ORIGINAL RESEARCH

A context-specific analysis of ethical principles relevant for AI-assisted decision-making in health care

Larissa Schlicht[1,2] · Miriam Räker[3,4]

Abstract

Artificial intelligence (AI)-assisted technologies may exert a profound impact on social structures and practices in care contexts. Our study aimed to complement ethical principles considered relevant for the design of AI-assisted technology in health care with a context-specific conceptualization of the principles from the perspectives of individuals potentially affected by the implementation of AI technologies in nursing care. We conducted scenario-based semistructured interviews focusing on situations involving moral decision-making occurring in everyday nursing practice with nurses ($N=15$) and care recipients ($N=13$) working, respectively, living in long-term care facilities in Germany. First, we analyzed participants' concepts of the ethical principles beneficence, respect for autonomy and justice. Second, we investigated participants' expectations regarding the actualization of these concepts within the context of AI-assisted decision-making. The results underscore the importance of a context-specific conceptualization of ethical principles for overcoming epistemic uncertainty regarding the risks and opportunities associated with the (non)fulfillment of these ethical principles. Moreover, our findings provide indications regarding which concepts of the investigated ethical principles ought to receive extra attention when designing AI technologies to ensure that these technologies incorporate the moral interests of stakeholders in the care sector.

Keywords Artificial intelligence · Nursing · Ethics · Stakeholder perspectives · Interview study

1 Introduction

The application of algorithms based on artificial intelligence (AI)[1] is spreading in the world of work and also in the health care sector. AI systems have the ability to imitate human problem solving, allowing them to assist with or perform tasks that require cognitive abilities (e.g., [2]). The transfer of agency from humans to AI-assisted technologies may, therefore, have a significant impact on social structures as well as practices in the health care context and the social and moral norms[2] manifested therein.

Recently, there has been an increase in the research and development of AI-assisted technologies for nursing care [5–7]. Against the backdrop of current challenges in professional care, such as the shortage of skilled workers, workforce aging and growing care needs resulting from increasingly aging societies and population growth [8], AI technologies promise to optimize nursing workflows, e.g., by providing automated tracking and analysis of care recipients'

---

[1] While there is, at least for the time being, no agreed definition of AI technologies, it is generally assumed that such technologies include computer-based systems that can, for a given set of objectives, influence their environment by producing outputs such as predictions, recommendations or decisions. The Organisation for Economic Cooperation and Development (OECD), for instance, specifies that an AI system "uses machine and/or human-based data and inputs to (i) perceive real and/or virtual environments; (ii) abstract these perceptions into models through analysis in an automated manner (e.g., with machine learning), or manually; and (iii) use model inference to formulate options for outcomes" [1, p. 7].

✉ Larissa Schlicht
   schlicht.larissa@baua.bund.de

1   German Federal Institute for Occupational Safety
    and Health, Dresden, Germany

2   Faculty of Humanities and Social Sciences, Karlsruhe
    Institute of Technology, Karlsruhe, Germany

3   German Federal Ministry of Health, Berlin, Germany

4   School of Public Health, Bielefeld University, Bielefeld,
    Germany

[2] A widely shared assumption in ethics as well as the social sciences is that *social norms* are exogenous informal rules that govern (and often constrain) behavior in groups and societies [3]. *Moral norms*, in particular, can be defined as ideals (i.e., moral imperatives) that prescribe how people—considered free to decide—should behave toward others and themselves. The influence of technologies on such norms can be understood as tendencies to condition their environments to behave or be organized according to the rationales of the norms [4, p. 47].

activities and health data as well as identifying options for clinical decision-making[3] [6, 10].

Nurses assist individuals in activities that "contribute to their health or recovery or to dignified death that they would perform unaided if they had the necessary strength, will, or knowledge" [11]. In this, they take responsibility for the well-being of humans who are limited in their decision-making ability and/or dependent on professional care.

Advocating for the interests and needs of those in need of care is a key aspect of professional care. Hence, nurses are often confronted with complex decisions that require the inclusion of multiple perspectives, taking into account the individual situation of care recipients. Frequently, their decisions have morally significant consequences (e.g., [12, 13]).

Areas of application of AI-assisted technologies already in use range from activity and health tracking to care coordination and communication; the systems are based on, e.g., computer vision, predictive modeling, natural language processing, or speech recognition [7]. It has been shown that such technologies can make assessments and processes more efficient—such as by early detection and prevention of adverse events or by reducing the time needed for documentation—enabling nurses to focus on humanistic aspects of care, including communication (e.g., [10]). Moreover, AI may offer the opportunity to make care services more personalized (by integrating individual health data) and to provide evidence-based health information for decision-making [14].

However, the implementation of AI-assisted technologies also creates new challenges. It has been discussed that the adoption of such technologies may be associated with adverse effects such as a depersonalization of the nurse-patient relationship [15] and impaired communication [16], thereby undermining the holistic approach to care practice. Depending on the system design and the field of application, the individuality of those in need of care could gradually disappear aside from what can be empirically captured (the so-called datafication of patients) (e.g., [17, 18]). Furthermore, a nonrepresentative selection of datasets and/or the quantification and categorization of data for the training of AI models contain the potential to discriminate against particular (sociodemographic) groups, such that their needs and characteristics are overlooked [19, 20]. Another possible drawback is that the reliance of nurses on algorithms could lead to a loss of the ability or willingness to critically reflect on their actions (e.g., [15]).

Overall, it can be said that the risk of neglecting care recipients' interests and (behavioral) repercussions within care processes is already present (i.e., independent of AI technology) but can be exacerbated by this technology and particularly by systems that have a direct impact on human–human relationships.

To ensure that the implementation of AI applications supports human agency in an ethically aligned way, there is a need to provide early identification of possible tendencies of implemented algorithms to contribute to a perpetuation or change of social structures and the moral norms anchored therein (e.g., [4]). At present, unintended consequences associated with a de-humanization or impersonalization of care are not being systematically assessed during the system design process though. Existing ethical guidelines for AI are usually formulated as highly abstract ethical principles that appear to be too indeterminate, i.e., normatively unambiguous to guide the design of technologies based on moral claims [21]. To effectively inform choices made during the design process, guidelines need to be specified for specific contexts of use.

This study complements ethical principles considered relevant for the design of AI-assisted technology in health care with a context-specific conceptualization of the principles from the perspectives of individuals potentially affected by the implementation of AI technologies in long-term care facilities in Germany. With this approach, we provide indications regarding which concepts of the investigated ethical principles ought to receive particular attention during the design of AI technologies to ensure that these technologies are not blind to the moral interests of stakeholders in the German care sector.

