural and pedagogical approach in that they do not presuppose and strive for a decision outcome that aligns with previously determined ethical principles as fixed benchmarks. Instead, they navigate users through their own ethical decision-making process without aiming to indoctrinate particular ethical principles or values (Lara and Deckers, 2020) or without acting like "an oracle [that] give[s] moral answers based on specific theories" (Bang et al., 2023; p.2). Therefore, the user of the system maintains a high level of participation throughout the ethical decision-making process while the system acts as a supportive instructor (Rodríguez-López and Rueda, 2023). Lara (2021) suggests 'SocrAI' can here serve as an example, which is a system "inspired by the dialectical method adopted by Socrates in his dialogues, which aimed to help his interlocutors to reach definitions of concepts, usually of some virtue, on their own" (p.41). The role of such technological assistants would be to provide objective information (such as predictions or decision alternatives) that helps kick off and structure the moral reasoning process without giving any concrete recommendations on what to do/decide.

IDSSs that adopt *outcome-oriented navigation* follow a directive approach in that they presuppose and target (and sometimes even execute) a specific ethical decision outcome conditioned on previously chosen and embedded principles. This predetermination is conducted on account of the system's users themselves or by the system's programmers. The former subtype will here be referred to as 'participatory', highlighting the user-centric nature, in that users themselves fix the moral criteria that should be followed as well as the moral goals that should be strived for by the systems during the decision-making process (Savulescu and Maslen, 2015). The latter subtype will here be called 'heteronomous', meaning that the involved programmers determine the system's underlying moral values, criteria and aspired goals of the ethical decision-making process (van de Voort et al., 2015). An example of a system that adopts outcome-oriented navigation could be an allocation tool that – based on embedded ethical principles such as 'sickest first' – provides recommendations to healthcare professionals on distributing scarce medical resources between patients. Thus, such systems act as counselors while only providing a low level of participation to their users throughout the decision-making process (Rodríguez-López and Rueda, 2023).

Each navigation type corresponds to a specific operation that manifests in particular informational, analytical or suggestive features adopted within the system. Informational features provide "users with information pertinent to the decision task without suggesting how to act" (Ebrahimi and Hassanein, 2021; p.3). Analytical features examine and report on particular characteristics of the user's decision. Suggestive are those features that make concrete decision recommendations (Ebrahimi and Hassanein, 2021). In the following, we will show how these features express themselves for process-oriented and outcome-oriented navigation.

#### 4.3.1. Features of process-oriented navigation

According to scholars, IDSSs that pursue process-oriented navigation would operate via the following three features.

First, within such systems, *informational features* could be applied. For one, this entails the *transmission of background information* to the user. This includes the supply of general facts about the process of and steps in moral reasoning (without advising specific actions in a particular case) (O'Neill et al., 2022) or the enumeration and explanation of any moral principles or theories (Seville and Field, 2000). Furthermore, such technologies can scan the environment and provide morally relevant information about the user's surroundings (Klincewicz, 2016) or reconstruct how/why the (dilemma) situation occurred in the first place (Tassella et al., 2023). For example, in the military sector, IDSSs "provide soldiers with enhanced information about morally salient features of the battlefield for use in their deliberation", such as the presence of civilians (Vallor, 2015; p.116). In addition, relevant information to be transmitted could constitute data on how others have decided in similar

**Table 3**
Overview of identified methods by which IDSSs influence human ethical decision-making. The two identified methods (i.e., 'process-oriented navigation' or 'outcome-oriented navigation') manifest in different technological features and the way they navigate users through the decision-making process.

| Types of methods | Description | Operation (i.e., corresponding features) | Exemplary research (alphabetical order) |
|---|---|---|---|
| Process-oriented navigation | The system navigates the user through the ethical *decision-making process* without drawing on fixed ethical benchmarks. | • Informational features<br>○ Transmission of background information<br>○ Indication of personal constraints<br>○ Overview of alternative options<br>○ Provisions of predictions<br>• Analytical features<br>○ Questioning of decision justification<br>○ Provision of feedback on an empirical basis, logical & ethical rigor of decision<br>○ Foreshadowing a decision's potential implications<br>• Suggestive features<br>○ Advice on implementation strategies for self-determined decision | Bang et al. (2023); Benzinger et al. (2023); Biller-Andorno et al. (2022); Cappuccio et al. (2021); Eisikovits and Feldman (2022); French and Lindsay (2022); Klincewicz (2016); Lara (2021); Lara and Deckers (2020); Mathieson (2007); O'Neill et al. (2022); Schwarz (2023); Seville and Field (2000); Shaikh (2020); Tassella et al. (2023); Vallor (2015); Yang et al. (2004) |
| Outcome-oriented navigation<br>- Participatory<br>- Heteronomous | The system navigates the user toward a *decision outcome* that is conditioned on predetermined ethical benchmarks (embedded by the system's user and/or programmer). | • Informational features<br>○ Reminding on user's duties and moral value of a particular activity<br>• Analytical features<br>○ Provision of feedback concerning consistency with user's ethical principles<br>○ Provision of feedback concerning consistency with 'absolute' ethical principles<br>• Suggestive features<br>○ Decision recommendation based on the user's ethical principles<br>○ Decision recommendation based on 'absolute' ethical principles | Ach and Beck (2023); Frank (2020); French and Lindsay (2022); Giubilini and Savulescu (2018); Klincewicz (2016); O'Neill et al. (2022); Savulescu and Maslen (2015); Tassella et al. (2023); Vallor (2013); van de Voort et al. (2015) |

situations in the past (Biller-Andorno et al., 2022). This highlights the possibility of offering integrated knowledge to users through technological systems. For example, decision-support technology in the healthcare sector can gather and transfer multi-stakeholder information such as patients' demographics or past treatments, nurses' or doctors' medical reports (Yang et al., 2004) as well as patients' preferences (Benzinger et al., 2023). Second, informational features can be adopted that *indicate personal constraints* to the user. For example, Lara (2021) proposes a virtual assistant monitoring users' physiology or mental states to generate alerts if factors (such as fatigue or stress) that could negatively affect their decision-making process are detected. Furthermore, corresponding features constitute presenting an *overview of alternative options* in combination with arguments for and against each alternative or viewed from different perspectives (Bang et al., 2023). Another informational feature involves the *provision of predictions*, for example, medical diagnoses in life-or-death situations (Shaikh, 2020).

Second, *analytical features* that were stated in this regard are the following. IDSSs can *question the justification for a decision*. For example, Lara and Deckers (2020) suggest that a corresponding system would process the entered information (i.e., decision) and engage in a conversation with the user to inquire about the provision of reasons for a particular choice. Similarly, thought-provoking questions such as "Would you like to see your decision reported in the newspaper?" (Mathieson, 2007; p.6) or "Are you sure we should be doing this?" (French and Lindsay, 2022; p.35) can be displayed to the user via the system. Another analytical feature constitutes the p*rovision of feedback on the empirical basis, logical and ethical rigor of a decision*, thereby challenging the user's assumptions. For example, IDSSs can analyze and indicate to the users to what extent their moral judgments are empirically refuted premises (Lara and Deckers, 2020). Similarly, by formalizing the user's arguments, logical fallacies and decision invalidity can be made transparent for the user (French and Lindsay, 2022). This is especially feasible for technological systems if they can continuously access and recognize user's decision-making patterns and inconsistencies (Seville and Field, 2000). Moreover, analytical features can be adopted that *foreshadow a decision's potential implications* by evaluating a user's decision action or inaction given resulting positive or negative consequences (Cappuccio et al., 2021). For example, "in Honda hybrid cars, a device called the eco-assist provides feedback on the driver's energy usage based on their driving style" (Frank, 2020; p.372).

Lastly, systems that pursue process-oriented navigation entail limited *suggestive features.* Namely, the only corresponding feature identified in past literature was the *advice on implementation strategies for self-determined decisions.* Here, the IDSS could inform the agent on "how to put into practice the morel decisions that they have [independently] reached" (Lara and Deckers, 2020; p.284).

#### *4.3.2. Features of outcome-oriented navigation*

According to scholars, IDSSs that adopt outcome-oriented navigation can operate via the following features.

First, only a limited number of *informational features* were identified by past literature in connection with this navigation type. These include sending *reminders of the user's duties and function*, for example, prompting military members to embrace rules such as the Uniform Code of Military Justice (French and Lindsay, 2022). Similarly, *reminders of the moral value of a particular activity* (such as conducting a daily act of kindness) can be featured by corresponding IDSSs (Frank, 2020).

Second, stated *analytical features* in this regard constitute the provision of feedback concerning a decision's consistency either with the user's own ethical principles or with externally determined, 'absolute' ethical principles. The *provision of feedback concerning a decision's consistency with the user's ethical principles* is more likely to emerge in the 'participatory' subtype, in that evaluations are conducted and reported, which highlight to what extent decisions correspond to the principles and values the user initially identified and entered as relevant (O'Neill et al., 2022). For example, Tassella et al. (2023) state such systems could warn users if anticipated decision actions go against their previously indicated values. On the other hand, the *provision of feedback concerning a decision's consistency with an 'absolute' value* is rather characteristic of the 'heteronomous' subtype, in which particular theories, principles or values are considered universally valid by the designer and are thus, integrated as a baseline guiding the ethical decision-making process. The 'absolute' theories that a user's decision is evaluated against could entail normative theories such as Kantian deontological theory or utilitarianism (Klincewicz, 2016) or sector-specific norms such as honor or courage in the military (Vallor, 2013).

Third, the *suggestive features* of systems that pursue outcome-oriented navigation operate similarly to their analytical features. On the one hand, a corresponding feature could encompass a *decision recommendation based on the user's own ethical principles*. For example, a corresponding IDSS could provide the user "with a range of options, signaling the one, which more closely complies with the [user's] moral standards" (Giubilini and Savulescu, 2018; p.175). On the other hand, a system could construct a *decision recommendation based on 'absolute' ethical principles* such as the previously stated normative theories (Mathieson, 2007). On this basis, such technological systems can give concrete decision/action suggestions or by, for example, providing "answers to first-order normative questions, such as 'should I report this to the authorities?' with a definite 'yes' or 'no'" (Klincewicz, 2016; p.179).

