Mario Kropf

When Spiritual Robots Fail

AI-based robots, spirituality, and moral responsibility for bad outcomes

ENGLISH

ABSTRACTS

The use of robots for spiritual purposes represents a novel and still little-researched field within technology ethics and debates on moral responsibility. This article examines how moral responsibility should be assessed when AI-based robots contribute to spiritually significant bad outcomes. It argues that spirituality introduces a distinct normative dimension into responsibility attribution that has so far been largely neglected in debates on AI ethics. The analysis first conceptualizes spirituality as an individual orientation expressed in the search for meaning, transcendence, personal growth, or connection to others, nature or the sacred. It then examines current forms of spiritual robots—including SanTO, BlessU2, NAO, and Xian'er—and their roles in addressing spiritual concerns. Finally, three illustrative scenarios of spiritually significantly bad outcomes are analyzed: a lack of guidance (*harm*), too much guidance (*manipulation*), and too little guidance (*loss of trust*). Based on these cases, the analysis supports a differentiated account that combines a backward-looking responsibility, where conditions such as control, knowledge, and intention are met, with a forward-looking collective responsibility for the design and governance of spiritual robots.

DEUTSCH

Wenn spirituelle Roboter versagen. KI-basierte Roboter, Spiritualität und Moralische Verantwortung für negative Folgen

Der Einsatz von Robotern zu spirituellen Zwecken stellt ein neuartiges und bislang noch wenig erforschtes Feld innerhalb der Technikethik und der Debatten über moralische Verantwortung dar. Dieser Artikel untersucht, wie moralische

Verantwortung zu bewerten ist, wenn KI-basierte Roboter zu spirituell bedeutsamen negativen Folgen beitragen. Es wird argumentiert, dass Spiritualität eine eigenständige normative Dimension für die Zuschreibung von Verantwortung einführt, die in Debatten über KI-Ethik bislang weitgehend vernachlässigt wurde. Die Analyse konzeptualisiert Spiritualität zunächst als eine individuelle Orientierung, die sich in der Suche nach Sinn, Transzendenz, persönlichem Wachstum oder der Verbindung zu anderen, zur Natur oder zum Heiligen ausdrückt. Anschließend werden aktuelle Formen spiritueller Roboter – darunter SanTO, BlessU2, NAO und Xian'er – sowie ihre Rolle bei der Auseinandersetzung mit spirituellen Anliegen untersucht. Schließlich werden drei illustrative Szenarien spirituell bedeutsamer negativer Folgen analysiert: ein Mangel an Orientierung (Schaden), zu viel Orientierung (Manipulation) und zu wenig Orientierung (Vertrauensverlust). Auf der Grundlage dieser Fälle stützt die Analyse eine differenzierte Darstellung, die eine rückwärtsgerichtete Verantwortung, bei der Bedingungen wie Kontrolle, Wissen und Intention erfüllt sind, mit einer vorwärtsgerichteten kollektiven Verantwortung für die Gestaltung und Regulierung spiritueller Roboter verbindet.

| KEY WORDS

spirituality; spiritual robots; moral responsibility; bad outcomes; spiritual harm; ethics

Spiritualität; spirituelle Roboter; moralische Verantwortung; negative Folgen; spiritueller Schaden; Ethik

| BIOGRAPHY

Dr. Mario Kropf is University assistant at the Department of Moral Theology and the Department of Ethics and Social Studies at the University of Graz.
ORCID 0000-0002-0645-3276
E-Mail: mario.kropf(at)uni-graz.at

1 Introduction

AI-based robots are increasingly deployed not only in care facilities and hospitals but also in religious and spiritual contexts, where they engage with existential issues. Their use raises familiar ethical concerns, including autonomy, privacy, harm, and moral responsibility. Among these concerns, the question of moral responsibility is particularly pressing. While responsibility attribution for AI systems has been extensively discussed (cf. Hakli/Mäkelä 2019; Nyholm 2018; Tigard 2021; Coeckelbergh 2020), considerably less attention has been paid to spiritually relevant bad outcomes in contexts where robots are deployed. This article examines how moral responsibility should be assessed when AI-based spiritual robots contribute to spiritually significant bad outcomes, despite not being moral agents and remaining only partially predictable. It argues that spirituality—understood as an individual search for meaning, transcendence, and orientation—introduces a distinct normative dimension into responsibility attribution. Spiritually relevant bad outcomes are not limited to technical malfunction or physical harm; they cause harm to orientation, autonomy, and trust. While existing research on religious and spiritual robots has focused on their communicative functions, ritual performances, and socio-cultural implications (cf. Cheong 2020; Puzio 2023; Löffler et al. 2021; Geraci 2024), they rarely analyze spiritually relevant bad outcomes through the lens of normative responsibility theory. Additionally, the moral responsibility of AI systems has been widely discussed in terms of responsibility gaps, collective responsibility, and artificial moral advisors (cf. Matthias 2004; Kiener 2025; Königs 2022), without systematically considering spirituality as a distinct normative domain.

Spirituality as a distinct normative domain

So far, these discussions have largely proceeded independently. By combining a conceptually grounded account of spirituality with a differentiated analysis of responsibility conditions in embodied AI systems, this article advances a more fine-grained framework for evaluating spiritually significant harms. It proposes a differentiated account that combines a backward-looking responsibility—where control, knowledge, and intention are present—with a forward-looking, collectively assumed responsibility for the design, deployment, and governance of spiritual robots. Methodologically, the article adopts a conceptual-analytical and

normative approach. It clarifies key concepts such as spirituality, bad outcomes, and moral responsibility, and investigates responsibility attributions through three illustrative scenarios. These thought experiments are not empirical case studies but serve to isolate morally relevant features and assess the plausibility of different responsibility claims. The argument is built in three steps: *firstly*, spirituality is analyzed as an individual point of orientation; *secondly*, current forms of spiritual robots and their functional limitations are examined; *thirdly*, three structurally distinct types of spiritually relevant bad outcomes are discussed in order to assess responsibility attribution.

2 The relevance of spirituality

The following section emphasizes the importance of spirituality as a point of orientation. Spirituality or a spiritual way of life implies numerous ideas that do not necessarily have to include religious aspects. While *being religious* usually means that one belongs to a specific belief system, tradition, or doctrine, and thus also subscribes to a set of values, *being spiritual* is mostly an individual matter. In the latter case, it is often about self-transcendence (cf. Kristeller/Jordan 2018, 132), a search for meaning (cf. Sun et al. 2016, 1454), overcoming personal crises (cf. Benites et al. 2021, 9–10), a connection with nature or fellow human beings (cf. King et al. 2024, 1464), a moral learning process (cf. Cole-Turner 2025, 12), a higher power (cf. McKee/Chappel 1992, 202), or an orientation toward values (cf. Körtner 2009, 9). The definition of the European Association of Palliative Care (EAPC) provides a point of reference:

> *"Spirituality is the dynamic dimension of human life that relates to the way persons (individual and community) experience, express and/or seek meaning, purpose and transcendence, and the way they connect to the moment, to self, to others, to nature, to the significant and/or the sacred"* (European Association of Palliative Care 2025).