2 The need to contextualize AI ethics frameworks

The need to develop norms and standards to achieve ethically aligned AI systems is being critically discussed by various organizations (e.g., [22, 23]), the private sector (e.g., [24]) and researchers (e.g., [25, 26]). Consequently, numerous ethical guidelines for AI have been developed in recent years [27, 28]. However, these guidelines seem to be rarely considered in practice [29]. This cannot solely be explained by the number of frameworks to choose from and/or the limited (sanction) mechanisms to date reinforcing their normative claims. An obstacle to the effective translation of ethical principles into practice is the high degree of epistemic uncertainty regarding the risks and opportunities associated with the (non-)fulfillment of ethical principles. To resolve this uncertainty, context-specific conceptualizations of the proposed principles, e.g., via bottom-up case studies with relevant stakeholders, are needed [30, 31]. The current guidelines are usually formulated as highly abstract principles that "leave much room for interpretation as to how they can be practically applied in specific contexts of use such as LTC [long-term care]" [32, p. 2].

---

[3] In clinical settings, such systems are often referred to as AI-based clinical decision support Systems (CDSS), which are designed to aid health professionals by generating patient-specific assessments or recommendations, taking into accounts the specific characteristics of individual patients [9].

Correspondingly, it is widely agreed that the design of technologies implemented in socially sensitive areas, such as the healthcare sector, should not solely be informed by predefined normative principles (adapted to the technology's abilities) but also by local phenomena (i.e., thick ethical concepts)[4] that appear morally salient to those that are potentially affected by the implementation of such technology (e.g., [31, 34, 35]). To adequately assess and operationalize stakeholders' perspectives, several researchers have stressed the need for a stronger investigation not only of stakeholders' situated conceptualizations of proposed principles (e.g., [36]) but also of possible associations of these conceptualizations with specific tasks [37]. Existing approaches that aim to translate stakeholder perspectives into design requirements in a principled manner, such as value-sensitive design (VSD) [38] or participatory design (PD) [39], usually do not consider aspects that address ethical principles' realization through situational factors embedded in specific real-life contexts (e.g., [40]).

Aiming to complement ethical principles with context-specific perspectives of individuals potentially affected by AI-assisted decision-making, we focus on the framework proposed by Beauchamp and Childress [41]. A mapping review conducted by Floridi et al. [42] of the literature on ethical guidelines for AI in health care suggests that the key principles incorporated by many AI initiatives are consistent with the ethical principles proposed herein.[5]

The most influential framework in health care practice (hereafter, referred to as principles of biomedical ethics) proposes the following four prima facie principles for the ethical evaluation of health care practice:

- *Beneficence*: all norms, dispositions and actions aiming to benefit or promote the well-being of other persons [41, pp. 217–218]. It "(1) present[s] positive requirements of action, [that] (2) need not always be followed impartially, and (3) generally do not provide reasons for legal punishment when agents fail to abide by them" [ibid., p. 219].
- *Nonmaleficence*: obligates to abstain from causing harm to others [ibid., p. 155]. It is conceptualized as "(1) … negative prohibitions of action, [that] (2) must be followed impartially, and (3) provide moral reasons for legal prohibitions of certain forms of conduct" [ibid., p. 219].
- *Respect for autonomy*: both the negative obligation that autonomous actions should not be subjected to controlling constraints and the positive obligation to disclose information as well as to promote the capacities for autonomous choice [ibid., p. 105]. The realization of the principle is assumed to require liberty (independence from controlling influences) and agency (capacity for intentional action) [ibid., p. 100].
- *Justice*: broadly defined as the obligation to fairly distribute benefits, risks and costs under conditions of scarce resources [ibid., pp. 13, 250]. In the absence of social consensus on specific theories of justice (such as utilitarian, libertarian, communitarian, egalitarian, capability and well-being theories), policies are expected to integrate various elements of these theories on a case-by-case basis [ibid., p. 313].

Further references on ethical principles considered relevant in care contexts occur in nursing theories with their respective value orientations [43], in professional codes of ethics (e.g., [44, 45]) and to some extent in other (bioethical) approaches of health care ethics [46–48]. In particular, relational theories of health care and nursing, such as the ethics of care [49–51], make normative claims against the principles of biomedical ethics. Based on the assertion that social relationships and the recognition of the vulnerability of those in need of care should be the focus of ethical considerations of care work, the principle respect for autonomy, in particular, is thought to be based on an overly individualistic view of human beings. We assume that such perspectives are not necessarily incompatible with the principles of biomedical ethics; instead, they could be integrated into the framework, along with the context-specific conceptualization as well as adaptation of the principles. In fact, Beauchamp and Childress conceptualized their principles as an analytical framework of general norms derived from common morality[6] that serves as a practical instrument for moral reasoning [41, p. 17] and requires further specification to provide direct guidance within specific contexts [ibid., p. 9].[7]

---

[4] Thick ethical concepts denote descriptive features of a situation that may be (considered) morally relevant. Within moral philosophy, situational ethicists emphasize the importance of taking such features into account to determine considerations relevant to the ethical evaluation of a particular situation [33].

[5] Note: Floridi et al. identify *explicability* as an additional, frequently discussed, instrumental (i.e., aiming to realize intrinsic principles), principle. In this study, however, we initially aimed to investigate intrinsic principles (versus instrumental principles).

[6] Common morality is considered to contain "moral norms that are abstract, universal and content-thin (such as "Tell the truth")" (ibid., p. 5), in contrast to particular moralities, which are considered to "present concrete, nonuniversal, and content-rich norms (such as "Make conscientious oral disclosures to and obtain a written informed consent from all human research subjects")" (ibid., p. 5).

[7] Specifically, Beauchamp and Childress argue that the "content of […] rules and principles is too abstract to determine the specific acts that we should and should not perform. In the process of specifying and balancing norms and in making particular judgments, we often must take into account factual beliefs about the world, cultural expectations, judgments of likely outcome, and precedents to help assign relative weights to rules, principles, and theories" ([41, p. 427]). The framework can thus be seen as a hybrid model combining top-down approaches, which provide deductive generation of moral justification, and bottom-up approaches, which assume that moral justification ought to be derived inductively through contextualized reasoning [52].

However, we narrowed our search space to the three principles of *beneficence*, *respect for autonomy* and *justice*. *Nonmaleficence* requires intentional avoidance of actions that (may) cause harm and are, therefore, legally prohibited (ibid. p. 219).[8] In our study, however, we wanted to encourage participants to reflect on decision-making situations in which their moral intuitions are (presumably) not primarily guided by internalized rules of conduct. More importantly, we decided not to include a scenario prompting reflection on the principle of *nonmaleficence* because we aimed to respond to the (potential) vulnerability of participants in the care-recipient group and minimize the risk of causing psychological/emotional harm (such as feeling uncomfortable, embarrassed, or upset) to them [53, 54]. Due to the mutual relations between the principles, it must still be assumed that some participant statements may also be related to the principle of *nonmaleficence*.

3 Research questions

While former studies have assessed, e.g., medical students' views of the *principles of biomedical ethics* (based on four scenarios) [55], the influence of the principles on health care practitioners' attitudes toward AI technology [56] and student rankings of the principles within decision-making in ethical scenarios [57], to our knowledge, no qualitative study has assessed whether the principles are morally salient to direct stakeholders in the German care sector. Moreover, no study to date has examined which situational factors of specific real-life contexts are thought to promote the actualization of ethical principles by stakeholders. As outlined in previous section, it is assumed that such complementary data will help to translate ethical principles into practice. Therefore, the present study first aimed to illuminate the established *principles of biomedical ethics* from the perspective of direct stakeholders in the German care sector, nurses and care recipients (to ensure that multiple perspectives are factored into the analysis [58]). To meet this goal, we formulated the following research questions:

Q1: Are the principles of *beneficence*, *respect for autonomy* and *justice* morally salient to participants?