### 4.4. Direct outcomes

Based on past literature, four different immediate outcomes on individuals were identified that result from the use of particular technologies, namely: deliberation enhancement, motivation enhancement, autonomy enhancement as well as action enhancement (as illustrated in Table 4). All of these outcomes are illuminated in the following sections.

#### 4.4.1. Deliberation enhancement

Deliberation enhancement is here apprehended as the increased ability to engage in deliberative, reason-based ethical decision-making processes. More specifically, the outcomes identified in past literature in this regard are the users' recognition of a multitude of integrated data, awareness of the moral dimension of a particular situation, the realization of personal fallacies as well as their detection of a greater solution space.

The first outcome is the user's *recognition of a multitude of integrated data*, which allows informed decision-making. For example, an epistemic increase may arise due to the user's identification and access to clear, high-caliber and consolidated information (Manders-Huits, 2006), for instance, through a graphical user interface (Schwarz, 2023). In this line, De Boer and Kudina (2021) state that using machine learning in medical practices positively affects the observation capacities of medical professionals in diagnosis processes, which in turn unlocks the "potential to skyrocket the availability and quality of individualized medicine" (Benzinger et al., 2023; p.2). Next to information about the situation at hand, through IDSSs, users can additionally recognize data relating to the fundamentals of ethical decision-making. For example, users may generate knowledge of various normative theories and principles (Lara and Deckers, 2020) or identify what constitutes logical reasons (Lara, 2021).

Second, past literature has raised the outcome of users' *awareness concerning the moral dimension of a situation*. In that some IDSSs point toward morally relevant facts within the environment, individuals may be more mindful of what is taking place in the present (Ebrahimi and Hassanein, 2021). Following, individuals "suffer from less cognitive

**Table 4**
Overview of identified direct and secondary outcomes that can result from IDSSs. While the direct outcomes cover the immediate influences on ethical decision-making on an individual level, secondary outcomes refer to indirect and long-term repercussions for ethical decision-making on a societal level.

| Types of outcomes | | Description | Expressions | Exemplary research (alphabetical order) |
|---|---|---|---|---|
| Direct | Deliberation enhancement | *Ability* to engage in *deliberative, reason-based ethical decision-making* processes | • Recognition of a multitude of integrated data<br>• Awareness concerning the moral dimension of situations<br>• Realization of personal fallacies<br>• Detection of greater solution space | Ach and Beck (2023); Bang et al. (2023); Benzinger et al. (2023); Biggar (2023); Biller-Andorno et al. (2022); Cappuccio et al. (2021); De Boer and Kudina (2021); Ebrahimi and Hassanein (2021); Frank (2020); French and Lindsay (2022); Inthorn et al. (2015); Lara (2021); Lara and Deckers (2020); Liu et al. (2022); Manders-Huits (2006); Schwarz (2023); Vallor (2013); van der Waa et al. (2020) |
| | Motivation enhancement | *Willingness* to *execute established (ethical) decision intention* | • Perception of decision proximity<br>• Acknowledgment of decision ownership | Biggar (2023); De Boer and Kudina (2021); De Cremer and Narayanan (2023); Ebrahimi and Hassanein (2021); Lara (2021); Lara and Deckers (2020); Renic and Schwarz (2023); Vallor (2015); van de Voort et al. (2015) |
| | Autonomy enhancement | *Ability* to *decide in accordance with one's own ethical principles* | • Achievement of decision that aligns with the user's ethical principles<br>• Achievement of decision that aligns with external ethical principles | Ach and Beck (2023); Bang et al. (2023); Berber (2023); De Boer and Kudina (2021); Lara (2021); Lara and Deckers (2020); Liu et al. (2022); O'Neill et al. (2022); Susser (2019); Yang et al. (2004) |
| | Action enhancement | *Execution* of *ethical decision* | • Change toward an ethical decision outcome<br>• Change toward an unethical decision outcome<br>• No change in ethical decision outcome at all | Berber (2023); Biller-Andorno et al. (2022); Cappuccio et al. (2021); De Cremer and Narayanan (2023); Ebrahimi and Hassanein (2021); Frank (2020); Giubilini and Savulescu (2018); Klincewicz (2016); Lara (2021); Liu et al. (2022); Ogunbiyi et al. (2021); Renic and Schwarz (2023); Schwarz (2018); van der Waa et al. (2020) |
| Secondary | Moral deskilling | *Loss of skill* at *making ethical decisions* due to lack of experience and practice | • Sufficient practice due to cultivating ethical decision-making<br>• Insufficient practice due to narrowing and outsourcing ethical decision-making | Ach and Beck (2023); Biggar (2023); De Cremer and Narayanan (2023); Eisikovits and Feldman (2022); Erler and Müller (2021); Green (2019); Lara (2021); Lara and Deckers (2020); Liu et al. (2022); Renic and Schwarz (2023); Schwarz (2023); Vallor (2013); Volkman and Gabriels (2023) |
| | Restricted moral progress | *Limited* discovery and application of *new values* | • Static account of morality, limited to embedded ethical principles<br>• Loss of personal skills necessary for moral innovation | Frank (2020); French and Lindsay (2022); Lara and Deckers (2020); Manders-Huits (2006); O'Neill et al. (2022); Renic and Schwarz (2023); Schwarz (2018); Schwarz (2023); Volkman and Gabriels (2023) |
| | Moral responsibility gap | Vacuum, in which decision outcomes are *no one's moral responsibility* | • Human users cannot be held responsible for decision outcomes<br>• Technological systems cannot be held responsible for decision outcomes | Bang et al. (2023); Berber (2023); Boddington (2021); De Boer and Kudina (2021); French and Lindsay (2022); O'Neill et al. (2022); Renic and Schwarz (2023); Schwarz (2018); van de Voort et al. (2015); van der Waa et al. (2020) |

overload as the need for sufficient situational understanding is reduced [to those cues that are morally relevant]" (van der Waa et al., 2020; p.212). Ebrahimi and Hassanein (2021) show corresponding empirical evidence that highlights providing morally relevant information (i.e., demographics of data subjects) during the use of a data analytics tool enhances users' moral recognition.

Third, technologies can assist users in *realizing their own personal fallacies*. Scholars have stated that IDSSs "effectively help humans pay attention to their shortcomings […], supporting the reflective consciousness and self-control" (Cappuccio et al., 2021; p.18). For example, users could identify and subsequently overcome cognitive and affective limitations and biases (Frank, 2020) and be invited to engage in self-critical reflections (Biller-Andorno et al., 2022). On the other hand, scholars in the respective literature also suggest the exact opposite, namely, that particular technological systems disguise or deliberately appeal to users' personal fallacies. For example, individuals may be forced to refrain from extensive ethical deliberation but decide based on heuristics due to a limited time horizon, in which a user can veto a system's decision suggestion (Vallor, 2013).

Fourth, another outcome constitutes the user's *detection of a greater solution space*. Namely, by being assisted with particular decision-support tools, users can recognize "solutions that are not limited by a restricted frame of mind" (Inthorn et al., 2015; p.183) but are broadened with multiple perspectives (Bang et al., 2023). Instead of believing to be in binary, 'either/or' dilemmas, IDSSs show users alternative decision/action options (French and Lindsay, 2022).

#### 4.4.2. Motivation enhancement

Motivation enhancement is here conceptualized as the increased willingness to execute one's established (ethical) decision intention. Two pertinent outcomes emerging from the use of particular IDSSs were identified in past literature: the users' perception of the decision proximity and their acknowledgment of decision ownership.

For one, users' *perception of decision proximity* may arise from technological systems if the proximity to affected data subjects is emphasized (van de Voort et al., 2015). Again, referring back to the study by Ebrahimi and Hassanein (2021), it can be witnessed that by increasing users' perceived proximity to the subjects of their decision (by providing subjects' demographics or pictures), their moral motivation rises. Such closeness and cognitive empathy can be accelerated with the deployment of virtual reality by "allowing us to 'experience' the realities of (particularly) distant others more vividly, and to imagine much better how our actions and omissions affect them" (Lara and Deckers, 2020; p.285). Other technologies, however, are said to (geographically) distance human decision-makers from their decision implications and potentially affected parties, such as a soldier relying on a drone to decide whether to strike and kill (Vallor, 2015). Similarly, a non-provocative (i.e., non-anthropomorphic) design of the decision-making tools may undermine the user's elicitation of emotions and thus, motivation (Lara, 2021). Biggar (2023), on the other hand, argues that "being safely removed from the theater of operations" (p.72) allows users (e.g., pilots of autonomous weapons) to be less affected by emotions such as pain, fear of anger and instead, access their practical judgment.

Moreover, certain IDSSs and their corresponding features are said to result in the user's *acknowledgment of ownership for a decision.* For example, when engaging in a 'neutral' dialogue with a virtual assistant, individuals may perceive the resulting decision recommendation as their own, motivating them and making them proud to follow through with this decision (Lara, 2021). However, under certain conditions, external artifacts such as AI advisors can become part of an individual's mind so the device's reasoning processes are to be equated with the individual's processes (Erler and Müller, 2021). Indeed, Krügel et al. (2023) found out that even users of ChatGPT tend to proclaim moral stances that the technology has suggested as their own. Other scholars argue that the use of technology in (ethical) decision-making processes changes the way humans (i.e., the users) understand themselves and their obligations (De Boer and Kudina, 2021). More specifically, individuals may develop less 'ethical' intentions because of acting behind the veil of complex technology (Green, 2019) or feel morally disengaged as they "perceive machines – rather than themselves – as being in charge of driving decision-making" (De Cremer and Narayanan, 2023; p.6), thus renouncing decision ownership.