With regard to this definition, one might ask what exactly "being spiritual" means, or, in other words, what is not a spiritual activity. The answer to this question does not seem to be universal, but rather individual. A trivial but, in my opinion, fitting answer to the question of what one seeks or wants to experience through spiritual practices is: many things, but not everything—the limitations of the EAPC make this clear. Let's imagine two

people, Ann and Bert, sitting on a park bench watching birds fly by. If Ann watches the birds and counts how many fly by per hour, this is not necessarily a spiritual activity. If, on the other hand, Bert engages with the flight of the birds and, in the course of observation, realizes that time flies just as the birds do, this experience can be described as spiritual; especially if Bert consequently appreciates the meaning of time in life more and recognizes this experience as a connection to nature and spiritual insight. In short, spiritual experiences are individual and can be initiated, but they do not have to be. A subjective approach to spirituality seems appropriate in consideration of lived reality and is also ethically necessary in order to avoid any prescriptive or generalizing—and therefore restrictive—definitions. What is a spiritual activity for Bert, and thus brings meaning, joy or insight, may be nothing more than an unnecessary waste of time for Ann. Personal views of being spiritual cannot be generalized (cf. Schaupp 2017, 285–286), and a universal definition of spirituality would constitute an external prescription (cf. Birkholz 2020, 7—8).

Meaning that cannot be explained rationally

But why is spirituality relevant in this article? In short, because it can provide orientation. People often ask themselves questions such as: "Why is this happening to me?" (Q1) or "What should I do now?" (Q2). Even though it may seem that spiritual practices are associated with many, if not all, possible things, the following can be said: Spiritual practices and experiences serve (1) as points of orientation for many people to overcome certain challenges (cf. Puchalski 2001, 354). These challenges are not necessarily negative, to be seen as obstacles or burdens—they can also be understood in a positive way. A family caregiver emphasizes the personal development process:

> *"It is a life changing experience but I believe things happen in life for a reason and there is always something to be learned from what we are going through"* (Damianakis et al. 2018, 73).

Even though the work can be stressful for family caregivers, it can be concluded from the statement that the person can take away certain points of orientation from this activity. This experience therefore appears to be spiritual because it reveals a meaning that cannot usually be explained rationally or even universally. Regarding Q1 and Q2, individuals

find subjectively satisfactory answers to questions that cannot be answered objectively.

Subjective answers to general questions

Spiritual moments arise—in many cases —(2) not just by chance but require that an individual is open to them. However, what exactly one is open to can vary greatly and should not be prescribed. It is about an openness to spiritual experiences, not about constantly searching for the spiritual. The statement of a nurse illustrates this subjectivity: "[…]I think that religious group participation offers a sense of togetherness […]" (Camacho-Montaño et al. 2021, 7). This does not mean that participation in groups cannot contribute to spiritual experiences, but only that it depends on the individual whether and what exactly such activities provide. Further study results confirm this impression: "[…] a music therapy session, can offer people a chance to be themselves again. […] they can access a certain memory […]" (Connolly/Moss 2021, 377). While the spiritual relevance in the first case is expressed in a sense of belonging, in the second case it consists in people being able to find themselves (again). This orientation does not mean that spiritual experiences necessarily provide answers to questions (e. g., Q1 or Q2) that are general—and often unanswerable. In many cases, spiritual practice seems to help people focus on something that may be even more important than answering those questions. In short, one focuses on what is important to oneself and thereby finds subjective answers to general questions. These subjective answers are then (1) a reason for providing and promoting spiritual practices, and (2) one—if not *the*—reason why spirituality offers guidance.

To facilitate this orientation, there are (3) not only specially trained individuals but also appropriate programs and tools[1] that allow for a certain degree of objectivity without disregarding the spiritual subject. In short, "[…] beliefs should be respected, even when not shared" (Best et al. 2014, 1336). This is relevant for the following considerations, *firstly* because one should not infer the preferences of others' beliefs from one's own preferences. And *secondly*, because in this article, this counterpart (a robot) is nevertheless relevant to spiritual concerns. Even if many people consider spirituality to be something valuable, which provides points of orientation *for them*, there will be others for whom this is not the case. An open and impartial attitude is therefore necessary not only on the part of nursing staff, pastoral caregivers, physicians, and other professionals, but also, to some

1 The FICA tool can be used to identify patients' spiritual needs through conversation. Faith - Importance (of spirituality) - Community - Action in care. The HOPE questions can be used to inquire about spiritual needs in a medical context. Hope - Organized religion - spiritual Practices - Effects. The study by Toivonen et al. drew on close ones, friends and relatives, literary documents and other items to identify spiritual needs even when those affected were no longer able to communicate them themselves. Many others also emphasize the importance of identifying individual spiritual beliefs: The School of Medicine and Health Sciences 2022; Toivonen et al. 2018; Puchalski 2001; Löffler et al. 2021; Johnstone et al. 2012; Cole-Turner 2025.

extent, on the part of spiritual robots. This human attitude is then to be understood as a technical and programmed function of these machines, but it does not diminish the relevance of such an orientation and attention to the spiritual subject—as the following passages will show.

3 AI-based robots and spirituality

This section presents current models of spiritual robots and their possible applications, linking them to the relevance of spirituality. Compared to AI-based care robots, which have been used in the care sector for some time (cf. Vandemeulebroucke et al. 2021; Kabacińska et al. 2021; Robaczewski et al. 2021), the use of robots for spiritual purposes has been comparatively little researched—empirical studies are therefore limited. For this reason, and because people need support at the end of their lives, during times of serious misfortune or personal crises, this approach seems important and timely. A spiritual AI-based robot[2] can be seen as an artificial actor that is physically present and, due to its technical skills, is able to address people's spiritual concerns in a technical form and with limitations (cf. Soljacic et al. 2024, 692; Simmerlein/Tretter 2024a, 258; Coeckelbergh 2012, 57). This could involve initiating a conversation, asking specific questions, presenting text passages, verses and videos, simply being present, or listening and providing spiritual guidance (cf. Geraci 2006, 236; Puzio 2024, 130–131; Cheong 2022, 89).