Q2: How do participants conceptualize the principles? Which situational factors (in particular, demands) do participants regard as promoting the actualization of their concepts of these principles in situations involving moral decision-making occurring in everyday nursing practice?

We further aimed to provide initial indications of which concepts of the investigated ethical principles ought to receive particular attention when designing AI technologies to ensure that they are not blind to the moral interests of stakeholders in the German care sector. We, therefore, analyzed participant expectations regarding the actualization of their concepts of the principles in the context of AI-assisted decision-making in the third question.

Q3: Which potential influences do participants anticipate from the use of AI-assisted technology in situations involving moral decision-making (care tasks) with regard to the actualization of their concepts of the principles?

4 Methods

We conducted scenario-based semistructured interviews (see [59, 60]) focusing on situations involving moral decision-making occurring in everyday nursing practice. With this approach, we prompted participants to reflect upon the three ethical principles of *beneficence*, *respect for autonomy* and *justice* as well as the potential influences of AI-assisted technology on the actualization of the principles.

4.1 Participants

In total, semistructured interviews were conducted with 15 nurses and 15 care recipients between October 2021 and May 2022. In the care-recipient group, 2 interviews were excluded due to insufficient comprehensibility of their statements, resulting in 13 analyzable interviews. Recruitment took place through telephone and e-mail inquiries to long-term care facilities within Germany. Participants in the nurse group had to be employed as registered nursing professionals. Participants in the care-recipient group had to be at least 18 years old, without cognitive or communicative impairment (in everyday social life in the facility) and to have already received care for at least 1 year. Their demographic characteristics are reported in Table 1.

---

[8] Examples of specific moral rules that are supported by the principle of *nonmaleficence* include, for example, "do not kill", "do not cause pain or suffering", and "do not incapacitate" (ibid., p. 156).

Table 1 Participants' sociodemographic characteristics

Nurses (N = 15)	N	Care recipients (N = 13)	N
Age (years)		Age (years)	
18–30	3	18–30	0
31–45	6	31–45	3
46–55	2	46–55	1
56+	3	56+	9
Gender		Gender	
Female	9	Female	7
Male	6	Male	6
Diverse	0	Diverse	0
Function		Length of stay in a long-term care facility (years)	
Nursing staff	5	1–4	6
Nursing management	8	5+	7
Other	2		
Occupational experience (years)			
< 5	4		
5–9	0		
10–19	6		
20–29	4		
30+	1		

4.2 Procedure

For the nurse group, the duration of interviews ranged from 60 to 90 min.[9] We originally planned to conduct the interviews on-site (i.e., at the facility in which the participants lived or worked); however, in some cases, this was not possible due to the coronavirus disease 2019 (COVID-19) pandemic. Therefore, some interviews were conducted digitally. In the care-recipient group, the length of the interviews was limited to 60 min. Most of these interviews were conducted at the nursing home in which the care recipients lived at the time. Interview audio was recorded using a conventional voice recorder. All interviews were conducted by one of the authors.

4.3 Scenarios

With a multidisciplinary group of researchers and a registered nurse, we developed three scenarios, depicting different care tasks associated with moral decision-making as potential fields of application for AI technology [5–7]. The scenarios were revised after pilot testing with two

individuals. To assess the ecological validity of the scenarios, participants were asked whether they experienced decision-making situations in their (professional) everyday life similar to those described in the scenarios. Overall, agreement was high for all scenario variants.[10]

The first scenario (see Scenario SI 1) describes a situation in the field of basic care (bodily care), and the second scenario (see Scenario SI 2) describes a situation in the field of social care (interaction and relationship). In both scenarios, a nurse must decide whether to follow the expressed will of a person in need of care or to perform a care task against his or her will, i.e., the nurse must weigh the principles of *respect for autonomy* and *beneficence*. The third scenario (see Scenario SI 3) describes a situation in which workflows must be prioritized (organization of workflows) due to staff shortages; specifically, a nurse has to decide between caring for one person (who needs emotional support) or caring for many (as part of routine on-site care). This scenario prompts reflection on the principle of *justice*.

Two versions of each scenario were presented, one in which the nurse decides with the support of an AI-assisted technology and one in which the nurse makes the decision without this technology.

Analysis of results related to Q1 and Q2 was primarily based on statements referring to scenarios *without* AI technology; in contrast, analysis of results related to Q3 primarily focused on statements referring to scenarios *with* AI technology. The presentation of both versions of each scenario was designed to increase the salience of the difference between the two situations. The resulting six situations were presented to participants in written form or, if necessary, aloud.

For each situation, participants were asked to answer questions concerning (a) possible implications of the outlined decision, (b) their moral evaluation of the outlined decision, and (c) their rationale for the evaluation made in (b). In addition, the participants were asked to describe their conception of *good care*. In order not to influence the moral reasoning of the participants and to be able to assign their statements inductively to the ethical principles, the participants were not given the definitions of the principles.

4.4 Data analysis

The recorded interviews were first pseudonymized and then transcribed. A content analysis following that of Kuckartz [62] using MAXQDA analysis software [63] was carried out. Participants were pseudonymized as follows: nurses were labeled as G1, G2,…, and G15; and care recipients

---

were labeled as R1, R2,…, and R13. The transcripts were analyzed by a stepwise construction of codes. Initial main codes were derived deductively from our research questions; further main codes and subcodes were derived inductively from the data. Together with a third researcher, we independently performed coding; occasional differences in our codes were discussed and resolved within the research team.

5 Results

5.1 Contextualization of biomedical ethics principles

In the qualitative content analysis, participant moral reasoning clearly reflected the three principles of *beneficence*, *respect for autonomy* and *justice* (Q1). However, the results also suggested that the principles' definitions may need to be extended to care-specific concepts.

Superordinate findings regarding participants' contextualized perspectives of the principles (Q2) are described below (principle concepts are italicized). Tables of all key aspects associated with the principles (including situational factors considered to promote the actualization of their concepts of the principles) as well as corresponding anchor quotations are provided in the Supplementary Information.

5.1.1 Beneficence

Participants' concepts of *beneficence* were highly multifaceted. Many facets referred to the relationship between the nurse and care recipient as well as specific caring actions. In other words, participants seemed to think of the principle as a dynamic process within care procedures that also impacts the actualization of the other principles.

The participants largely agreed that the overarching aim of *beneficence* is, on the one hand, the prevention of (physical) harm as well as the satisfaction of basic needs and, on the other hand, the promotion of care recipients' emotional well-being. This conceptualization is largely consistent with the definition of Beauchamp and Childress (2019).