#### 4.4.3. Autonomy enhancement

Autonomy enhancement here pertains to the improved ability to decide in accordance with one's own ethical principles. Scholars within past literature have stated that two different outcomes can emerge from IDSSs in this regard: either the achievement of a user's decision that aligns with their own ethical principles or the achievement of a user's decision that aligns with external ethical principles.

Concerning the outcome of users' *achievement of decisions that align with their own ethical principles*, Lara (2021) sketched a virtual assistant "that not only respects but also increases moral autonomy" (p.42). In particular, through dialogue, this assistant instructs users so that they can follow the criteria of deliberative rigor and contemplate their own ethical principles (Lara, 2021). A similar result was suggested by Yang et al. (2004), highlighting the value of automated decision tools to help clients clarify their value systems and eventually make corresponding decisions. Such technologies pursue an "individual value alignment function" (O'Neill et al., 2022; p.23) and can be considered as an "efficient assistant, a mid-wife who helps users to give birth to a decision that is completely theirs" (Lara and Deckers, 2020; p.285).

By contrast, other scholars have highlighted that certain IDSSs result in users' *achievement of decisions that align with external ethical principles*, thus threatening personal autonomy (Susser, 2019). Particularly, technological systems that implement nudges aim to trigger users' heuristics instead of improving users' attention to their deeply held values (Lara, 2021). Such an "invisible influence threatens th[e] ideal [of personal autonomy] by inducing people to act for reasons they don't understand, and therefore can't endorse" (Susser, 2019; p.406). A similar outcome emerges when medical professionals adopt diagnosis and treatment advice from complex machine learning technologies, in that they cannot comprehend and articulate the underlying reasoning process (De Boer and Kudina, 2021). The eventual implication for humans' autonomy would entail blind, inauthentic acceptance of external values and recommendations without any reflection to what extent they personally identify with them (Lara, 2021).

#### 4.4.4. Action enhancement

Action enhancement here means the final execution of an ethical decision. As identified in this review, contemporary scholars pictured three different corresponding outcomes that can emerge when using particular decision-support technologies, these are: users' decision change toward an ethical outcome, users' decision change toward an unethical outcome or no resulting change in the ethical decision outcome at all.

On the one hand, past research highlights that IDSSs can generate humans' *change toward an ethical decision outcome*. For example, they can allow users to be intrinsically moral (van der Waa et al., 2020), change the decision context so that it is easier for individuals to act morally (Klincewicz, 2016) or "make the unacceptable unpalatable" (Lara, 2021; p.42). Especially, more sophisticated, persuasive technologies that offer dynamic choice architectures are ascribed with the ability to eventually "engineer the environment for maximum moral behavior" (Frank, 2020; p.372). Cappuccio et al. (2021) introduce the example of a robot tortoise, Shelly, which taught children to refrain from violent and abusive actions. On an extreme level, using such technologies that take over most of the humans' ethical decision-making process could entail the realization of optimal moral choices on our behalf (Giubilini and Savulescu, 2018).

On the other hand, scholars criticize that IDSSs may contribute to individuals' *change toward an unethical decision outcome*. For one, "a poorly designed system can reinforce current flaws in decision-making and introduce new ones" (Biller-Andorno et al., 2022; p.10). Therefore, through the use of AI, individuals may unintentionally engage in unethical actions (Ogunbiyi et al., 2021). For instance, even well-intentioned decision-makers may conduct discriminatory decisions when advised by a tool that is based on training data holding an embedded (racial) bias (Ebrahimi and Hassanein, 2021). Even more so, systems may spur unethical decision outcomes when used by malevolent individuals who see a potential to engage in unethical decisions without being recognized (Schwarz, 2018) or when they act in the form of a cheerleader reaffirming their own (unethical) belief systems (Liu et al., 2022).

With a similar skepticism, past literature sketches the possibility that decision-making tools may bring *no change in ethical decision outcomes at all*. This is especially the case for values and ethical principles that individuals deeply hold: either these values would be implemented as a baseline for the technologies' decision support and thus encourage the user's biased decisions or these values would be set in stone even when presented counterarguments or opposing values (Klincewicz, 2016). As Volkman and Gabriels (2023) state, "[m]erely making agents more consistent and thoughtful is not sufficient to produce moral enhancement, since users might input the wrong values from the start and become more consistently wrong as a result" (p.10).

### 4.5. Secondary outcomes

Past literature has indicated that relying on IDSSs for ethical decisions may not only generate immediate influences for the user but can also imply indirect and long-term consequences for the ethical decisions of individuals and society as a whole (as illustrated in Table 4). More specifically, these are moral deskilling, restricted moral progress and the emergence of moral responsibility gaps.

#### 4.5.1. Moral deskilling

Moral deskilling is "the loss of skill at making moral decisions due to lack of experience and practice" (Green, 2019; p.2), which has been associated with delegating ever more ethical decisions to technological systems. On the one hand, IDSSs may help to *cultivate practical wisdom and habituation in making ethical* decisions (Volkman and Gabriels, 2023) and thereby work contrary to moral deskilling. In particular, it is argued that, with their dialectical training, such systems help users exercise their deliberative capacities so that, over time, they can make ethical decisions self-sufficiently (Lara, 2021). On the other hand, it is indicated that "if a moral AI advisor did provide us with sound ethical advice, it would […] fail to authentically enhance our capacity for moral reasoning […] due to insufficient practice" (Erler and Müller, 2021; p.10). Namely, as *ethical decisions are outsourced* to IDSSs, which suggest concrete decisions, individuals may unlearn being critical about the processes by which these decision suggestions are produced (Lara and Deckers, 2020). "This is sometimes referred to as 'ethical muscle memory' representing the idea that it is easier to do the right thing if you have engaged with a situation in advance and are already familiar with the ethical landscape of the problem" (Schwarz, 2023). Similarly, it was indicated that relying on technological systems during an ethical decision-making process narrows the skill set required for humans to simply supervise, approve or veto suggested decisions. Cultivating moral skills only in a limited number of aspects will imply that humans eventually cannot make qualified moral judgments on their own (Vallor, 2013). Compared to its technological counterparts, society as a whole will increasingly adopt a passive role in ethical decision-making processes, blindly following technologies' instructions that are delivered on a plate. This entails a form of regression, "which is a version of the general 'autopilot problem'" (Erler and Müller, 2021; p.10). In line with this ambiguous debate concerning IDSSs' influence on moral deskilling, Biggar (2023) argues that IDSSs will not universally deskill their users. Instead, particular virtues (such as physical courage among soldiers) may decrease, but this loss could be offset by the cultivation of other virtues (such as prudence) (Biggar, 2023).

#### 4.5.2. Restricted moral progress

Moral progress refers to "the discovery and application of new values or sensitization to new sources of harm" (Frank, 2020; p.374). Past literature stated relying on IDSSs for ethical decisions can lead to restricted moral progress. The underlying assumption is that there must be moral pluralism and dissent to allow moral progress (Lara and Deckers, 2020). However, when relying on technologies during decision-making processes, humans may become less aware of alternatives compared to those offered to them by the system (Manders-Huits, 2006), "ultimately reducing morality to the output of some algorithm" (Volkman and Gabriels, 2023; p.3). These alternatives prioritize those moral aspects that can be quantified and turned into code while neglecting those that exceed analytical cost-benefit analysis. Because of this technological solutionism, humans' ethical decision-making is limited to 'ethics-as-science' (Schwarz, 2018), which, in turn, incentivizes the dehumanization of affected individuals (such as the targets of autonomous weapons) (Renic and Schwarz, 2023). Furthermore, using IDSSs for ethical decisions may jeopardize moral progress, as the systems are compliant and set on existing norms and laws (Volkman and Gabriels, 2023). This way, "we would be left with a *static account of morality*", although it can be expected that "[m]oral judgments that are made today may no longer be acceptable at some other time" (Lara and Deckers, 2020; p.279). This restricted perspective may increase with the time that IDSSs are being deployed. For example, it is assumed that "the longer that autonomous systems are used in decision-making capacities […], the more likely it becomes that the humans will feel less sure of themselves when it comes to questioning the autonomous system and challenging its authority" (French and Lindsay, 2022; p. 61). Thus, also on an individual level, deploying IDSSs for ethical decisions can constrict users' moral growth (O'Neill et al., 2022) as they no longer have the cognitive and affective *skills necessary for moral innovation* (Frank, 2020).

#### 4.5.3. Moral responsibility gap

To be held morally responsible, an actor must have control over the outcomes of an event and must be knowledgeable about action alternatives and the expected outcomes of particular actions (van de Voort et al., 2015). Using IDSSs drives unclarity into who is responsible for a particular decision outcome or output that the system offers (Bang et al., 2023) since neither the system nor the human user fulfills the conditions necessary for being morally responsible. For example, IDSSs may bias users so that the adopted values are not actually intended by the users (van der Waa et al., 2020), *limiting their control over and moral agency* for resulting decisions (Boddington, 2021). Furthermore, users might be unable to comprehend or retrace the underlying decision logic of technological systems due to algorithmic opacity, which outlaws them "[t]o serve as responsible epistemic agents" (De Boer and Kudina, 2021; p.252). Any human connected to a certain IDSS "(e.g., engineers, legislators, users […]) would morally distance themselves from the decisions made by the [system]" (Berber, 2023; p.3). Similarly, *technological systems "might not be the sort of thing that can be punished or* *adequately held to account* for its claims and recommendations" (O'Neill et al., 2022; p.31). For one, this is due to the inability to verify the accuracy of a technologically determined decision (Schwarz, 2018). What results from this is that no one can truly be held accountable for decision outcomes (French and Lindsay, 2022). Instead, a moral vacuum or responsibility gap is created (Schwarz, 2018).

## 5. Discussion: an integrated framework & implications

This discussion section will summarize and connect the previously sketched findings by providing an integrated framework and proposing interrelations (see 5.1. Proposed interrelations). This way, we can answer research questions 1 and 2 by showing by what concrete mechanisms (i.e., sources and methods of influence) IDSSs shape humans' ethical decision-making in which particular way (i.e., direct and indirect outcomes). Moreover, theoretical and practical implications are outlined to illustrate how this study extends previous scholarly work and what potential governance measures for companies that develop pertinent IDSSs are (see 5.2. Theoretical and practical implications).