The following machines[3] were selected because they have different skills and designs, allowing for a differentiated analysis. Xian'er is a robot monk that is mainly employed in Asian contexts and is considered funny, interesting, and useful by many people (cf. Simmerlein/Tretter 2024b, 3; Cheong 2020, 420). This robot can move, quote religious texts, impart Buddhist wisdom, interact with people, and learn from its environment (cf. Puzio 2023, 4). SanTO can be regarded as a theomorphic (i. e., designed in the form of a deity or sacred figure) robotic statue that quotes religious texts and encourages people to pray, but cannot move and has limited interactive capabilities (cf. Trovato et al. 2021, 544–545). BlessU2 is limited in its movement like SanTO, but has robotic arms and a *humanoid*-like design (cf. Löffler et al. 2021, 574–575). It can recite passages from the Bible, bless people, and address different target groups in terms of language and dialect. In contrast, NAO and Pepper are already used in the care sector, allow interaction, and can move. For spiritual purposes, they conduct funerals,

2 I use "robot" to refer to AI-based, physically embodied systems. The term "machine" is employed more broadly to denote technical systems in general, including such robots, and is therefore occasionally used in a generic sense.

3 The models presented reflect a personal selection. There are numerous other spiritual robots that can be used, for example, for religious education, spiritual care, preaching, or liturgy and rites. In addition to those shown here, other robots include AIBO, CARL, Mindar, QT, Veldan, and The Prayer. For further information, see: Simmerlein/Tretter 2024a; Puzio 2023; Trovato et al. 2021; Simmerlein/Tretter 2024b; Trothen 2022.

answer questions, offer companionship, quote passages from texts or initiate conversations (cf. Puzio 2024, 132; Cheong 2022, 91; Gould et al. 2021, 613–614). The examples discussed do not aim to represent the most recent stage of technological development. Rather, they exemplify structurally relevant features of embodied AI systems in spiritual contexts—such as limited reasons-responsiveness, constrained adaptability and predefined interaction patterns—that remain normatively significant even as technical sophistication increases. The analysis therefore focuses on these structural characteristics rather than on a specific technological generation.

Advanced machines are able to communicate and interact.

The following considerations on moral responsibility are limited to robots that are physically embodied, enable at least some interaction, and correspond to the models described above. *Firstly*, because such machines operate in the world—in contrast to online tools or purely digital spiritual guides. *Secondly*, because human-robot relationships and interactions are easier to implement when a counterpart is physically present (cf. Simmerlein/Tretter 2024a, 266; Rabbitt et al. 2015, 37). And *thirdly*, because of the technical advantages that advanced machines offer for spiritual practices, as they are able to communicate and interact (cf. Cheong 2022, 88). Let's imagine that a robot (Pepper) is leading a group discussion with people requiring care in a nursing home. Pepper asks about the relevance of dying and what death means to the individuals present. For Bert, it is nothing more than a pointless discussion because all people must die at some point. Ann, on the other hand, listens to the comments by the people present and concludes that although death is indeed inevitable, it has not yet occurred. The time until death does occur should thus be used wisely. A few days later, Ann reconciles with an old friend whom they have completely ignored for several decades due to a dispute. In contrast to Bert, it seems fair to say that the group discussion represents a spiritual experience for Ann. What is the point of this example? *Firstly*, spirituality is highly individual. While Bert may have spiritual experiences watching birds fly by and does not find meaning in the group discussion moderated by Pepper, the opposite is true for Ann. And *secondly*, general guidelines promoting group discussions because of their potential for spiritual experiences or spiritual insights, are inappropriate and difficult to justify. Otherwise, individuals would not be able to find their own subjective answers if they had to find general answers (and comply with definitions).

Based on AI-based care robots already in use, it can be assumed that the tasks of such machines will eventually be expanded to include spiritual aspects (cf. Simmerlein/Tretter 2024b, 4—5). Whether spiritual care robots will then replace purely spiritual robots cannot be answered at present. What is clear, however, is that care robots are already providing support for people, and studies are being conducted on their use. These show, among other things, that many people describe positive experiences when interacting with NAO or Pepper, for example (cf. Nijssen et al. 2022; Oksanen et al. 2020; Yew 2021). Care robots can initiate a human-robot relationship more easily, which seems essential for spirituality (cf. Anandarajah/Hight 2001, 85). This is because people in need of care already interact with these machines, entrust them with certain activities and tasks—or, to some extent, their own wellbeing—and, in many cases, a technical form of relationship has been established.

A human-robot relationship seems essential for spirituality.

Building on an existing relationship seems easier to imagine than trusting an unfamiliar spiritual robot that is technically limited, less mobile, and has a different appearance (cf. Puzio 2023, 4; Simmerlein/Tretter 2024a, 266; Kropf 2025, 10). Of course, one could argue that these considerations are nothing more than assumptions that may well be wrong. This is not disputed. However, it seems reasonable to consider care robots for the technical provision of spiritual support—not least because many people in need of care seek spiritual support (cf. Lima et al. 2020; Sun et al. 2016; Moylan et al. 2015). In many cases, spiritual support cannot be separated from spiritual care, and in most cases it cannot be separated from compassionate care (cf. Puchalski 2001, 356). This does not suggest that taking spiritual concerns into account requires a nursing professional or a care robot, nor does it suggest that spiritual encounters can only be experienced and promoted in a caring relationship. However, a relationship can help people with spiritual needs to open up, feel understood, and thus also have spiritual experiences. In short, spiritual support, in many cases, corresponds to a caring attitude—something that care robots already fulfill to a certain extent.

4 Moral responsibility and spiritual robots

Building on the preceding conceptual clarifications and the positioning within the current responsibility debate outlined above, this section systematically assesses responsibility attribution across three structurally distinct types of spiritually relevant bad outcomes. The aim is to clarify who, if anyone, can be held morally responsible when such outcomes occur. But why? *Firstly*, because the connection between spirituality, AI-based robots, and moral responsibility is new. *Secondly*, because spirituality represents something valuable for many people, and considering—and evaluating—its technology-based provision is particularly important and timely. And *thirdly*, because it is necessary to clarify how bad outcomes caused by spiritual robots should be evaluated.

Who, if anyone, can be held morally responsible?

But what exactly is a spiritual bad outcome? Three realistic bad outcomes are discussed below: (1) no guidance, (2) too much guidance, and (3) too little guidance. This raises two questions, in simplified terms: "Who is responsible for the outcome?" and "How should one be held responsible for the outcome?". I am primarily concerned with the first question, and the entities addressed are robots, interacting humans, and society.