As shown in Table SI 1, participant statements regarding critical requirements to achieving these aims (in situations involving moral decision-making) can be broadly grouped into three categories, namely *recognizing needs*, *assuming responsibility* and *meeting needs*. These requirements provide a nuanced understanding of the principle of *beneficence* within the context of long-term care. In particular, participants highlighted demands that specified "positive requirements for actions" [ibid., p. 204]. Participants pointed out that the recognition of care recipients' needs is the basis for the realization of subsequent aspects and demands on nurses to, inter alia, holistically assess care recipients' needs, e.g., "Caring requires perceiving the persons in need of care as comprehensively as possible. Their wishes, needs, problems" (G9).

The assumption of responsibility, preceding the performance of concrete nursing actions, was viewed as closely linked to the demand of obtaining extended information on the (health) condition of patients as well as weighing possible consequences associated with the available options for action. In addition, many participants highlighted that communication plays a central role for building trust within this stage of caring processes: "If we talk to the patients, for example, explain why a particular treatment is important, the patients usually allow the treatment to be carried out" (G15).

Finally, participants stressed that meeting care recipients' needs often requires nurses to respond to their patients according to a given situation and, if necessary, to adapt their (planned) actions accordingly, e.g., "The art of nursing involves applying abstract knowledge to the person and the specific situation" (G9).

5.1.2 Respect for autonomy

Participants' contextualized understanding of *respect for autonomy* was roughly categorized into the concepts of *individual autonomy* and *relational autonomy*, which differ in their respective aims and demands (see Table SI 2).

In line with the definition of Beauchamp and Childress (2019), many participants argued that *respect for autonomy* requires care recipients to be self-determined as well as free from interference when making decisions, e.g., "Respect for autonomy requires that I regard the person in need of care as the decision-maker" (G9). Correspondingly, participants emphasized that nurses should trust in patients' decision-making competency and, if necessary, improve their ability to make fully informed and independent decisions, e.g., "It is important to promote competence to make their own decisions… To do this, we often have to provide information" (G7). Limits to this understanding of patient autonomy are identified in associated risks of self-endangerment and harm for uninvolved persons[11] as well as regarding care recipients with cognitive impairments.

At the same time, many participants pointed out that patients' exercise of agency is usually embedded in social relationships and that patients may not be capable of claiming the right to autonomy. Accordingly, some participants reasoned that patient autonomy may also be preserved by retaining a person's sense of identity rather than independence, particularly with cognitively impaired persons. Thus,

---

[11] Note that in this case, participant statements could also be attributed to the principle of *nonmaleficence.*

autonomy should be understood as a relational process involving the demand to holistically assess care recipients' individual situation and motives. Several participants argued that nurses should consider the possibility of internalized incapacitation. Moreover, participants assumed that (relational) autonomy can also be preserved within shared decision-making. Relatedly, many emphasized the possible demand of ascertaining care recipients' motives and needs through nonverbal communication as well as through consulting colleagues, e.g., "To strengthen the autonomy of people in need of care, it is important to talk to colleagues from other professional groups about particular residents. This opens up new perspectives" (G12).

5.1.3 Justice

As depicted in Table SI 3, participants identified *nondiscrimination* and, more particularly, *distributive justice*, i.e., the fair allocation of resources, as focal concepts of *justice* in everyday nursing practice. These concepts also fit well into the broad definition of *justice* proposed by Beauchamp and Childress (2019).

Several participants argued that their concept of *justice* prohibits treating people differently due to characteristics such as "their religion or the color of their skin" (G6).

Many participants emphasized the relevance of a fair allocation of time and attention to care recipients, presumably due to the frequent scarcity of nursing staff, demanding that health professionals set priorities. However, the participants held different views on what constitutes a *fair* distribution of these resources. While some reasoned that nurses "… shouldn't concentrate on an individual patient because [they] might get the impression that he or she needs [them] more than than other patients" (G4) (i.e., the *equality principle*), others articulated the view that the allocation of resources should be based on individual needs for basic care and/or social support (i.e., the *need principle*).

In addition, several participants mentioned that the realization of these concepts is not always achievable in (professional) everyday life. An obstacle to the realization of the first concept is seen in that some care recipients may be more "visible" than others. An idiosyncratic issue with allocating resources on a strictly needs-oriented approach is considered to be the fact that care recipients' ability to articulate their needs may be limited due to cognitive and/or communicative impairments.

5.2 Expected influence of AI technology on the actualization of the principles

Participant statements relating to their expectations regarding the actualization of the principles of *beneficence*, *respect for autonomy* and *justice* in the context of AI-assisted

decision-making are categorized below into identified risks and opportunities.

5.2.1 Identified risks

Participant-identified risks regarding the use of AI-assisted technology frequently relate to the principle of *beneficence* and, in particular, associated aspects concerning the nurse–patient relationship. Many participants were concerned that the adoption of such technologies could compromise the promotion of emotional well-being, which is one of the core aims of beneficence. For instance, one participant reasoned that "…the use of the device could lead to patients feeling that the nurse only looks at the screen and no longer talks to them" (G12).

The participants highlighted that the use of AI-assisted technology may negatively impact demands related to the recognition of care recipients' individual needs (i.e., *recognizing needs*). Risks identified in this context mostly addressed nurses' empathy for and awareness of the vulnerability of persons in need of care, both for care recipients in general and for care recipients with impaired communicative abilities, e.g., "[With this technology,] I think the nurse would no longer be as aware of what the person in need of care is expressing in a nonverbal manner" (G7). Similarly, some participants expressed the fear that AI assistance could discourage nurses from exploring care recipients' motives, such as in the event that a care recipient refused certain care procedures, e.g., "…when using such technology, nursing professionals … would tend to reflect less. They would spend less time thinking about what the other person wants" (G9). Moreover, care recipients expressed concern that the use of AI technology would disrupt interpersonal communication with nurses, as nurses might be preoccupied with operating the technology, e.g., "From my point of view, it is more personal and much more pleasant to talk to a nurse who is not simultaneously busy using such technology" (R5).

Mainly with regard to tasks in the context of social care and organization of workflows, individual participants articulated the worry that AI-based decision-support could impair the willingness of nurses to take responsibility for patients' well-being (i.e., *assuming responsibility*). One participant stated, "I think [the nurse] feels validated when using the technology and questions less whether a decision is appropriate" (G10). Relatedly, participants assumed that nurses' experiential knowledge would decrease as a consequence of regularly using such technology. While they reasoned that such may be suitable for providing orientation and confidence in (time) critical situations, they likewise expressed the view that the ability to weigh and balance risks and opportunities could gradually decrease, e.g., "I see a disadvantage in that you would probably tend

to think less independently and instead follow standard procedures" (G15). One care recipient, moreover, raised the concern that particularly inexperienced nurses may no longer learn to independently weigh possible consequences of decisions in situations with moral implications, e.g., "I believe it depends on how long a nurse has been in the profession. A person who hasn't been doing it for very long would certainly be highly influenced by the decision support [of AI technology]. Will that person ever be capable of making such decisions on his or her own?" (R5).

In the context of basic care, participants were also concerned with possible influences on patients' *autonomy*. As shown in the Sect. 5.1.2, many participants perceived that both *relational* and *individual autonomy* could be improved by communication. Relatedly, some individuals expressed discomfort about the possibility that information asymmetries and dependencies would increase if nurses "…don't engage in negotiation with the resident as much" (G1). One nurse explained, "Nurses are in a position of power over people in need of care. In uncertain situations, they enforce what they think is right. I think this disparity could become even greater [with such technology]" (G9).