### *5.1. Proposed interrelations*

As visualized in Fig. 3 and detailed in this section, nine takeaways concerning potential interrelations between the identified sources, operation types and outcomes can be drawn.

As a starting point, influences commence within IDSSs due to two sources: they originate from the system's creators (i.e., its human developer, programmer and user) and/or from the technological system itself. Namely, the creator implicitly or explicitly shapes what features (i.e., informational, analytical or suggestive) are implemented into an IDSS and what the nature of these features is (e.g., neutral or value-laden). Thus, in the development process, the system's creators frame whether the IDSS will favor process-oriented or outcome-oriented navigation. In addition, inherent properties of the system, such as its complexity or efficiency, undermine or reinforce to what extent these embedded features can effectively influence human ethical decision-making. For example, a high degree of an IDSS's complexity/opaqueness impedes users from actively participating in the deliberation process. This may manifest as the difficulty of reconstructing the rationale behind a system's recommendations. Instead, individuals are compelled "to trust in the [IDSS's] cognitive and rational superiority" and adopt the suggested recommendations, aligning with the phenomenon of automation bias (Renic and Schwarz, 2023). To give another example, a system's sophistication in efficiently and reliably analyzing the user's environment will impact to what extent morally relevant factors in a situation can be recognized and communicated to the user. These illustrations show that neither the IDSS's creators nor the system independently determine the eventual influence on the users' ethical decision-making; instead, they collaboratively shape this influence. This dual and reciprocal influence aligns with the technological mediation theory. Therefore, regarding IDSSs, the first takeaway can be postulated as follows:

Takeaway 1 – Dual sources of influence:

- Influences of IDSSs on humans' ethical decision-making can originate from their human creators as well as the technological system itself.
- To what extent the influences of the human creators and the system manifest within the IDSS depend on their reciprocal impact on each other.

Second, the manner in which influences materialize within IDSSs can be broadly separated into two different types/methods. These differ in their operation (i.e., process-oriented navigation or outcome-oriented navigation), which manifests in corresponding informational, analytical and suggestive features. Compared to process-oriented navigation, the features of outcome-oriented navigation are less neutral in that they are charged with the ethical principles of the system's users (i.e., participatory) and/or the system's programmers (i.e., heteronomous). For example, transmitting mere, objective facts such as past medical reports represents an informational feature of process-oriented

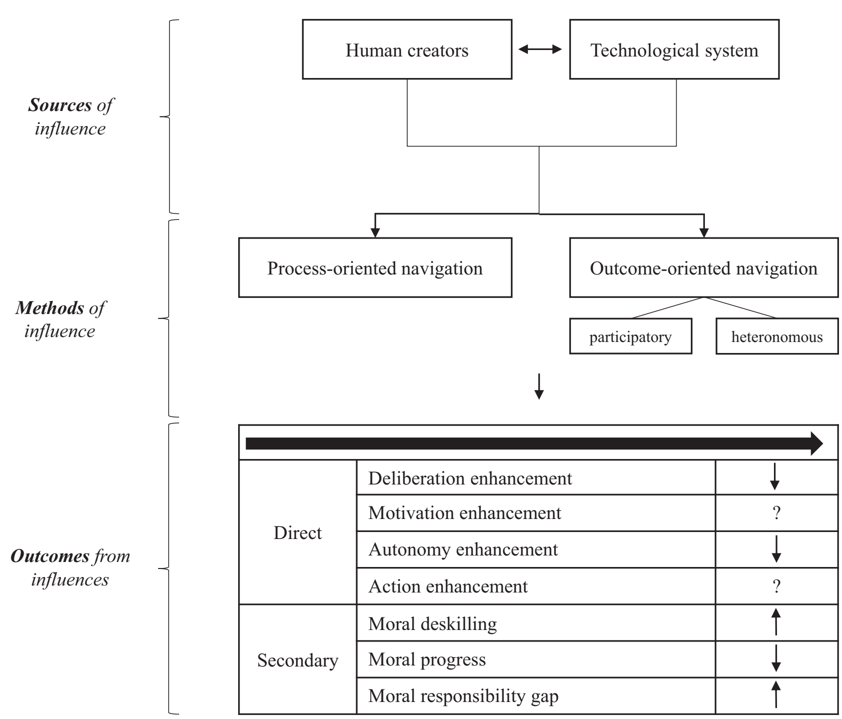


**Fig. 3.** Proposed integrated framework of the impact of IDSSs on humans' ethical decision-making, displaying potential interrelations between the sources, methods of influence and the resulting direct and secondary outcomes.

navigation. These types of IDSSs rather fall into the category of an ideal observer (Lara, 2021). By contrast, an exemplary informational feature of outcome-oriented navigation would entail reminding users of their duties, deduced from value-laden data (e.g., the Uniform Code of Military Justice) that was previously embedded into the IDSS as a reference point. Therefore, IDSSs that adopt outcome-oriented navigation direct the user toward a predetermined decision outcome. This decision outcome is predetermined in the sense that the systems draw on embedded reference points (such as particular ethical principles that are selected before usage) to guide their operation. On the other hand, IDSSs that adopt process-oriented navigation follow a procedural and pedagogical approach by guiding users through their ethical decision-making process without aiming to indoctrinate particular predetermined values. The user of the system remains active throughout the decision-making process. Overall, the following takeaway can be established:

Takeaway 2 – Two methods of influence:

- IDSSs can influence humans' ethical decision-making through process-oriented navigation or outcome-oriented navigation.
- While process-oriented navigation is characterized by neutral informational, analytical and suggestive features, the features of outcome-oriented navigation are rather directive and value-laden as specified by the IDSS's user (i.e., participatory) and/or programmer (i.e., heteronomous).
- As a result, process-oriented navigation aims to guide users through their ethical decision-making process, while outcome-oriented navigation aims to direct users toward a predetermined ethical decision outcome.

These influences ultimately alter the decision-making process and morality of individual users and society as a whole. For one, deliberation enhancement (i.e., the improved ability to engage in deliberative, reason-based ethical decision-making processes) can be attributed more to IDSSs that pursue process-oriented navigation rather than outcome-oriented navigation. The argument here is that through the multitude of information and analytical features that process-oriented navigation entails, individuals are facilitated to engage in a moral reasoning/deliberation process. For one, this can be explained by referring back to rationalist models of ethical decision-making, which declare moral awareness as the required outset to kick off the ethical decision-making process in the first place (e.g., Lewis, 1989; Rest, 1986). In line with this, one feature of process-oriented navigation is transmitting morally relevant information, which increases the user's awareness that the situation at hand carries a moral dimension, which in turn initiates the reasoning process. Furthermore, through analytical features such as challenging the justification or ethical rigor of a user's decision, users are more likely to realize if their intuitions or decision heuristics are biased. This way, with IDSSs that follow process-oriented navigation, "reason may not be completely enslaved to our passions" (Volkman and Gabriels, 2023; p.4) as it is typically assumed according to intuitionist models of ethical decision-making (e.g., Haidt, 2001). By contrast, IDSSs that strive for outcome-oriented navigation do not help the user learn to reason (ethically) in this way but force users to refrain from extensive ethical deliberation (Lara and Deckers, 2020; Vallor, 2013).

This, in turn, spurs the emergence of moral deskilling. As a reminder, referring back to virtue ethics, repeated moral practice (for example, in moral reasoning) is necessary to excel at making ethical decisions in the long run (Vallor, 2013). Eisikovits and Feldman (2022) even argue that this moral deskilling may not only concern the user's ethical decision-making in the specific (professional) context for which technological systems are used but spill over to, for example, the ability to make ethical decisions in private life. Since – in contrast to process-oriented navigation – outcome-oriented navigation undermines the habituation of engaging in deliberation and moral reasoning, individuals will lose the ability to make ethical decisions on their own. Therefore, the following takeaways can be formulated:

Takeaway 3 – Deliberation enhancement:

- An IDSS's contribution to humans' deliberation enhancement decreases when moving from process-oriented to outcome-oriented navigation.

Takeaway 4 – Moral deskilling:

- An IDSS's contribution to humans' moral deskilling increases when moving from process-oriented to outcome-oriented navigation.

For motivation enhancement (i.e., higher willingness to execute the established decision intention), it is unclear to what extent its occurrence differs between the two types of methods (i.e., process-oriented to outcome-oriented navigation). For example, it has been argued that users' corresponding motivation is likely to rise when they perceive high levels of participation and, consequently, ownership of a particular decision (Lara, 2021), which, in principle, can be mapped more so to IDSSs that adopt process-oriented navigation. However, one could similarly expect that even with IDSSs that exercise outcome-navigation, users may perceive high levels of decision ownership, which has been supported by past studies (Krügel et al., 2023). More importantly, – independent of the particular navigation type employed by the designer – motivation enhancement seems to be shaped by the systems' properties, such as immersiveness, which appeal to individuals' emotional states. This aligns with intuitionist models of ethical decision-making (e.g., Haidt, 2001), which emphasize the significant impact of affects and empathy in the decision process. Similarly, the moral intensity theory (Jones, 1991) can be consulted to argue why immersiveness may lead to motivation enhancement in that such properties create proximity to the effects (e.g., through virtual reality applications) (Lara and Deckers, 2020), resulting in higher levels of motivation to decide (in a particular way). Overall, the following can be assumed:

Takeaway 5 – Motivation enhancement:

- An IDSS's contribution to humans' motivation enhancement is less dependent on its navigation type but rather determined by its inherent properties (e.g., immersiveness) that appeal to the users' emotions.