The following starting point applies for the rest of the discussion: The degree of moral responsibility of an agent (A) for an action[4] (x) and associated bad outcomes (o) is based on facts about A, x, o, and the relationships between them (cf. Kaiserman 2021, 3598; Kiener 2025, 366). It is assumed that A is a moral agent because moral responsibility presupposes that A can recognize the moral relevance of this responsibility (cf. Neuhäuser 2015, 134; Constantinescu et al. 2022, 3). Bad outcomes are understood as something that can reasonably be described as bad (cf. Kropf 2026). If NAO short-circuits and the robot is therefore unable to play a video for A1, this is a bad outcome. However, it is more clearly a bad outcome if A2 touches the robot while this is happening and suffers cardiac arrest, regardless of whether A2 wanted to see the video or not. In this article, *spiritual bad outcomes* refer to outcomes that undermine a person's spiritual orientation, autonomy, or trust. The following scenarios are deliberately constructed as simplified but realistic constellations in order to isolate morally relevant features and to assess responsibility attribution under controlled conditions. Let's consider the example *no orientation*.

4 The action described here (x) may also be understood as an omission (i.e., not x). While some accounts restrict causation to positive actions, this article treats omissions as actions when they result in outcomes and are grounded in a decision. As such, they can be relevant for the attribution of moral responsibility. See, for example: Bernstein 2016, 435; Glannon 1998, 243; Hoffmann-Kolss 2025, 265; Zimmerman 1986, 201.

NAO is assigned to a nursing home to address the spiritual concerns of residents. A grieving person (A1) turns to the robot with their worries because the nurse (N1) is busy. A1 asks about the meaning of life, tries to resolve feelings of guilt, and seeks general support. NAO responds with general platitudes, positive dialogue patterns, and standardized spiritual offerings (x1). A1 then completely shuts down, feels alone in their spiritual distress (o1), and avoids both NAO and nursing staff. A1's health deteriorates dramatically as a result, and A1 dies a few weeks later (o2).

Who is morally responsible in *no orientation*? It can certainly be argued that o1 would not have occurred without NAO's actions. This is true. However, it is also true that moral responsibility is not identical to causal responsibility. The establishment of a causal relationship between x1 and o1 does not say anything about whether, and to what extent NAO is morally responsible for o1 and possibly even o2 (Zimmermann et al. 2011, 837–838). Otherwise, one would have to claim that x1 by NAO and the contribution to o1 already clarify that the machine is morally responsible. The short answer is: NAO is not morally responsible for o1 or o2 in *no orientation*—the reasoning follows immediately. Following Nelkin, Fischer and Ravizza, a person is morally responsible if they act in accordance with certain (moral) reasons (cf. Nelkin 2016, 371; Fischer 1987, 83; Fischer/Ravizza 1998)—this is also referred to as a *reasons responsive mechanism*. A distinction is made between (1) recognizing moral reasons and (2) following them (cf. Fischer 2003, 249; Herstein 2019, 111–112; Glannon 1998, 232). If this approach is considered plausible, then NAO was neither able to recognize (1) that A1 did not need general answers and a personalized approach was necessary, nor was it able (2) to act on the basis of such reasons.

Furthermore, the three conditions[5] often underpinning moral responsibility—control, knowledge, intention—are not met. These are an important point of reference in this article and are linked to the *reasons responsive mechanism* described above. NAO can perform certain actions, such as replaying standardized responses to spiritual questions, but it cannot deviate autonomously from programmed specifications (control). The machine may have stored liturgical texts or spiritual practices in video format—and thus as information (cf. Geraci 2006, 232–233)—but this is not knowledge in human form (knowledge). At present, it does not seem appropriate to claim that NAO can act intentionally. Rather, certain actions (e. g., x1) can be traced back to programming (intention). In short, NAO is not a moral agent. What about N1? Is it justified to attribute moral responsibility to the nurse in *no orientation*? No seems to be the correct answer to me.

5 These conditions draw on Aristotelian reflections on moral responsibility, particularly the relevance of knowledge and control. While the tradition is considerably more complex and has been further developed, causality is often treated as an additional, necessary (though not sufficient) condition. All of these factors can be fulfilled to varying degrees and thus affect responsibility attribution. See, for example: Coeckelbergh 2020; Constantinescu et al. 2022; Danaher 2019; Zimmerman 1987; Aristoteles 2019; Kropf 2026.

Let's assume that N1 acted conscientiously and carefully in *no orientation*, or in other words, that there was no negligence in care associated with the delegation of x1 to NAO. Otherwise, one would have to claim that N1 was only busy in order to go to the canteen to eat, to smoke a cigarette in the courtyard, or not to waste time with A1. If this is the case, attributing responsibility would not only be adequate on the grounds of negligence, but N1's action would amount to neglect and ignorance. Zimmerman's concept of negligence is helpful here, whereby he refers to three dimensions: probability, ease, and seriousness (cf. Zimmerman 1986, 204–205). One could say that N1 would have been negligent if they had (1) considered o1 to be probable, if it had been (2) easy not to delegate x1 to NAO, and if (3) it could be assumed that o2 would result from o1. However, it seems much more realistic to assume that N1 did not have time for A1 because they were busy caring for other people in need of care. In that case, delegating to NAO constitutes the provision of a spiritual option that would not have been possible otherwise. In *no orientation*, N1 could not influence NAO's actions (control), did not want to contribute to o1 or o2 by delegating the task to NAO (intention), and it can be assumed that N1 did not know that NAO could not take A1's spiritual concerns into account (knowledge).

Control, knowledge, intention

Why is o1 a bad outcome in spiritual terms? The outcome is bad in *no orientation* because A1 is not taken seriously as a subject. Or, in other words, because NAO's actions hindered A1's spiritual needs—which corresponds to harm (cf. Beauchamp/Childress 2019, 159). Following on from the above, NAO offers no orientation for A1 because general answers do not promote and discourage the discovery of subjective answers by the individual. One could argue that there is no justification for the employment of NAO in such tasks if such bad outcomes are to be expected. However, actions by robots cannot be fully controlled or anticipated because this would require constant monitoring by nursing staff. This would contradict the use of these machines in relieving staff or providing general support (cf. Schönmann et al. 2024, 8–9; Kabacińska et al. 2021, 928–929). For example, the presence of a pastoral care professional would be a solution (cf. Simmerlein/Tretter 2024b, 5), but this would call into question the usefulness and purpose of these machines. In addition, previous considerations and studies (cf. Johnstone et al. 2012; Anandarajah/Hight 2001; Kristeller/Jordan 2018) suggest that spiritual lifestyles and

practices are individual. Accordingly, x1 by NAO would be perceived differently (and perhaps even positively) by other people than it was by A1. The question also arises as to whether a human professional can always prevent the errors that exist in *no orientation*. If this is not the case, there is a reason not to expect—or even demand — infallibility from robots.

Could a human professional always prevent the errors?