Finally, several participants noted that the introduction of AI technology could negatively impact the objective of considering individual (subjective) needs when allocating resources (i.e., *need principle*, see the Sect. 5.1.3), e.g., "Since the system is fed by data, [I assume that] in case of doubt, it would recommend caring for the higher number of care recipients regardless of the individual feelings of those in need of care. Very pragmatic" (G12). This perception that in situations involving aspects of distributive justice, the individual situation of those in need of care might disappear from view aside from measurable data, is closely related to identified risks relating to *beneficence*. One person stated that "…such a decision must always be made after weighing all the individual points that play a role in a given situation. …a computer can't grasp how someone feels inside" (G3).

5.2.2 Identified opportunities

In addition to possible risks, the interviewed nurses and care recipients also identified several opportunities arising from the use of AI-assisted technology. Again, considerations primarily focused on *beneficence*. In particular, with regard to basic care tasks participants reasoned that the use of such technology could prevent physical harm. Many participants assumed a positive influence of AI assistance on the empirical basis of decisions made under uncertainty, e.g., "Such applications would certainly provide added value not just by shortening the decision-making process but also, I think, above all ensuring that decisions are more empirically sound" (G12).

While several participants were concerned that the ability to weigh benefits and risks associated with different caring actions could decrease with regular use of AI technology (see the Sect. 5.2.1), some also expressed the hope that the expanded information base would provide assurance and guidance to inexperienced nurses in (time-sensitive) critical situations, e.g., "…I think in situations in which it is important to act quickly, a system like this could be very helpful for new colleagues. Because you really, yes, sometimes you don't know what to do for a moment" (G10). Some participants further assumed that this decision-support could motivate nurses to reconsider their intuitions, e.g., "In order to reflect on your own intuition, I think such a system is actually quite useful. At least, if the various aspects that are important in certain situations are highlighted" (G6).

In addition to the potential support of AI technology in situations requiring nurses to weigh their options to prevent (physical) harm (a key aspect of *assuming responsibility*), one care recipient envisioned that this technology could support a holistic assessment of patients' needs in the first place, e.g., "Nurses are different. Some make little effort to recognize what is going on in a person in need of care. …such technology could, perhaps, identify more precisely where the shoe pinches" (R3).

In the context of social and basic care, several participants identified a further opportunity arising from the expanded information base associated with AI-assisted technology: "I think such technology could provide reassurance to some residents because they can get additional information, sort of like a second opinion" (G13). The participants reasoned that, in this manner, AI technology could promote care recipients' ability to make informed choices and improve their perceived self-determination (i.e., *individual autonomy*), e.g., "It would be good if there was a bit more transparency in the interaction between the nursing staff and the residents. With such technology, some residents would probably be more likely to be convinced because they would see that the information referred to was not made up but documented" (G13).

Another positive aspect mentioned by participants was a potential benefit regarding a fair(er) distribution of resources. Referring to tasks related to organizing workflows, participants noted that the adoption of AI technology may provide a more objective basis for workflow prioritization (when the technology is informed by patient needs), e.g., "Such systems can have a positive effect. Because with them, I think, you are less driven by emotions but more objective, that is, really guided to what is needed" (G5). Relatedly, participants expressed the hope that, depending on the system design, the technology could strengthen the concept of *nondiscrimination* (see the Sect. 5.1.3); in other words, that resources could be distributed independently of

the visibility of individual care recipients and instead guided by their need for care.

Overall, participants mainly perceived advantages in the adoption of AI technology for expanding and increasing the objectivity of nursing professionals' (information) bases for clinical decision-making.

6 Discussion

To complement ethical principles considered relevant for the design of AI-assisted technology in health care with a context-specific conceptualization of the principles of individuals potentially affected by the implementation of AI-assisted technology, we first investigated stakeholders' contextualized perspectives on three principles: *beneficence*, *respect for autonomy* and *justice* (Q1 and Q2). Building upon this analysis, we investigated participant expectations regarding the actualization of their concepts of the principles in the context of AI-assisted decision-making. Thus, we provided initial indications regarding which principles ought to receive particular attention when designing AI technologies for nursing care.

Our analysis of participant reasoning in situations involving moral decision-making that occur in everyday nursing practice indicates that nurse and care recipient perspectives are largely compatible with the principles of *beneficence*, *respect for autonomy* and *justice*. Thus, these three *principles of biomedical ethics* are applicable to the field of nursing care and are a suitable launching point to explain and categorize nurse and care recipient beliefs and reasoning in situations involving moral decision-making in the fields of basic care, social care and organization of workflows (Q1).

Moreover, these results demonstrate that a qualitative analysis of stakeholder reflections on ethical principles based on scenarios depicting care tasks associated with moral decision-making (Q2) provide a more nuanced understanding (i.e., context-specific conceptualization) of the principles as well as their actualization through situational factors and, in particular, demands.

The results confirm that the principles' definitions need to be specified for as well as adapted to care-specific requirements. Participant concepts of *beneficence* were largely consistent with the definition of Beauchamp and Childress [41]; particularly, participants highlighted the demands to recognize care recipients' needs and to assume responsibility for the identified needs. With regard to *respect for autonomy*, many participants noted that autonomy may require that care recipients are free from controlling influences and/or that (capacities for) autonomous choice are promoted [ibid., p. 105] (the concept of *individual autonomy*). Other participants argued though that patient autonomy can also be ensured by preserving a person's sense of identity as well

as utilizing *shared* decision-making (the concept of *relational autonomy*). Caregiver and care recipient concepts of *justice* were, again, broadly in line with the definition of Beauchamp and Childress. Many participants referred to "the obligation to fairly distribute benefits, risks and costs under conditions of scarce resources" [ibid., pp. 13, 250]. Our analysis additionally suggests that the participants hold different views on what constitutes a *fair* distribution. While some advocated an equal allocation of resources to each care recipient (the *equality principle*), others argued for an allocation of resources based on individual needs for basic care and/or social support (the *need principle*).

Hence, our analysis indicates that a stakeholder-oriented specification of the principles allows for, or even requires, integration of specific theories of healthcare and nursing. Notably, statements relating to demands perceived as critical to the actualization of *beneficence* closely corresponded to Tronto's assumption that beneficent care should be regarded as a dynamic process and, in particular, assessed along different phases [51]. Moreover, participant concepts of *respect for autonomy* referring to it as a relational process closely relate to feminist reconceptualizations of autonomy (e.g., [64, 65]). Such accounts highlight the importance of interpersonal or social conditions and a person's sense of identity for the realization of autonomy, which therefore contrast with an individualistic interpretation of autonomy.