Concerning autonomy enhancement (i.e., the improved ability to decide in accordance with one's own ethical principles), it can be assumed that process-oriented navigation holds positive impacts for the user, while outcome-oriented navigation rather contributes negatively to a user's autonomy. For example, due to its "axiological neutrality" (Rodríguez-López and Rueda, 2023; p.6) and through its analytical features that request the user to self-reflect, process-oriented navigation can help users clarify their value systems and achieve reflective equilibrium by which their ultimate decisions are consistent with deeply held values (Klincewicz, 2019). On the other hand, IDSSs pursuing outcome-oriented navigation rather push or nudge users toward adopting values and decisions that are externally predetermined, usually advocating a particular ethical position (Volkman and Gabriels, 2023). Even if the resulting decisions were less 'evil', they would emerge at the expense of the users' free will (Berber, 2023). Those decisions that derive from 'participatory' IDSSs also limit a user's autonomy since they are advised on "transient and noisy snapshot of [own] values" (Liu et al., 2022; p.441). What results is blind and unreflected acceptance of the recommendations that the IDSS provides. This limited decision autonomy creates moral responsibility gaps because the user can no longer be considered the complete morally responsible agent. This gap intensifies the more passive the user becomes in the ethical decision-making process. In other words, "the less human oversight and the greater decision-making influence AI systems have, the greater the problem of legal and ethical responsibility grows" (French and Lindsay, 2022), which is the case when moving from process-oriented to outcome-oriented navigation. Thus, we formulate the following takeaways:

Takeaway 6 – Autonomy enhancement:

- An IDSS's contribution to humans' autonomy enhancement decreases when moving from process-oriented to outcome-oriented navigation.

Takeaway 7 – Moral responsibility gap:

- An IDSS's contribution to the moral responsibility gap increases when moving from process-oriented to outcome-oriented navigation.

Lastly, action enhancement (i.e., whether an ethical decision is eventually executed) cannot distinctively be mapped to a particular navigation type but may (not) transpire from both. For one, scholars argue that through features that are attributed to process-oriented navigation, individuals are, in principle, provided with the fundamentals (e.g., deliberative capabilities) enabling them to arrive at an ethical decision. However, these do not necessarily entail that an ethical decision indeed results, for example, when the user's intentions are malicious. Similarly, IDSSs that pursue ('participatory') outcome-oriented navigation may not lead to any action enhancement or moral progress since they will only help to reaffirm and encourage previously embedded belief systems. Some scholars argue technologies pursuing operations similar to the here depicted 'heteronomous' outcome-oriented navigation could lead to ethical decisions. For example, Klincewicz (2016) argues that to sufficiently address the moral lag problem, "the moral AI would have to play not only an advisory or facilitative role, but also a normative one" (p.177) by suggesting concrete actions that the system derives from embedded ethical benchmarks. However, according to the motivated cognition thesis, if the IDSSs' recommendations do not appear to align with their own moral intuitions, the user will not act on them anyway, suggesting no actual change in the decision outcome will result. In any case, blindly following a system's recommendation may lead to a user's decision that is "performed merely in accordance with duty rather than from duty" (Frank, 2020; p.381). This "undermines the very nature of morality" (Lara and Deckers, 2020; p.280), so the resulting decision cannot be considered 'ethical' after all. In addition, this could lead to restricted moral progress or morale on an individual and collective level, as the discovery and application of new values are limited to the outputs of IDSSs (which is not the case for process-oriented navigation). Thus, the last takeaways are:

Takeaway 8 – Action enhancement:

- An IDSS's contribution to humans' action enhancement can be attributed to process-oriented navigation as well as outcome-oriented navigation.
- Since outcome-oriented navigation limits the user's decision autonomy, it is questionable whether the resulting decision can be considered 'ethical' and is compatible with the core of what constitutes morality.

Takeaway 9 – Moral progress:

- An IDSS's contribution to moral progress decreases when moving from process-oriented to outcome-oriented navigation.

### 5.2. Theoretical and practical implications

This study aimed to establish clarity in the conceptions and (directions of) influences that have been sketched in previous literature, which investigated the impact of IDSSs on humans' ethical decision-making. As a reminder, previous studies have suggested that different kinds of IDSSs lead to different kinds of moral enhancements. For example, Lara and Deckers (2020) suggested that IDSSs can be distinguished into three types: (1) systems that interact with users throughout their moral reasoning process in a dialectic manner so that the users can arrive at "better" decisions themselves; (2) systems that provide advice to users based on the users' predetermined values and (3) systems that supplant the entire moral reasoning process without any active participation of users. Based on our literature review, we derived a similar distinction by differentiating between 'process-oriented navigation', 'participatory outcome-oriented navigation' and 'heteronomous outcome-oriented navigation'. As we carved out explicit features that each navigation type characterizes (see Table 3), we extend and refine the previously proposed distinction by Lara and Deckers (2020). Similarly, in this study, we break down the previously established types of moral enhancement by Liu et al. (2022) (i.e., 'broad AI moral enhancement' and 'narrow AI moral enhancement') by specifying four concrete enhancements that are achieved along the entire ethical decision-making process of users and three related repercussions for society overall. We link these outcomes to each navigation type (see 5.1. Proposed interrelations) and thereby add on earlier studies by showing through which features exactly, IDSSs can facilitate or inhibit particular types of moral enhancement and long-term consequences.

Overall, *we provide tools (*i.e*., vocabulary, classifications and an integrated framework proposing interrelations) to facilitate discussion and future research* about the influence of IDSSs on humans' ethical decision-making. More specifically, the nine established takeaways can serve as propositions to be empirically tested in longitudinal studies or expert interviews in the future. Enriching the respective research field with different methodological investigations is important when considering the low number of existing empirical studies to this date (see Fig. B.1).

Next to these theoretical contributions, this study holds important implications for practice. Namely, it provides technology companies with information on what are (un)intended sources of influence within IDSSs (i.e., the system's human creators and the system itself), what are different operations and features that can be adopted in the design of systems (i.e., process-oriented navigation and outcome-oriented) and what are corresponding ramifications for the ethical decision-making on an individual and societal level. These insights can serve as checklists to guide companies in their development or governance processes when producing IDSSs that aim to assist humans' ethical decision-making. For example, when considering the negative consequences that can be expected from (heteronomous) outcome-oriented navigation (see 5.1. Proposed interrelations), it could be argued that (where possible) companies should *refrain from building such technology and focus on IDSSs that pursue process-oriented navigation*. Essentially, even if all technical bugs or biases were fixed and IDSS indeed led to decisions that are 'more ethical', it is expected that outcome-oriented IDSS still entail detrimental ramifications for critical human capacities (i.e., ethical decision-making and free will). Thus, contemporary scholars state such systems "ought to be used only as an aid or tool for human operators – never as a replacement for them" (French and Lindsay, 2022; p.72) by helping them to reach better decisions themselves instead of taking over these decisions completely (Lara and Deckers, 2020).

If companies still wish to develop IDSSs that favor outcome-oriented navigation, the findings of this study can *point to areas for which companies will need to establish countermeasures*. Namely, according to this study, these measures will need to counteract the negative impacts on users' deliberation and autonomy enhancement, as well as attenuate the emergence of moral deskilling, restricted moral progress and moral responsibility gaps. For example, to prevent users lose their ability to decide in accordance with their own ethical principles (i.e., autonomy enhancement) when engaging with an IDSS that operates on outcome-oriented navigation, the system should disclose the moral stances that fed into its calculation and led to the moral advice (Biller-Andorno et al., 2022). This transparency may enable users to realize whether the underlying reasoning process and resulting moral advice resonate with their own ethical standards. Similarly, to inhibit the unfolding of moral responsibility gaps, systems that deploy outcome-oriented navigation could clearly state that users always have "to assess the machine outputs for truth, soundness, and moral acceptability" and bear ultimate accountability as they are the ones "responsible for what […] they choose to do with the models' output" (Bang et al., 2023; p.7).

## 6. Conclusion and future research directions

Individuals make wide use of technology to assist their decision-making, even in the most critical areas that entail ethical decision-making, such as in healthcare, criminal justice or the military sector (Ogunbiyi et al., 2021; Vallor, 2013). Therefore, scholars have called for exploring the corresponding implications for humans' ability and independence when making ethical decisions (Scherer and Neesham, 2020). This article aimed to follow this request by synthesizing insights from existing literature, which deals with the influence of intelligent decision-support systems on humans' ethical decision-making. Based on the established literature synthesis, an integrated framework and potential interrelations were proposed, depicting answers to the initial research questions. That is to say, it is demonstrated to what extent particular IDSSs hinder or facilitate humans' ethical decision-making on an individual level (i.e., deliberation enhancement, motivation enhancement, autonomy enhancement and action enhancement) as well as on a societal level (i.e., moral deskilling, restricted moral progress and moral responsibility gaps). Furthermore, it is demonstrated by what mechanisms they do so by illustrating underlying sources (i.e., human creators and the technological system) and different methods/navigation types (i.e., process-oriented or outcome-oriented navigation) and their corresponding features. Thereby, this article generates important theoretical and practical insights. For one, our study contributes to the scholarly community by establishing clarity in the conceptions and (directions of) influences that have been sketched in previous literature, providing a foundation for future research. For practitioners such as technology companies, the insights of this study can serve as guides of what needs to be contemplated/addressed during the design process of pertinent IDSSs.

This research is not without limitations. Our focused selection excludes parts of the literature, which could offer additional insights. For instance, only publications were considered as 'in scope' that dealt with systems aiming to assist or fully automate ad hoc humans' ethical decision-making, while publications that discussed other technologies, such as games or virtual environments designed for ex-ante education of humans' ethical decision-making were neglected. In the future, the contemplation of such articles could prove useful in validating and expanding this article's findings (e.g., identified features). For example, Zarglayoun et al. (2022) discovered that incorporating social reinforcement messages in serious video games yielded higher levels of socio-moral reasoning. Moreover, in our literature analysis, we omitted considering contextual and individual factors that may moderate the impact of IDSSs on humans' ethical decision-making. For example, individual attributes that may affect the relationship could be the user's personal experience in making ethical decisions (Ogunbiyi et al., 2021). For the sake of analyzing the influence relationship in-depth as opposed to in breadth, such external variables were not picked up in this literature review but could be incorporated in future replication studies. Finally, due to the mainly theoretical nature of identified relevant publications (see Fig. B.1), an important limitation of this literature review is the lack of empirical evidence to support the findings and propositions. It is expected and endorsed that the here postulated interrelations will be tested and empirically validated in future studies. After all, as highlighted by the range and nature of the here identified consequences, IDSSs are technologies that "threaten to undo a foundational human capacity [(i.e., ethical decision-making) and, thus] deserve closer [moral] scrutiny" (Eisikovits and Feldman, 2022; p.197). Delving deeper into comprehending the range of influences of IDSSs on human ethical decision-making, along with the underlying sources and methods involved, will enable us to pursue a more reflected, human-centered approach to the use, governance and development of IDSSs.