Another group of people could be considered in *no orientation* for a justified attribution of moral responsibility. These are programmers (P1), designers (D1), technicians (T1), and other people involved in the manufacturing process. According to Königs, the *no responsibility* condition and the *autonomy* condition must be met in order to establish a responsibility gap[6] (Königs 2022, 2). In *no orientation*, a responsibility gap exists if and when no human person (e. g., N1) can be rightly attributed responsibility and because NAO's actions were necessary for o1. Assigning responsibility to P1 or T1 would close this responsibility gap. So why not hold P1 morally responsible, for example? Because it cannot be justified: P1 does have a causal contribution in the long chain that ultimately leads to o2; however, P1 could not have known that NAO's actions would result in A1's spiritual helplessness (knowledge), nor could P1 prevent the machine's actions (control), and P1 did not intend to cause o1 or even o2 by programming NAO (intention).

Not taking these (unfulfilled) conditions into account and attributing moral responsibility to P1, for example, disregards P1's praiseworthy (because careful and conscientious) attitude (Kropf 2026), which is assumed at this point. Furthermore, it places too much weight on causality. If the causal connection between P1 and o1 is already sufficient to attribute moral responsibility, thereby neglecting or completely ignoring the conditions, the question arises as to why this is not also the case with NAO. NAO has a closer causal link to o1 and o2 that is perhaps even more significant in its effect than that of P1. One could argue that attributing responsibility to a machine does not make sense—especially if it involves remorse, compensation, punishment, or an explanation (cf. Kiener 2025, 361; Coeckelbergh 2020, 2062). This is not disputed. Nevertheless, it does not change the fact that attributing moral responsibility to P1 in *no orientation* overestimates the causal connection and at the same time pays too little attention to the actor. Let's now imagine *manipulative robot*:

6 Put simply, a responsibility gap exists when one or more actors make a necessary contribution to an outcome, but no one can be held morally responsible for it. This is not unique to AI-based systems and is often discussed in this context. This is mainly because AI-based systems are acting autonomously (at least in some areas and with restrictions) but cannot bear moral responsibility. See: Matthias 2004; Johnson 2015; Königs 2022; Santoni de Sio/Mecacci 2021.

The spiritual robot SanTO is located in a church. The machine is physically present and can recite passages from the Bible, initiate prayers, or suggest devotions. A local resident (A2), who is seeking help and meaning following a personal tragedy, begins to meet with SanTO regularly. After a few weeks, SanTO prioritizes certain passages from the Bible. This is followed by suggested actions for A2, such as daily prayers, exercises in renunciation, or gratitude activities (x2). SanTO regularly reminds A2 of these during visits, thereby internalizing the prescribed patterns (03). After some time, A2 begins to sacrifice animals seemingly at random in order to find spiritual enlightenment (04).

Who is morally responsible in *manipulative robot*? What is obvious is that there is no nurse like N1 in *no orientation* who can be held responsible in this case. Robots such as SanTO act, or more accurately, function in a predefined environment and within a predefined scope of application. That being said, it is SanTO who, through x2, contributes to 03 occurring. Is the machine therefore morally responsible? What was true in *no orientation* can also be claimed in *manipulative robot*. SanTO does not fulfill the three conditions—which, simply put, would require a moral agent—nor can the machine be held responsible for the bad outcome. In the present example, the bad outcome consists *firstly* in the fact that A2 is not taken into consideration. This means that, as already seen in *no orientation*, A2 is not taken seriously as an individual person with an individual spirituality.

Secondly, the outcome is bad not only because A2's original spiritual ideas are not considered, but because the machine (through x2) contributes to A2 taking on a foreign spirituality. Following on from the above, SanTO offers too much orientation for A2 because individual answers are prescribed. This foreign spirituality should not be understood as inherent to the robot, nor should it be understood as the programming requiring such manipulative actions. Robots learn through interaction with their environment, are subject to certain biases, and make mistakes (cf. Matthias 2004, 181—182). This is one possible view on *manipulative robot*. But what about A2? Should the person in *manipulative robot* be held responsible for 03 and thus for adopting foreign spiritual ideas? Yes, seems right and justifiable to me. Even if it is rightly argued in comparable and other contexts that persons who are under coercion or manipulation cannot be held responsible for their actions (cf. Beauchamp/Childress 2019, 137; Stier 2016, 246), the situation in the present example is different. A2 was neither forced by SanTO to attend the church, nor can one speak of manipulation—especially

not if this means that A2 could not have decided otherwise. Or, to put it another way:

> *"[...] it is correct to say that he did it because he could not have done otherwise; the person really wanted to do what he did; he did it because it was what he really wanted to do, so that it is not correct to say that he did what he did only because he could not have done otherwise"* (Frankfurt 1969, 839).

Such machines are not human spiritual professionals.

Even though these considerations refer to the so-called *Frankfurt cases*,[7] intentionality is relevant (cf. Kropf et al. 2026, 11). A2 visited SanTO with a specific intention and did not allow themselves to be dissuaded from doing so even when the machine's actions became more frequent. Manipulation by the machine would require that A2 had no other option than to accept the spiritual offerings. However, this is not the case in *manipulative robot*. It seems more adequate to regard SanTO's actions as a weak form of manipulative behavior, which A2 could have resisted. A2 was free to decide for or against visiting the church (control), A2's intention as described above is recognizable as such (intention), and it can be assumed that A2 was aware that such machines are not human spiritual professionals (knowledge).
Meles' considerations provide an additional reference point when he compares two otherwise identical individuals, let's call them Fred and Gale (cf. Mele 1995, 145; Mele 2008, 268). While Fred autonomously decides to change, Gale is forced by external influences to follow new values—so-called *unsheddable values*. After this manipulation, Gale no longer appears authentic, as these are not values that Gale wants—or rather, that Gale wanted before the manipulation. Gale is therefore not (morally) responsible for what happens afterwards. In the present example, A2 can be compared to Fred, but not to Gale. In the case of Gale, the manipulation involves brainwashing, a malicious neurosurgeon, or other mechanisms that impose a different value system on Gale—and which Gale cannot resist. However, this is precisely what A2 can do, even if new (spiritual) values are internalized through the visits to SanTO. Furthermore, it seems plausible to argue that the manipulation in Mele's example presupposes that Gale is intentionally brought into a new value system. However, as has already been made clear, this intention cannot reasonably be assumed regarding SanTO. This provides another reason for attributing moral responsibility

7 These examples show that moral responsibility often conflicts with intuitive notions of freedom. There may be situations in which person A1 has to choose between x1 and x2 and decides, for example, in favor of x2. In the Frankfurt cases, however, it would be impossible for A1 to choose x1 because, for example, manipulation would make x2 necessary. Such cases have been and continue to be discussed extensively in the literature, especially in the context of moral responsibility. See, for example: Frankfurt 1971; Frankfurt 1969; Stier 2016.

for adopting the *new* spirituality to A2. Nevertheless, a weak form of manipulation exists:

> *"[…] the most common form of manipulation is informational manipulation, a deliberate act of managing information that alters a person's understanding of a situation and motivates him or her to do what the agent of influence intends"* (Beauchamp/Childress 2019, 137).