With regard to Q3, our analysis showed that the participants anticipated risks as well as opportunities relating to the actualization of their concepts of all three principles, and especially their concepts of *beneficence*, in the context of AI-assisted decision-making. In particular, care recipients reasoned that the use of AI-assisted technology could disrupt interpersonal relations as well as communication (the concept of *recognizing needs*) [15]. Both groups assumed that there would be a negative impact on nurses' experiential knowledge (the concept of *assuming responsibility*) [ibid.] and that the technology could discourage nurses from exploring care recipients' motives. On the other hand, participants envisioned that such technology, particularly with regard to basic care tasks, could prevent physical harm, e.g., by providing evidence-based health information for decision-making in uncertain conditions [14] and by motivating nurses to reconsider their intuitions (the concept of *assuming responsibility*).

Possible influences on the realization of *respect for autonomy* mainly included two aspects: On the one hand, an increase in information asymmetry is considered to reduce care recipient autonomy, but, on the other hand, increases in information to promote care recipients' ability to make informed decisions, thereby strengthening their autonomy (the concept of *individual autonomy*). Moreover, participants expected that adopting AI technology in tasks related to organizing workflows could negatively

impact the consideration of individual (subjective) needs when distributing resources (the *need principle*); however, adoption of this technology may improve the distribution of resources independent of the visibility of individual care recipients (the concept of *nondiscrimination*).

In conclusion, our study generated prospective understanding of how AI-assisted technologies might modify social structures and practices as well as existing asymmetries within care contexts. Participants reasoned that such technologies may improve and augment nurse abilities, assist in the identification of novel solutions to well-known problems such as discrimination, and help to coordinate complexity (e.g., within tasks that demand situational weighing). However, at the same time, participants warned that AI technology carries the inherent risk of unintended side effects, such as an objectification and rationalization of the nurse–care recipient relationship.

6.1 Implications for future research

The study results underscore the importance of a context-specific conceptualization of ethical principles relevant for AI-assisted decision-making to address the current epistemic uncertainty regarding the risks and opportunities associated with the (non)fulfillment of ethical principles. Moreover, existing guidelines not only appear too vague to guide the design of technologies based on ethical principles but they are also blind to stakeholders' individual needs and interests. To ensure that ethical guidelines for AI assistance are sensitive to the interests and needs of stakeholders, AI technology guidelines should be determined within specific contexts. Linked to this, we recommend also future studies to consider both nurse and care recipient perspectives when generating bottom-up knowledge regarding the actualization of ethical principles in the context of AI-assisted decision-making. While considering ethical requirements within situations involving moral decision-making falls within the responsibility of nurses, their fulfillment also needs to be assessed by care recipients.

In addition, future studies might need to assess in greater detail how the implementation of AI-assisted technology may alter nurse tasks and impact their perceived moral coercion. The use of digital care services can be associated with moral distress (e.g., [66]), i.e., the experience of not being able to act according to personal and professional values [67], frequently reported by nurses [68]. However, to date, no studies have focused on the influences of AI-based systems.

Our study, moreover, suggests that the ethical principles of *beneficence*, *respect for autonomy* and *justice* provide suitable guidance for the development of care-specific indicators that can help to align AI-assisted technologies

(in the field of nursing) with stakeholders' moral interests. To specify such indications with regard to more concrete design considerations for AI systems and according relevant instrumental principles (such as *explicability* [42]), a prerequisite is to integrate interdisciplinary and transdisciplinary perspectives (e.g., from the social sciences, computer science and occupational sciences) to provide a (rich)er understanding of the coconstruction of technological and social phenomena (see also, e.g., [69, 70]). In addition, the specification of the ethical assessment of using AI-assisted technologies for care (e.g., by methods from the field of technology assessment [71]) requires broader knowledge of the technological possibilities of specific AI-assisted applications.

Further research is also needed to determine if stakeholders' reflections on moral decision-making situations associated with different bioethical principles (such as 'integrity', 'autonomy', 'vulnerability' or 'dignity', as proposed by Rendtoff [47] or Häyry [48] as specific European principles) can broaden the set of ethical principles considered relevant for the design of AI. Similarly, different analytic methods such as the grounded theory may help to identify further ethical principles relevant for the design of AI-assisted technology in nursing care.

Ultimately, it may be necessary to develop innovative system design approaches that enable the integration of ethical principles during an iterative process throughout technologies' entire lifecycle. Traditional engineering processes and current risk analysis methods do not allow a continuous assessment of possible risks, i.e., there is no open feedback loop between operators and system designers. As many algorithms underlying AI technologies are able to adapt to their environment (and given the black-box nature of frequently applied deep learning models), it would be useful if information on the extent to which technologies already in use affect social structures were available during the design process [72].

6.2 Study limitations

The results of this study must be interpreted in light of some limitations. First, while there is ample reason to *prospectively* deliberate on the potential consequences of emerging technologies, individuals who are unfamiliar with such technologies may have a limited understanding of the technologies' abilities and their impacts on everyday (professional) life. In our study, this is particularly likely in the care-recipient group. Second, although our scenarios were designed to be comparable to real-life situations, the addition of different context-specific information could result in different principle-related statements. Third, we decided not to include a scenario prompting reflection on the principle of *nonmaleficence* because we aimed to respond to the

(potential) vulnerability of participants in the care-recipient group. However, future studies could explore nurses' moral reasoning regarding *nonmaleficence*.

7 Conclusion

Artificial intelligence (AI)-assisted technologies may exert a profound impact on social structures and practices in health care contexts. Our study helped to translate ethical principles considered relevant for the design of AI-assisted technology in health care into practice. In particular, our analysis provides a context-specific conceptualization as well as adaptation of the well-established *principles of biomedical ethics* in the context of long-term care and, building upon this, generates bottom-up knowledge regarding the actualization of the ethical principles in AI-assisted decision-making in care contexts. Thus, we provided initial indications regarding which concepts of the investigated ethical principles ought to receive extra attention when designing AI technologies to ensure that these technologies are not blind to the moral interests of stakeholders in the care sector.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

Consent to participate Informed consent, including consent to publish the results in an anonymized form, was obtained from all individual participants included in the study.

Ethics approval The study was approved by the Data Protection Office and Ethics Committee of the German Federal Institute for Occupational Safety and Health. All procedures in this study involving human participants were performed in accordance with the 1964 Declaration of Helsinki and its later amendments. All study participants provided written informed consent. Data management and analysis were consistent with the European Union's General Data Protection Regulation.