## Funding

The authors gratefully acknowledge the financial support from the Technical University of Munich - Institute for Ethics in Artificial Intelligence (IEAI). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IEAI or its partners.

## Availability of data and material

Descriptive analyses of the identified publications that underlie this literature review can be found in the Appendix.

## CRediT authorship contribution statement

**Franziska Poszler:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization. **Benjamin Lange:** Writing – review & editing, Validation.

## Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose.

## Acknowledgements

This article is based on the bachelor's thesis of Franziska Poszler. The authors would like to thank the reviewers for their valuable comments and suggestions.

## Appendix A. Overview of the literature search process

| Database | Link | Query | Hits 12.01.2024 | Hits (without dublicates) 12.01.2024 | Hits: Titel & abstract analysis | Hits: Full paper analysis | Hits: Forward & backward search (until 12.01.2024) | Total | Comments |
|---|---|---|---|---|---|---|---|---|---|
| **Scopus** | https://www.scopus.com/ | TITLE-ABS-KEY ( ( "ethical decision making" OR "moral decision making" OR "ethical reasoning" OR "moral reasoning" OR "moral psychology" OR "technological mediation" OR "technomoral change" OR "moral deskilling" OR "moral enhancement" OR "moral character" OR "moral virtue" OR "moral progress" OR "moral thinking" OR "moral agency" OR "moral improve*" OR "moral advice" OR "moral responsib*" ) AND ( "artificial intelligence" OR "AI" OR "algorithm*" OR "automation" OR "robot*" OR "decision-support system*" OR "decision-support tool*" OR "moral advisor*" OR "decisional guidance" OR "machine*" OR "moral technolog*" OR "virtual assistant" OR "digital assistant" OR "ethics assistant" OR "moral expert*" OR "autonomous system*" ) ) AND PUBYEAR > 1974 AND PUBYEAR < 2024 AND ( LIMIT-TO ( LANGUAGE , "English" ) ) | 1.171 | | | | | | |
| **Web of Science** Science Citation Index Expanded (SCI-EXPANDED; 1945-present) Social Sciences Citation Index (SSC; 1985-present) Conference Proceedings Citation Index - Science (CPCI-S; 1990-present) Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH;1990-present) | https://apps.webofknowledge.com/ | TS=(("ethical decision making" OR "moral decision making" OR "ethical reasoning" OR "moral reasoning" OR "moral psychology" OR "technological mediation" OR "technomoral change" OR "moral deskilling" OR "moral enhancement" OR "moral character" OR "moral virtue" OR "moral progress" OR "moral thinking" OR "moral agency" OR "moral improve*" OR "moral advice" OR "moral responsib*" ) AND ("artificial intelligence" OR "AI" OR "algorithm*" OR "automation" OR "robot*" OR "decision-support system*" OR "decision-support tool*" OR "moral advisor*" OR "decisional guidance" OR "machine*" OR "moral technolog*" OR "virtual assistant" OR "digital assistant" OR "ethics assistant" OR "moral expert*" OR "autonomous system*" )) | 493 | | | | | | Refined by: LANGUAGES: ( ENGLISH ); no publication listed yet for 2024; thus, no refinement to year (<2024) necessary |
| **IEEE Explore** | https://ieeexplore.ieee.org/Xplore/home.jsp | ("Abstract":"ethical decision making" OR "Abstract":"moral decision making" OR "Abstract":"ethical reasoning" OR "Abstract":"moral reasoning" OR "Abstract":"moral psychology" OR "Abstract":"technological mediation" OR "Abstract":"technomoral change" OR "Abstract":"moral deskilling" OR "Abstract":"moral enhancement" OR "Abstract":"moral character" OR "Abstract":"moral virtue" OR "Abstract":"moral progress" OR "Abstract":"moral thinking" OR "Abstract":"moral agency" OR "Abstract":"moral improve*" OR "Abstract":"moral advice" OR "Abstract":"moral responsib*") AND ("Abstract":"artificial intelligence" OR "Abstract":"AI" OR "Abstract":"algorithm*" OR "Abstract":"automation" OR "Abstract":"robot*" OR "Abstract":"decision-support system*" OR "Abstract":"decision-support tool*" OR "Abstract":"moral advisor" OR "Abstract":"decisional guidance" OR "Abstract":"machine" OR "Abstract":"moral technolog*" OR "Abstract":"virtual assistant" OR "Abstract":"digital assistant" OR "Abstract":"ethics assistant" OR "Abstract":"moral expert*" OR "Abstract":"autonomous system*") | 58 | 1.211 | 112 | 34 | 11 | **45** | no publication listed yet for 2024; thus, no refinement to year (<2024) necessary |
| | | ("Document Title":"ethical decision making" OR "Document Title":"moral decision making" OR "Document Title":"ethical reasoning" OR "Document Title":"moral reasoning" OR "Document Title":"moral psychology" OR "Document Title":"technological mediation" OR "Document Title":"technomoral change" OR "Document Title":"moral deskilling" OR "Document Title":"moral enhancement" OR "Document Title":"moral character" OR "Document Title":"moral virtue" OR "Document Title":"moral progress" OR "Document Title":"moral thinking" OR "Document Title":"moral agency" OR "Document Title":"moral improve*" OR "Document Title":"moral advice" OR "Document Title":"moral responsib*") AND ("Document Title":"artificial intelligence" OR "Document Title":"AI" OR "Document Title":"algorithm*" OR "Document Title":"automation" OR "Document Title":"robot*" OR "Document Title":"decision-support system*" OR "Document Title":"decision-support tool*" OR "Document Title":"moral advisor" OR "Document Title":"decisional guidance" OR "Document Title":"machine" OR "Document Title":"moral technolog*" OR "Document Title":"virtual assistant" OR "Document Title":"digital assistant" OR "Document Title":"ethics assistant" OR "Document Title":"moral expert*" OR "Document Title":"autonomous system*") | 9 | | | | | | |
| | | SUM: | 1.731 | | | | | | |

**Fig. A.1.** Literature search process including consulted databases, search queries and corresponding hits.

## Appendix B. Descriptive/structural analysis of literature

**Table B.1**
Overview of research outlets by discipline and journal/publication type.

| Primary subject category/discipline | Journal/publication type | No. of publications (%) | |
|---|---|---|---|
| Arts and Humanities | Moral Philosophy and Politics | 1 | 3 (7,5 %) |
| | Studies in Logic, Grammar and Rhetoric | 1 | |
| | Techné: Research in Philosophy and Technology | 1 | |
| Computer Science | AISB Quarterly | 1 | 12 (30 %) |
| | Book (chapter) | 1 | |
| | Conference Proceeding | 7 | |
| | Frontiers in Artificial Intelligence | 1 | |
| | Information and Management | 1 | |
| | International Journal of Social Robotics | 1 | |
| Engineering | Book (chapter) | 1 | 3 (7,5 %) |
| | Conference Proceeding | 2 | |
| Ethics & Philosophy | AI and Ethics | 2 | 19 (47,5 %) |
| | BMC Medical Ethics | 1 | |
| | Book (chapter) | 3 | |
| | Conference Proceeding | 1 | |
| | Ethics and Information Technology | 1 | |
| | Journal of Medical Ethics | 1 | |
| | Neuroethics | 1 | |
| | Philosophies | 1 | |
| | Philosophy & Technology | 2 | |
| | Science and Engineering Ethics | 3 | |
| | Theoretical Medicine and Bioethics | 1 | |
| | White paper | 2 | |
| Multidisciplinary | Nature | 1 | 1 (2,5 %) |
| Social Sciences | Journal of International Political Theory | 1 | 2 (5 %) |
| | Social Sciences | 1 | |
| | | Total: 40 (100 %) | |

As illustrated in Table B.1, 42,2 % of the relevant publications were issued in journals, books and conferences within the discipline of Ethics & Philosophy. Following this, the discipline of Computer Science ranks second with 12 (26,7 %) pertinent publications. 7 (15,5 %) articles were published in outlets of the discipline of Arts and Humanities. The Social Sciences and Engineering discipline each issued 3 (6,7 %) publications in this research area. Lastly, within one multidisciplinary journal 1 (2,2 %) corresponding article was published.

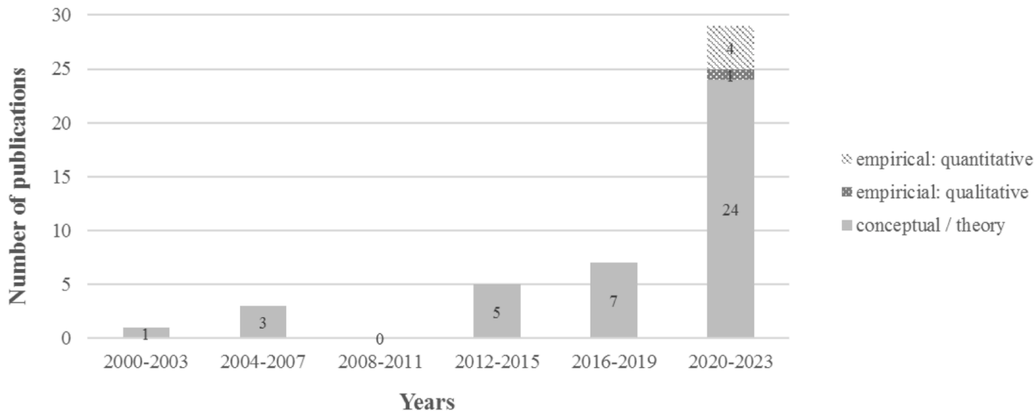


**Fig. B.1.** Publications (and utilized methodology) over time (2000–2023).

As illustrated in Fig. B.1, publications focusing on the influence of IDSSs on humans' ethical decision-making have increased over time. Especially in the past four years, pertinent research has more than quadrupled from 7 publications between 2016 and 2019 to overall 29 publications between 2020 and 2023. In addition, while all publications until the end of 2019 are purely conceptual/theoretical, since 2020, scholars have additionally started utilizing empirical methodologies within their research (5 respective publications in total).