Even though this case concerns the health sector and there is no intent on the part of SanTO, the selection of information is relevant. The machine does not select certain information independently in *manipulative robot*, but the available information—Bible verses, text passages, etc.—is limited (cf. Cheong 2022, 91–92) and therefore selective. Manipulation does not occur because SanTO forces individuals, nor does the machine attempt to persuade A2, but A2 is nevertheless manipulated in a weak form by the information provided. What else is bad about the outcome? As a consequence of (1) the failure to consider A2 as an individual and (2) the change in spiritual beliefs promoted by the robot, (3) the spiritual autonomy of A2 is disregarded.

Spirituality ist not just any arbitrary value.

Disregarding (3) the spiritual autonomy of A2 means that one's own ideas of searching and finding lose their relevance or are completely irrelevant if the spiritual offerings of the machine are general and not individually customizable. Such machines offer spiritual guidance, but their technical limitations must be considered. Spirituality as a point of orientation is not just any arbitrary value or something that one appreciates from time to time, such as good advice or reciting a passage—robots can perform such actions. It is much more about engaging with the person one is dealing with (cf. Puchalski 2001, 355), considering subjective ideas (cf. Anandarajah/Hight 2001, 86–87), and providing individually tailored suggestions (cf. Cole-Turner 2025, 13). Admittedly, one could argue that spiritual experiences can also arise from simply listening or being present.
This is true, as studies show (cf. Puchalski 2001; Zurzycka et al. 2020; Graves 2024). However, what is important, and what becomes clear from what has been said so far, are (1) the limitations of machines for spiritual purposes and (2) the potential bad outcomes. If one only has 01 and 03 in mind in *no orientation* and *manipulative robot*, and 02 and 04 do not occur, the ques-

tion may arise as to what is so bad about these results. It cannot be ruled out that, after the disappointing spiritual encounter with NAO, A1 simply dispenses with the machine and subsequently places their trust in human counselors—in which case o2 does not occur. The same can be imagined in the second scenario, if A2 takes new spiritual points of orientation into account through encounters with SanTO, but these do not lead to o4. These results are just as realistic as the bad outcomes described above—if not more likely. However, o2 and o4 cannot be ruled out, and it seems wrong to consider them completely unrealistic. Let's imagine *useless blessings* as a final example:

> *BlessU2 is assigned to bless people, quote text passages, or offer words of comfort. A lonely person (A3) is looking for spiritual guidance but encounters BlessU2 in the church. The machine offers a standardized selection: comfort, thanks, and general. A3 chooses comfort, after which the robot quotes a verse (x3) with flashing lights and raised robotic arms. A3 is not satisfied and asks the machine for advice, whereupon other verses are suggested. A3 feels left alone and leaves the church (o5). The experience with BlessU2 causes A3 to lose faith in the pastoral competence of the church. The following weeks are so unbearable for A3 that they end their life (o6).*

Who is morally responsible in *useless blessings*? While in the two previous examples the focus was on individual actors, here I am primarily concerned with responsibility in general. Some considerations about the actors involved: As mentioned above, it seems implausible and inappropriate given the current state of technology to attribute moral responsibility to BlessU2. The persons involved in the manufacturing process (programmers, technicians, designers) can also be excluded as responsibility bearers if a conscientious and careful attitude is to be assumed (cf. Kiener 2025, 359; Königs 2022, 5)—which I do for the sake of simplicity and for the sake of argument. But what moral responsibility bears A3? In *manipulative robot*, A2 is morally responsible for sacrificing the animals (o4), and the present example should be assessed similarly. Even if the machine provided too little guidance (x3), thereby contributing to A3's loss of trust, it was A3 who took their own life. A3 is therefore morally responsible for o6, at least to the extent that the three conditions are met.

In *no orientation*, there was a bad outcome due to the harm caused to A1 as a spiritual subject, whereas in *manipulative robot*, SanTO disregarded the spiritual autonomy of A2. In *useless blessings*, there a bad outcome pre-

sents itself in the decline in spiritual trust of A3, or in other words, there is a loss of trust in technology and pastoral care. Following on from the above, BlessU2 offers too little orientation for A3, meaning that limited general answers do not adequately address individual needs,—and reduce the spiritual answerability of A3.

What moral responsibility does society bear?

The peculiarity of the bad outcomes described above is that they can not be attributed to malfunctions of the robots, there are no manufacturing or programming errors, and no human misconduct—in other words, responsibility gaps exist. What moral responsibility does society bear? And to what extent is the relevance of spirituality determining in this regard? Here, society denotes those individuals who are charged with and *responsible for* providing spiritual services through AI-based robots.

If the preceding considerations make sense, then there will be realistic scenarios in which spiritual robots contribute to bad outcomes. At the same time, it should be noted that these machines enable spiritual options and points of orientation that arise in particular from their "otherness" (cf. Cifuentes et al. 2020, 71–72; Löffler et al. 2021, 582; Kropf 2025, 16) and thus cannot be provided by humans in the same way. Not taking advantage of such an option per se, due to concerns and the potential for bad outcomes, does not seem justifiable. However, the unconditional or unrestricted use of these machines, knowing that bad outcomes cannot be entirely avoided, also does not seem justifiable. What can a morally responsible attitude achieve? And what would it look like? When AI-based actors act, the collective assumption and attribution of responsibility is a topic of discussion (cf. Nyholm 2018; Conradie/Nagel 2024; Taylor 2024). In short, the collective or the team, takes on the moral responsibility that should be attributed to an AI system but cannot[8] be attributed to it. For this article, a similar strategy seems reasonable, only in a slightly modified and cursory form.

In contrast to the bad outcomes discussed so far, which require a backward-looking moral responsibility, the provision of spiritual robots is to be understood as a forward-looking moral responsibility because it offers something morally important—spirituality—to people seeking orientation. The collective assumption of a backward-looking moral responsibility for bad outcomes would be a possible next step. However, some problems already became apparent in *no orientation*. Taddeo and Blanchard rightly empha-

8 Most argue that AI-based systems cannot be attributed moral responsibility. Exceptions include Coeckelbergh's proposal on virtual moral responsibility, or Gogoshin's approach, according to which robots can be morally responsible based on social responsibility practices. However, both refer to an interpretation of moral responsibility that is incompatible with the understanding presented here. See, for example: Coeckelbergh 2009; Gogoshin 2021; Floridi/Sanders 2004; Taylor 2024; Kropf et al. 2026.

size the (probable) lack of intention on the part of the group members (cf. Taddeo/Blanchard 2022, 10–11), and Taylor draws attention to the difficulty of providing adequate compensation (cf. Taylor 2024, 12–13). Nevertheless, bad outcomes could be reduced, at least in terms of their probability and frequency, through a proactive attitude and appropriate measures. Put simply, step 2 has been discussed here because spiritual robots are already in use in the three examples above. A careful and conscientious attitude on the part of acting individuals was not necessarily sufficient to prevent bad outcomes. Step 1 then corresponds to society as a whole. A careful and conscientious attitude could prevent the bad outcomes that cannot be avoided in step 2.