References

1. Organisation for Economic Co-Operation and Development: Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 (2019). Accessed 10 Nov 2022
2. Parker, S.K., Grote, G.: Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Appl. Psychol. **71**(4), 1171–1204 (2022). https://doi.org/10.1111/apps.12241
3. Bicchieri, C., Muldoon, R., Sontuoso, A.: Social norms. In: Zalta, E. (ed.) The Stanford Encyclopedia of Philosophy (Winter 2018 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/social-norms (2018). Accessed 15 Nov 2022
4. Brey, P.: Values in technology and disclosive computer ethics. In: Floridi, L. (ed.) The Cambridge Handbook of Information and Computer Ethics, pp. 41–58. Cambridge University Press, Cambridge (2009)
5. Buchanan, C., Howitt, M.L., Wilson, R., Rooth, R.G., Risling, T., Bambford, M.: Predicted influences of artificial intelligence on the domains of nursing: scoping review. JMIR Nurs. (2020). https://doi.org/10.2196/23939
6. Seibert, K., Domhoff, D., Bruch, D., Schulte-Althoff, M., Fürstenau, D., Biessmann, F., Wolf-Ostermann, K.: Application scenarios for artificial intelligence in nursing care: rapid review. J. Med. Internet Res. (2021). https://doi.org/10.2196/26522
7. von Gerich, H., Moen, H., Block, L.J., Chu, C.H., DeForest, H., Hobensack, M., Michalowski, M., Mitchell, J., Nibber, R., Olalia, M.A., Pruinelli, L., Ronquillo, C.E., Topaz, M., Peltonen, L.M.: Artificial intelligence-based technologies in nursing: a scoping literature review of the evidence. Int. J. Nurs. Stud. (2021). https://doi.org/10.1016/j.ijnurstu.2021.104153
8. World Health Organization: State of the world's nursing 2020: investing in education, jobs and leadership: WHO guidance. https://www.who.int/publications/i/item/9789240003279 (2020). Accessed 10 Nov 2022
9. Kawamoto, K., Houlihan, C.A., Balas, E.A., Lobach, D.F.: Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Br. Med. J. (2005). https://doi.org/10.1136/bmj.38398.500764.8F
10. Ng, Z.Q.P., Ling, L.Y.J., Chew, H.S.J., Lau, Y.: The role of artificial intelligence in enhancing clinical nursing care: a scoping review. J. Nurs. Manag. **30**(8), 3654–3674 (2021). https://doi.org/10.1111/jonm.13425
11. International Council of Nurses: Nursing definitions. https://www.icn.ch/nursing-policy/nursing-definitions. Accessed 30 Nov 2022
12. Rainer, J., Schneider, J.K., Lorenz, R.A.: Ethical dilemmas in nursing: an integrative review. J. Clin. Nurs. **27**(19–20), 3446–3461 (2018). https://doi.org/10.1111/jocn.14542
13. Suhonen, R., Stolt, M., Habermann, M., Hjaltadottir, I., Vryonides, S., Tonnessen, S., Halvorsen, K., Harvey, C., Toffoli, L., Scott, P.A.: Ethical elements in priority setting in nursing care: a scoping review. Int. J. Nurs. Stud. **88**, 25–42 (2018). https://doi.org/10.1016/j.ijnurstu.2018.08.006

14. Morley, J., Machado, C.C.V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., Floridi, L.: The ethics of AI in health care: a mapping review. Soc. Sci. Med. (2020). https://doi.org/10.1016/j.socscimed.2020.113172

15. Rubeis, G.: The disruptive power of artificial intelligence. Ethical aspects of gerontechnology in elderly care. Arch. Gerontol. Geriatr. (2020). https://doi.org/10.1016/j.archger.2020.104186

16. Rogers, W.A., Draper, H., Carter, S.M.: Evaluation of artificial intelligence clinical applications: detailed case analyses show value of healthcare ethics approach in identifying patient care issues. Bioethics 35(00), 623–633 (2021). https://doi.org/10.1111/bioe.12885

17. Mittelstadt, B., Floridi, L.: The ethics of big data: current and foreseeable issues in biomedical contexts. Sci. Eng. Ethics 22, 303–341 (2016). https://doi.org/10.1007/s11948-015-9652-2

18. Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., Floridi, L.: The ethics of algorithms: key problems and solutions. AI Soc. 37, 215–230 (2021). https://doi.org/10.1007/s00146-021-01154-8

19. Beil, M., Proft, I., van Heerden, D., Sviri, S., van Heerden, P.V.: Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med. Exp. (2019). https://doi.org/10.1186/s40635-019-0286-6

20. Siala, H., Wang, Y.: SHIFTing artificial intelligence to be responsible in healthcare: a systematic review. Soc. Sci. Med. (2022). https://doi.org/10.1016/j.socscimed.2022.114782

21. Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., Floridi, L.: Operationalising AI ethics: barriers, enablers and next steps. AI Soc. 38, 411–423 (2021). https://doi.org/10.1007/s00146-021-01308-8

22. European Commission: Proposal for a Regulation of the European Parliament and the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. COM/2021/206 final. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence (2021). Accessed 10 Jan 2023

23. Organisation for Economic Co-Operation and Development: Scoping of the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO). OECD Digital Economy Papers, No. 291. https://www.oecd-ilibrary.org/science-and-technology/scoping-the-oecd-ai-principles_d62f618a-en (2019). Accessed 10 Nov 2022

24. Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Sánchez, A.N., Raji, D., Rankin, J.L., Richardson, R., Schultz, J., West, S.M., Whittaker, M.: AI Now 2019 Report. AI Now Institute, New York. https://ainowinstitute.org/AI_Now_2019_Report.pdf (2019). Accessed 27 Oct 2022

25. Bostrom, N., Yudkowsky, E.: The ethics of artificial intelligence. In: Frankish, K., Ramsey, W.M. (eds.) The Cambridge Handbook of Artificial Intelligence, pp. 316–334. Cambridge University Press, Cambridge (2014)

26. O'Neil, C.: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, 1st edn. Crown Publishers, New York (2016)

27. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. 1, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2

28. Prem, E.: From ethical AI frameworks to tools: a review of approaches. AI Ethics (2023). https://doi.org/10.1007/s43681-023-00258-9

29. Vakkuri, V., Kemell, K.K., Kultanen, J., Abrahamsson, P.: The current state of industrial practice in artificial intelligence ethics. IEEE Softw. 37(4), 50–57 (2020). https://doi.org/10.1109/MS.2020.2985621

30. Le Dantec, C.A., Poole, E.S., Wyche, S.P.: Values as lived experience: evolving value sensitive design in support of value discovery. In: CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1141–1150 (2009). https://doi.org/10.1145/1518701.1518875

31. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. 1, 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4

32. Lukkien, D.R.M., Nap, H.H., Buimer, H.P., Peine, A., Boon, W.P.C., Ket, J.C.F., Minkman, M.M.N., Moors, E.H.M.: Toward responsible artificial intelligence in long-term care: a scoping review on practical approaches. Gerontologist 63(1), 155–168 (2021). https://doi.org/10.1093/geront/gnab180

33. Richardson, H.S.: Moral Reasoning. In: Zalta, E. (ed.), The Stanford Encyclopedia of Philosophy (Fall 2018 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral (2018). Accessed 15 Nov 2022

34. Independent High-level Expert Group on AI: Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 (2019). Accessed 10 Nov 2022

35. World Health Organization: Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. World Health Organization, Geneva (2021)

36. Pommeranz, A., Detweiler, C., Wiggers, P., Jonker, C.: Elicitation of situated values: need for tools to help stakeholders and designers to reflect and communicate. Ethics Inf. Technol. 14(4), 285–303 (2012). https://doi.org/10.1007/s10676-011-9282-6

37. van Wynsberghe, A.: Designing robots for care: care centered value-sensitive design. Sci. Eng. Ethics 19, 407–433 (2013). https://doi.org/10.1007/s11948-011-9343-6