## References*


* Ach, J.S., Beck, B., 2023. Transhumanism and moral enhancement. In: Jotterand, F., Ienca, M. (Eds.), The Routledge Handbook of the Ethics of Human Enhancement. Routledge, New York, pp. 267–281.

Auernhammer, J., Boess, S., Cheung, M., Cain, R., 2020. Human-centered AI: The Role of Human-centered Design Research in the development of AI. In: Synergy-DRS International Conference 2020. https://doi.org/10.21606/drs.2020.282.

Badea, C., 2022. Have a break from making decisions, have a MARS: the multi-valued action reasoning system. In: International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer International Publishing, pp. 359–366. https://doi.org/10.1007/978-3-031-21441-7_31.

Bandura, A., 1991. Social cognitive theory of moral thought and action. In: Kurtines, W. M., Gewirtz, J.L. (Eds.), Handbook of Moral Behavior and Development, vol. 1. Erlbaum, pp. 54–103.

* Bang, Y., Lee, N., Yu, T., Khalatbari, L., Xu, Y., Cahyawijaya, S., Fung, P., 2023. Towards Answering Open-ended Ethical Quandary Questions.

* Benzinger, L., Ursin, F., Balke, W.T., Kacprowski, T., Salloch, S., 2023. Should artificial intelligence be used to support clinical ethical decision-making? A systematic review of reasons. BMC Med. Ethics 24 (1), 1–9. https://doi.org/10.1186/s12910-023-00929-6.

* Berber, A., 2023. Automated decision-making and the problem of evil. AI Soc. 1-10. https://doi.org/10.1007/s00146-023-01814-x.

* Biggar, N., 2023. An ethic of military uses of artificial intelligence: sustaining virtue, granting autonomy, and calibrating risk. Conatus - J. Philos. 8 (2), 67–76. https://doi.org/10.12681/cjp.34666.

* Biller-Andorno, N., Ferrario, A., Joebges, S., Krones, T., Massini, F., Barth, P., Arampatzis, G., Krauthammer, M., 2022. AI support for ethical decision-making around resuscitation: proceed with care. J. Med. Ethics 48 (3), 175–183. https://doi.org/10.1136/medethics-2020-106786.

* Boddington, P., 2021. AI and moral thinking: how can we live well with machines to enhance our moral agency?. AI Ethics 1 (2), 109–111. https://doi.org/10.1007/s43681-020-00017-0.

Brocke, J.V., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A., 2009. Reconstructing the giant: on the importance of rigour in documenting the literature search process. In: ECIS 2009 Proceedings. https://aisel.aisnet.org/ecis2009/161.

Campbell, R., 2017. Learning from moral inconsistency. Cognition 167, 46–57. https://doi.org/10.1016/j.cognition.2017.05.006.

* Cappuccio, M.L., Sandoval, E.B., Mubin, O., Obaid, M., Velonaki, M., 2021. Can robots make us better humans?. Int. J. Soc. Robot. 13 (1), 7–22. https://doi.org/10.1007/s12369-020-00700-6.

Carlson, D.S., Kacmar, K.M., Wadsworth, L.L., 2009. The impact of moral intensity dimensions on ethical decision-making: assessing the relevance of orientation. J. Manag. Issues 21 (4), 534–551.

Corley, K.G., Gioia, D.A., 2004. Identity ambiguity and change in the wake of a corporate spin-off. Adm. Sci. Q. 49 (2), 173–208. https://doi.org/10.2307/4131471.

Cushman, F., Young, L., Hauser, M., 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. Psychol. Sci. 17, 1082–1089.

Danaher, J., 2019. Why internal moral enhancement might be politically better than external moral enhancement. Neuroethics 12 (1), 39–54. https://doi.org/10.1007/s12152-016-9273-8.

* De Boer, B., Kudina, O., 2021. What is morally at stake when using algorithms to make medical diagnoses? Expanding the discussion beyond risks and harms. Theor. Med. Bioeth. 42 (5), 245–266. https://doi.org/10.1007/s11017-021-09553-0.

* De Cremer, D., Narayanan, D., 2023. How AI tools can—and cannot—help organizations become more ethical. Front. Artif. Intell. 6, 1093712. https://doi.org/10.3389/frai.2023.1093712.

DeGrazia, D., 2014. Moral enhancement, freedom, and what we (should) value in moral behaviour. J. Med. Ethics 40 (6), 361–368. https://doi.org/10.1136/medethics-2012-101157.

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

* Eisikovits, N., Feldman, D., 2022. AI and phronesis. Moral Philos. Politics 9 (2), 181–199. https://doi.org/10.1515/mopp-2021-0026.

* Erler, A., Müller, V.C., 2021. AI as IA: human enhancement through artificial intelligence (AI) for intelligence augmentation (IA)? In: Iena, M., Jotterand, F. (Eds.), The Routledge Handbook of the Ethics of Human Enhancement. Routledge Taylor & Francis Group.

Fereday, J., Muir-Cochrane, E., 2006. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 5 (1), 80–92. https://doi.org/10.1177/160940690600500107.

Firth, R., 1952. Ethical absolutism and the ideal observer. Philos. Phenomenol. Res. 12 (3), 317–345. https://doi.org/10.2307/2103988.

* Frank, L.E., 2020. What do we have to lose? Offloading through moral technologies: moral struggle and progress. Sci. Eng. Ethics 26 (1), 369–385. https://doi.org/10.1007/s11948-019-00099-y.

* French, S.E., Lindsay, L.N., 2022. Artificial intelligence in military decision-making: avoiding ethical and strategic perils with an option-generator model. In: Koch, B., Schoonhoven, R. (Eds.), Emerging Military Technologies. Brill Nijhoff, pp. 53–74. https://doi.org/10.1163/9789004507951_007.

Gioia, D.A., Corley, K.G., Hamilton, A.L., 2013. Seeking qualitative rigor in inductive research: notes on the Gioia methodology. Organ. Res. Methods 16 (1), 15–31. https://doi.org/10.1177/1094428112452151.

* Giubilini, A., Savulescu, J., 2018. The artificial moral advisor. The "ideal observer" meets artificial intelligence. Philos. Technol. 31 (2), 169–188. https://doi.org/10.1007/s13347-017-0285-z.

Gonzalez Fabre, R., Camacho Ibáñez, J., Tejedor Escobar, P., 2021. Moral control and ownership in AI systems. AI Soc. 36, 289–303. https://doi.org/10.1007/s00146-020-01020-z.


* The asterisk-labeled references represent the 45 publications identified and analyzed as the foundation of this systematic literature review.