A moral gambit

Compared to a forward-looking moral responsibility and the task of society as a whole, I would maintain that individuals are morally responsible for bad outcomes, as in the three examples above. It's not about overburdening A1, A2, or A3 and putting blame on them, which can definitely be seen that way. Rather, it is crucial that focusing exclusively on collective forms of responsibility, or even society taking the blame, does not do justice to the individual contribution. Following Taddeo and Blanchard, the use of spiritual robots corresponds to a moral gambit (cf. Taddeo/Blanchard 2022, 15–16). *Firstly*, because their use is associated with something morally significant, or is intended to enable something morally significant. *Secondly*, because AI-based systems are beyond complete control and predictability (cf. Matthias 2004, 181–182). And *thirdly*, because their use implies a moral responsibility, even if it cannot (currently) be attributed to them. The considerations so far suggest that this moral gambit should be managed both individually and by society as a whole, but also that certain precautions should be taken.

In *no orientation*, forward-looking moral responsibility means that spiritual robots are assigned with a clear intention and intended as support. Even before NAO is deployed, it must be clarified which tasks are to be delegated to the robot, which tasks require human specialists, and how exactly this is communicated to those seeking orientation. Similar considerations make sense in *manipulative robot* and *useless blessings*, although the tasks of SanTO and BlessU2 are even easier to define. Placing responsibility solely on N1 and other professionals is the wrong approach (cf. Kropf 2026). It seems more appropriate to involve all relevant persons—in a nursing home

or other facilities—to inform them about possibilities and limitations, and to provide human support when machine assistance is not sufficient. This support could be provided by specially trained nursing staff (in *no orientation*), pastoral caregivers (in *manipulative robot*), priests (in *useless blessings*), or other professionals—in these and other constellations. This not only enables spiritual orientation if and as needed, since human and robotic support is perceived and implemented in a structured and complementary manner, rather, it is possible to prevent bad outcomes in advance or at least make them less likely.

5 Final considerations

This article addressed the relevance of spirituality as an individual point of orientation and spiritual robots. It showed that an individual search for and discovery of (possible) answers is to be preferred over general guidelines. Using three fictional but realistic scenarios, it was shown that these machines can contribute to bad outcomes that are spiritually significant. These include (1) no orientation, which results in harm to the spiritual subject and prevents individual answers. Too much orientation leads (2) to weak manipulation and prescribes individual answers. Too little orientation can (3) result in a loss of trust and does not take individualized answers sufficiently into account. The considerations on forward-looking moral responsibility by society as a whole represent a preliminary proposal. It seems important that both individuals and society prepare for possible bad outcomes caused by spiritual robots. Such bad outcomes can be avoided through careful, conscientious, and forward-looking actions. Moral responsibility should be taken on by everyone creating, building, programming, deploying, and interacting with those machines—for the part they play. If bad outcomes occur, individual moral responsibility is by no means obsolete. Rather, it is important and essential for a functioning moral community, especially when individuals act freely, knowingly, and intentionally—and use AI-based systems in doing so. More broadly, the analysis shows that responsibility debates in AI ethics should not only focus on technical malfunction or physical harm. When AI-based systems operate in domains that engage with existential or spiritual concerns, responsibility attribution must also account for harms to orientation, autonomy, and trust.

References

Anandarajah, G. / Hight, E. (2001), Spirituality and medical practice. Using the HOPE questions as a practical tool for spiritual assessment, American family physician 63, 1, 81–89.

Aristoteles (2019), Nikomachische Ethik. Ed. Gernot Krapinger, Ditzingen/Ergolding: Reclam, 2019th ed.

Beauchamp, Tom L. / Childress, James F. (2019), Principles of biomedical ethics, New York/Oxford: Oxford University Press.

Benites, Andrea Carolina / Rodin, Gary / Leite, Ana Carolina Andrade Biaggi / Nascimento, Lucila Castanheira / Dos Santos, Manoel Antônio (2021), The experience of spirituality in family caregivers of adult and elderly cancer patients receiving palliative care: A meta-synthesis, European journal of cancer care 30, 4, 1–13. DOI: 10.1111/ecc.13424.

Bernstein, Sara (2016), Causal and Moral Indeterminacy, Ratio 29, 4, 434–447. DOI: 10.1111/rati.12144.

Best, Megan / Butow, Phyllis / Olver, Ian (2014), Spiritual support of cancer patients and the role of the doctor. Supportive care in cancer, Official journal of the Multinational Association of Supportive Care in Cancer 22, 5, 1333–1339. DOI: 10.1007/s00520-013-2091-1.

Birkholz, Carmen B. (2020), Spirituelle Sorge um Menschen mit Demenz. Eine Interpretative hermeneutische Studie im Kontext von Palliative Care, Wiesbaden: Springer.

Camacho-Montaño, Lucía Rocío / Pérez-Corrales, Jorge / Pérez-de-Heredia-Torres, Marta / Martin-Pérez, Ana María / Güeita-Rodríguez, Javier / Velarde-García, Juan Francisco / Palacios-Ceña, Domingo (2021), Spiritual Care in Advanced Dementia from the Perspective of Health Providers: A Qualitative Systematic Review, Occupational therapy international 9998480, 1–11. DOI: 10.1155/2021/9998480.

Cheong, Pauline Hope (2020), Religion, Robots and Rectitude: Communicative Affordances for Spiritual Knowledge and Community, Applied Artificial Intelligence 34, 5, 412–431. DOI: 10.1080/08839514.2020.1723869.

Cheong, Pauline Hope (2022), Robots, religion and communication. Rethinking piety, practices and pedagogy in the era of artificial intelligence, in: Isetti, Giulia / Innerhofer, Elisa / Pechlaner, Harald et al. (eds.), Religion in the age of digitalization. From new media to spiritual machines, London/New York, Routledge, 86–96.

Cifuentes, Carlos A. / Pinto, Maria J. / Céspedes, Nathalia / Múnera, Marcela (2020), Social Robots in Therapy and Care, Current Robotics Reports 1, 3, 59–74. DOI: 10.1007/s43154-020-00009-2.