38. Friedman, B., Hendry, D.: Value Sensitive Design: Shaping Technology with Moral Imagination. The MIT Press, Cambridge (2019)

39. Robertson, T., Simonsen, J.: Participatory design: an introduction. In: Simonsen, J., Robertson, T. (eds.) Routledge International Handbook of Participatory Design, vol. 711, pp. 1–17. Routledge, New York (2013)

40. Manders-Huits, N.: What values in design? The challenge of incorporating moral values into design. Sci. Eng. Ethics 17, 271–287 (2011). https://doi.org/10.1007/s11948-010-9198-2

41. Beauchamp, T.L., Childress, J.: Principles of Biomedical Ethics, 8th edn. Oxford University Press Inc, New York (2019)

42. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., Vayena, E.: AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 28(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5

43. Kangasniemi, M., Pakkanen, P., Korhonen, A.: Professional ethics in nursing: an integrative review. J. Adv. Nurs. 71(8), 1744–1757 (2015). https://doi.org/10.1111/jan.12619

44. American Nurses Association: Code of Ethics for Nurses with Interpretive Statements. American Nurses Association, Silver Spring (2015)

45. International Council of Nurses: The ICN Code of Ethics for Nurses: Revised 2021. https://www.icn.ch/system/files/2021-10/ICN_Code-of-Ethics_EN_Web_0.pdf (2021) Accessed 10 Oct 2022

46. Veatch, R.M.: Reconciling lists of principles in bioethics. J. Med. Philos. 45(4–5), 540–559 (2020). https://doi.org/10.1093/jmp/jhaa017

47. Rendtorff, J.D.: Basic ethical principles in European bioethics and biolaw: autonomy, dignity, integrity and vulnerability—towards a foundation of bioethics and biolaw. Med. Health Care Philos. 5, 235–244 (2002). https://doi.org/10.1023/A:1021132602330

48. Häyry, M.: European values in bioethics: why, what, and how to be used. Theor. Med. Bioeth. **24**, 199–214 (2003). https://doi.org/10.1023/A:1024814710487

49. Gilligan, C.: In a Different Voice: Psychological Theory and Women's Development. Harvard University Press, Cambridge (1982)

50. Noddings, N.: Caring: A Feminine Approach to Ethics and Moral Education. University of California Press, Berkeley (1984)

51. Tronto, J.: Moral Boundaries: A Political Argument for an Ethic of Care. Routledge, New York (1993)

52. Ridge, M., McKeever, S.: Moral Particularism and Moral Generalism. In: Zalta, E. (ed.) The Stanford Encyclopedia of Philosophy (Winter 2020 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/moral-particularism-generalism. Accessed 10 Oct 2022

53. Deutsche Gesellschaft für Pflegewissenschaft e. V.: Ethikkodex Pflegeforschung der Deutschen Gesellschaft für Pflegewissenschaft. https://dg-pflegewissenschaft.de/wp-content/uploads/2017/05/Ethikkodex-Pflegeforschung-DGP-Logo-2017-05-25.pdf (2016). Accessed 10 June 2023

54. Council for International Organizations of Medical Sciences: International ethical guidelines for health-related research involving humans. https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf (2016). Accessed 10 June 2023

55. Hébert, P.C., Meslin, E.M., Dunn, E.V.: Measuring the ethical sensitivity of medical students: a study at the University of Toronto. J. Med. Ethics **18**, 142–147 (1992). https://doi.org/10.1136/jme.18.3.142

56. Wang, W., Chen, L., Xiong, M., Wang, Y.: Accelerating AI adoption with responsible AI signals and employee engagement mechanisms in health care. Inf. Syst. Front (2021). https://doi.org/10.1007/s10796-021-10154-4

57. Page, K.: The four principles: can they be measured and do they predict ethical decision making? BMC Med. Ethics (2012). https://doi.org/10.1186/1472-6939-13-10

58. Archibald, M.M., Barnard, A.: Futurism in nursing: technology, robotics and the fundamentals of care. J. Clin. Nurs. **27**(11–12), 2473–2480 (2018). https://doi.org/10.1111/jocn.14081

59. Carroll, J.: Scenario-Based Design: Envisioning Work and Technology in System Development. Wiley, New York (1995)

60. Nathan, L.P., Klasnja, P.V., Friedman, B.: Value scenarios: a technique for envisioning systemic effects of new technologies. In: CHI EA '07: Extended Abstracts on Human Factors in Computing Systems, pp. 2585–2590 (2007). https://doi.org/10.1145/1240866.1241046

61. Rand, D.G.: Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. Psychol. Sci. **27**(9), 1192–1206 (2016). https://doi.org/10.1177/0956797616654455

62. Kuckartz, U.: Qualitative Text Analysis: A Guide to Methods, Practice & Using Software. SAGE Publications Ltd (2014)

63. VERBI: Maxqda Standard 2020 Network. https://www.maxqda.com/ (2020). Accessed 20 Sept 2022

64. Donchin, A.: Understanding autonomy relationally: toward a reconfiguration of bioethical principles. J. Med. Philos. **26**(4), 365–386 (2001). https://doi.org/10.1076/jmep.26.4.365.3012

65. Stoljar, N.: Informed consent and relational conceptions of autonomy. J. Med. Philos. **36**(4), 375–384 (2011). https://doi.org/10.1093/jmp/jhr029

66. Frennert, S.: Moral distress and ethical decision-making of elder-care professionals involved in digital service transformation. Disabil. Rehabil. Assist. Technol. **18**(2), 156–165 (2020). https://doi.org/10.1080/17483107.2020.1839579

67. Morley, G., Ives, J., Bradbury-Jones, C., Irvine, F.: What is 'moral distress'? A narrative synthesis of the literature. Nurs. Ethics **26**(3), 646–662 (2019). https://doi.org/10.1177/0969733017724354

68. Oh, Y., Gastmans, C.: Moral distress experienced by nurses: a quantitative literature review. Nurs. Ethics **22**(1), 15–31 (2015). https://doi.org/10.1177/0969733013502803

69. Beimborn, M., Kadi, S., Köberer, N., Mühleck, M., Spindler, M.: Focusing on the human: interdisciplinary reflections on ageing and technology. In: Domínguez-Rué, E., Nierling, L. (eds.) Ageing and Technology: Perspectives from the Social Sciences, pp. 311–333. Transcript Verlag, Bielefeld (2016)

70. Goirand, M., Austin, E., Clay-Williams, R.: Implementing ethics in healthcare AI-based applications: a scoping review. Sci. Eng. Ethics (2021). https://doi.org/10.1007/s11948-021-00336-3

71. Grunwald, A.: Technology Assessment in Practice and Theory, 1st edn. Routledge, Oxford (2019)

72. Schlicht, L., Melzer, M., Rösler, U., Voß, S., Vock, S.: An integrative and transdisciplinary approach for a human-centered design of AI-based work systems. In: Proceedings of the ASME 2021 International Mechanical Engineering Congress and Exposition. Volume 13: Safety Engineering, Risk, and Reliability Analysis. The American Society of Mechanical Engineers, New York, Art 71261 (2021). https://doi.org/10.1115/IMECE2021-71261