* Green, B.P., 2019. Artificial intelligence, decision-making, and moral deskilling. In: Markkula Center for Applied Ethics Website. Retrieved from. https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/artificial-intelligence-decision-making-and-moral-deskilling/.
Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108 (4), 814–834.
Henslee, E., Wiinikka-Lydon, J., Yeaman, A., 2022. Virtue ethics in robotics: an ethics module to cultivate character and ethical reasoning. In: 2022 ASEE Annual Conference & Exposition.
Higgins, A., Power, C., Kohlberg, L., 1984. Student judgments of responsibility and the moral atmosphere of high schools: a comparative study. In: Kurtines, W., Gewirtz, J. L. (Eds.), Morality, Moral Behavior and Moral Development: Basic Issues in Theory. Wiley, pp. 74–106.
* Inthorn, J., Tabacchi, M.E., Seising, R., 2015. Having the final say: machine support of ethical decisions of doctors. In: van Rysewyk, S., Pontier, M. (Eds.), Machine Medical Ethics. Springer, pp. 181–206. https://doi.org/10.1007/978-3-319-08108-3_12.
Jones, T.M., 1991. Ethical decision making by individuals in organizations: an issue-contingent model. Acad. Manag. Rev. 16 (2), 366–395. https://doi.org/10.5465/amr.1991.4278958.
Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. Am. Psychol. 58 (9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697.
Kempt, H., Heilinger, J.C., Nagel, S.K., 2023. "I'm afraid I can't let you do that, doctor": meaningful disagreements with AI in medical contexts. AI Soc. 38 (4), 1407–1414. https://doi.org/10.1007/s00146-022-01418-x.
* Klincewicz, M., 2016. Artificial intelligence as a means to moral enhancement. Stud. Log. Gramm. Rhetor. 48 (1), 61. https://doi.org/10.1515/slgr-2016-0061.
* Klincewicz, M., 2019. Robotic nudges for moral improvement through stoic practice. Techné: Res. Philos. Technol. 23 (3), 425–455. https://doi.org/10.5840/techne2019122109.
* Krügel, S., Ostermaier, A., Uhl, M., 2023. ChatGPT's inconsistent moral advice influences users' judgment. Sci. Rep. 13 (1), 4569. https://doi.org/10.1038/s41598-023-31341-0.
Kvalnes, Ø., 2015. Moral dilemmas. In: Moral Reasoning at Work: Rethinking Ethics in Organizations. Palgrave Macmillan, London, pp. 9–17. https://doi.org/10.1007/978-3-030-15191-1.
* Lara, F., 2021. Why a virtual assistant for moral enhancement when we could have a Socrates?. Sci. Eng. Ethics 27 (4), 1–27. https://doi.org/10.1007/s11948-021-00318-5.
* Lara, F., Deckers, J., 2020. Artificial intelligence as a socratic assistant for moral enhancement. Neuroethics 13 (3), 275–287. https://doi.org/10.1007/s12152-019-09401-y.
Leung, L., 2015. Validity, reliability, and generalizability in qualitative research. J. Family Med. Prim. Care 4 (3), 324–327. https://doi.org/10.4103/2249-4863.161306.
Lewis, P.V., 1989. Ethical principles for decision makers: A longitudinal survey. J. Bus. Ethics 8, 271–278. https://doi.org/10.1007/BF00383341.
* Liu, Y., Moore, A., Webb, J., Vallor, S., 2022. Artificial Moral Advisors: A New Perspective From Moral Psychology. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 436–445. https://doi.org/10.1145/3514094.3534139.
* Manders-Huits, N., 2006. Moral responsibility and IT for human enhancement. In: Proceedings of the 2006 ACM Symposium on Applied Computing, pp. 267–271. https://doi.org/10.1145/1141277.1141340.
* Mathieson, K., 2007. Dioptra: an ethics decision support system. In: AMCIS 2007 Proceedings. http://aisel.aisnet.org/amcis2007/469.
Meredith, R., Arnott, D., 2003. On ethics and decision support systems development. In: PACIS 2003 Proceedings, p. 106.
* Ogunbiyi, N., Basukoski, A., Chaussalet, T., 2021. An exploration of ethical decision making with intelligence augmentation. Soc. Sci. 10 (2), 57. https://doi.org/10.3390/socsci10020057.
* O'Neill, E., Klincewicz, M., Kemmer, M., 2022. Ethical issues with artificial ethics assistants. In: Veliz, C. (Ed.), The Oxford Handbook of Digital Ethics. Oxford University Press.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int. J. Surg. 88, 105906 https://doi.org/10.1016/j.ijsu.2021.105906.
Phillips-Wren, G., 2012. AI tools in decision making support systems: a review. Int. J. Artif. Intell. Tools 21 (02), 1240005. https://doi.org/10.1142/S0218213012400052.
Phillips-Wren, G., 2013. Intelligent decision support systems. In: Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications, 25-44. https://doi.org/10.1002/9781118522516.ch2.
Poszler, F., Geisslinger, M., Betz, J., Lütge, C., 2023. Applying ethical theories to the decision-making of self-driving vehicles: a systematic review and integration of the literature. Technol. Soc. 102350 https://doi.org/10.1016/j.techsoc.2023.102350.
Rawls, J., 1971. A Theory of Justice. Harvard University Press.
* Renic, N., Schwarz, E., 2023. Crimes of dispassion: autonomous weapons and the moral challenge of systematic killing. Ethics Int. Aff. 37 (3), 321–343. https://doi.org/10.1017/S0892679423000291.
Rest, J.R., 1986. Moral Development: Advances in Research and Theory. Praeger.
* Rodríguez-López, B., Rueda, J., 2023. Artificial moral experts: asking for ethical advice to artificial intelligent assistants. AI Ethics 1-9. https://doi.org/10.1007/s43681-022-00246-5.
* Savulescu, J., Maslen, H., 2015. Moral enhancement and artificial intelligence: moral AI? In: Romportl, J., Zackova, E., Kelemen, J. (Eds.), Beyond Artificial Intelligence. Springer, pp. 79–95. https://doi.org/10.1007/978-3-319-09668-1_6.
Scherer, A.G., Neesham, C., 2020. New challenges to enlightenment: why socio-technological conditions lead to organized immaturity and what to do about it. In: Working Paper. https://doi.org/10.2139/ssrn.3753612.
Schwartz, M.S., 2016. Ethical decision-making theory: an integrated approach. J. Bus. Ethics 139, 755–776. https://doi.org/10.1007/s10551-015-2886-8.
* Schwarz, E., 2018. Technology and moral vacuums in just war theorising. J. Int. Political Theory 14 (3), 280–298. https://doi.org/10.1177/1755088217750689.
* Schwarz, E., 2023. Cybernetics at war: military artificial intelligence, weapon systems and the de-skilled moral agent 1. In: Gruszczak, A., Kaempf, S. (Eds.), Routledge Handbook of the Future of Warfare. Routledge, pp. 297–307.
* Seville, H., Field, D., 2000. What can AI do for ethics?. AISB Q. 31–34.
* Shaikh, S.J., 2020. Artificial Intelligence and Resource Allocation in Health Care: The Process-Outcome Divide in Perspectives on Moral Decision-Making. In: AAAI Fall 2020 Symposium on AI for Social Good.
Stefan, R., Carutasu, G., 2020. How to approach ethics in intelligent decision support systems. In: Innovation in Sustainable Management and Entrepreneurship: 2019 International Symposium in Management (SIM2019). Springer International Publishing, pp. 25–40. https://doi.org/10.1007/978-3-030-44711-3_3.
* Straßmann, C., Eimler, S.C., Arntz, A., Grewe, A., Kowalczyk, C., Sommer, S., 2020. Receiving robot's advice: does it matter when and for what?. In: International Conference on Social Robotics. Springer, pp. 271–283. https://doi.org/10.1007/978-3-030-62056-1_23.
* Susser, D., 2019. Invisible influence: artificial intelligence and the ethics of adaptive choice architectures. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 403–408. https://doi.org/10.1145/3306618.3314286.
* Tassella, M., Chaput, R., Guillermin, M., 2023. Artificial moral advisors: enhancing human ethical decision-making. In: 2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS). IEEE, pp. 1–5. https://doi.org/10.1109/ETHICS57328.2023.10155026.
Theurer, C.P., Tumasjan, A., Welpe, I.M., Lievens, F., 2018. Employer branding: a brand equity-based literature review and research agenda. Int. J. Manag. Rev. 20 (1), 155–179. https://doi.org/10.1111/ijmr.12121.
* Vallor, S., 2013. The future of military virtue: autonomous systems and the moral deskilling of the military. In: 2013 5th International Conference on Cyber Conflict. IEEE, pp. 1–15.
* Vallor, S., 2015. Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. Philos. Technol. 28 (1), 107–124. https://doi.org/10.1007/s13347-014-0156-9.
* van de Voort, M., Pieters, W., Consoli, L., 2015. Refining the ethics of computer-made decisions: a classification of moral mediation by ubiquitous machines. Ethics Inf. Technol. 17 (1), 41–56. https://doi.org/10.1007/s10676-015-9360-2.
* van der Waa, J.V.D., Diggelen, J.V., Cavalcante Siebert, L., Neerincx, M., Jonker, C., 2020. Allocation of moral decision-making in human-agent teams: a pattern approach. In: International Conference on Human-Computer Interaction. Springer, pp. 203–220. https://doi.org/10.1007/978-3-030-49183-3_16.
Verbeek, P.P., 2006. Materializing morality: design ethics and technological mediation. Sci. Technol. Hum. Values 31 (3), 361–380. https://doi.org/10.1177/0162243905285847.
Verbeek, P.P., 2008. Obstetric ultrasound and the technological mediation of morality: a postphenomenological analysis. Hum. Stud. 31 (1), 11–26. https://doi.org/10.1007/s10746-007-9079-0.
* Volkman, R., Gabriels, K., 2023. AI moral enhancement: upgrading the socio-technical system of moral engagement. Sci. Eng. Ethics 29 (2), 11. https://doi.org/10.1007/s11948-023-00428-2.
Wallach, W., Allen, C., 2009. Moral Machines. Oxford University Press.
Webster, J., Watson, R.T., 2002. Analyzing the past to prepare for the future: writing a literature review. MIS Q. 26 (2), xiii–xxiii.
* Wong, P.H., 2019. Rituals and machines: a Confucian response to technology-driven moral deskilling. Philosophies 4 (4), 59. https://doi.org/10.3390/philosophies4040059.
* Yang, L., Frize, M., Eng, P., Walker, R.C., Catley, C., 2004. Towards ethical decision support and knowledge management in neonatal intensive care. In: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 3420–3423. https://doi.org/10.1109/IEMBS.2004.1403960.
Zarglayoun, H., Laurendeau-Martin, J., Tato, A., Vera-Estay, E., Blondin, A., Lamy-Brunelle, A., Chaieb, S., Morasse, F., Dufresne, A., Nkambou, R., Beauchamp, M.H., 2022. Assessing and optimizing socio-moral reasoning skills: findings from the MorALERT serious video game. Front. Psychol. 12 https://doi.org/10.3389/fpsyg.2021.767596.
Zollo, L., Pellegrini, M.M., Ciappei, C., 2017. What sparks ethical decision making? The interplay between moral intuition and moral reasoning: lessons from the scholastic doctrine. J. Bus. Ethics 145, 681–700. https://doi.org/10.1007/s10551-016-3221-8.

**Franziska Poszler** is a doctoral candidate at the Chair of Business Ethics and a research associate at the Institute for Ethics in AI (IEAI) at the Technical University of Munich (TUM). Her research interests include machine ethics, value-sensitive design and moral psychology, with a focus on investigating reciprocal influences between emerging technologies, morality and ethical decision-making. Franziska completed a B.A. in Philosophy as well as a B.Sc. in Business Administration with a minor in Psychology at the Ludwig-

Maximilians-University of Munich. She received her M.Sc. in Organisational Behaviour from the London School of Economics and Political Science in the UK.

**Benjamin Lange** is a Junior Research Group Lead in the Ethics of AI at the Ludwig Maximilians University of Munich (LMU) and Munich Center for Machine Learning (MCML). He holds an Associate Researcher Position at the Oxford Uehiro Centre for Practical Ethics at the University of Oxford and is a member of the Zentrum für Ethik und Philosophie in der Praxis (ZEPP) at LMU. Previously, he was a Visiting Researcher at Google's Responsible Innovation and AI Ethics team. He received his PhD in Moral Philosophy from the University of Oxford. His research focuses on organizational and business ethics, the ethics of AI and technology, and foundational issues in normative ethics.