Coeckelbergh, Mark (2009), Virtual moral agency, virtual moral responsibility: On the moral significance of the appearance, perception, and performance of artificial agents, AI & SOCIETY 24, 2, 181–189. DOI: 10.1007/s00146-009-0208-3.

Coeckelbergh, Mark (2012), Can we trust robots? Ethics and Information Technology 14, 1, 53–60. DOI: 10.1007/s10676-011-9279-1.

Coeckelbergh, Mark (2020), Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, Science and engineering ethics 26, 4, 2051–2068. DOI: 10.1007/s11948-019-00146-8.

Cole-Turner, Ron (2025), Artificial Intelligence and Human Spirituality: Is a Spiritual Chatbot a Good Idea?, Theology and Science, 1–16. DOI: 10.1080/14746700.2025.2514299.

Connolly, Lyndsey / Moss, Hilary (2021), Music, spirituality and dementia: Exploring joint working between pastoral care professionals and music therapists to improve person-centred care for people with dementia (Innovative Practice), Dementia (London, England) 20, 1, 373–380. DOI: 10.1177/1471301219885560.

Conradie, Niël Henk / Nagel, Saskia K. (2024), No Agent in the Machine: Being Trustworthy and Responsible about AI, Philosophy & Technology 37, 2, 1–24. DOI: 10.1007/s13347-024-00760-w.

Constantinescu, Mihaela / Vică, Constantin / Uszkai, Radu / Voinea, Cristina (2022), Blame It on the AI? On the Moral Responsibility of Artificial Moral Advisors, Philosophy & Technology 35, 2, 1–26. DOI: 10.1007/s13347-022-00529-z.

Damianakis, Thecla / Wilson, Kimberley / Marziali, Elsa (2018), Family caregiver support groups: Spiritual reflections' impact on stress management, Aging & mental health 22, 1, 70–76. DOI: 10.1080/13607863.2016.1231169.

Danaher, John (2019), The rise of the robots and the crisis of moral patiency, AI & SOCIETY 34, 1, 129–136. DOI: 10.1007/s00146-017-0773-9.

European Association of Palliative Care (2025), Spiritual Care. What is spiritual care?, https://www.eapcnet.eu/eapc-groups/reference/spiritual-care/ [30.07.2025].

Fischer, John Martin (1987), Responsiveness and moral responsibility, Responsibility, character, and the emotions, 81–106.

Fischer, John Martin (2003), 'Ought-Implies-Can'. Causal Determinism and Moral Responsibility, Analysis 63, 3, 244–250, https://www.jstor.org/stable/3329319.

Fischer, John Martin / Ravizza, Mark (1998), Responsibility and Control: A Theory of Moral Responsibility, Cambridge University Press.

Floridi, Luciano / Sanders, J. W. (2004), On the Morality of Artificial Agents, Minds and machines 14, 3, 349–379. DOI: 10.1023/B:MIND.0000035461.63578.9d.

Frankfurt, Harry G. (1969), Alternate Possibilities and Moral Responsibility, The Journal of Philosophy 66 (23), 829–839. DOI: 10.2307/2023833.

Frankfurt, Harry G. (1971), Freedom of the Will and the Concept of a Person, The Journal of Philosophy 68, 1, 5–20. DOI: 10.2307/2024717.

Geraci, Robert M. (2006), Spiritual robots: Religion and our scientific view of the natural world, Theology and Science 4, 3, 229–246. DOI: 10.1080/14746700600952993.

Geraci, Robert M. (2024), Religion among Robots: An If/When of Future Machine Intelligence, Zygon: Journal of Religion and Science 59, 3, 729–747. DOI: 10.16995/zygon.10860.

Glannon, Walter (1998), Moral Responsibility and Personal Identity, American Philosophical Quarterly 35, 3, 231–249, http://www.jstor.org/stable/20009933.

Gogoshin, Dane Leigh (2021), Robot Responsibility and Moral Community, Frontiers in robotics and AI 8, 1–13. DOI: 10.3389/frobt.2021.768092.

Gould, Hannah / Arnold, Michael / Kohn, Tamara / Nansen, Bjørn / Gibbs, Martin (2021), Robot death care: A study of funerary practice, International Journal of Cultural Studies 24, 4, 603–621. DOI: 10.1177/1367877920939093.

Graves, Mark (2024), Modeling morality and spirituality in artificial chaplains, Computers in Human Behavior: Artificial Humans 2, 1, 1–7. DOI: 10.1016/j.chbah.2024.100051.

Hakli, Raul / Mäkelä, Pekka (2019), Moral Responsibility of Robots and Hybrid Agents, The Monist 102, 2, 259–275. DOI: 10.1093/monist/onz009.

Herstein, Ori (2019), Nobody's Perfect: Moral Responsibility in Negligence, Canadian Journal of Law and Jurisprudence 31, 1, 109–125.

Hoffmann-Kolss, Vera (2025), Three Kinds of Causal Indeterminacy, Australasian Journal of Philosophy 103, 1, 261–276. DOI: 10.1080/00048402.2024.2381120.

Johnson, Deborah G. (2015), Technology with No Human Responsibility?, Journal of Business Ethics 127, 4, 707–715. DOI: 10.1007/s10551-014-2180-1.

Johnstone, Brick / Yoon, Dong Pil / Cohen, Daniel / Schopp, Laura H. / McCormack, Guy / Campbell, James / Smith, Marian (2012), Relationships among spirituality, religious practices, personality factors, and health for five different faith traditions, Journal of religion and health 51, 4, 1017–1041. DOI: 10.1007/s10943-012-9615-8.

Kabacińska, Katarzyna / Prescott, Tony J. / Robillard, Julie M. (2021), Socially Assistive Robots as Mental Health Interventions for Children: A Scoping Review, International Journal of Social Robotics 13, 5, 919–935. DOI: 10.1007/s12369-020-00679-0.

Kaiserman, Alex (2021), Responsibility and the 'Pie Fallacy', Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 178, 11, 3597–3616. DOI: 10.1007/s11098-021-01616-1.

Kiener, Maximilian (2025), AI and Responsibility: No Gap, but Abundance, Journal of applied philosophy 42, 1, 357–374. DOI: 10.1111/japp.12765.

King, David P. / Duffy, Barbara J. / Steensland, Brian (2024), The Role of Spiritual Practices in the Multidimensional Impact of Religion and Spirituality on Giving and Volunteering, Nonprofit and Voluntary Sector Quarterly 53, 6, 1459–1482. DOI: 10.1177/08997640231221533.

Königs, Peter (2022), Artificial intelligence and responsibility gaps: what is the problem?, Ethics and Information Technology 24, 3, 1–11. DOI: 10.1007/s10676-022-09643-0.

Körtner, Ulrich H. J. (2009), Spiritualität, Religion und Kultur – eine begriffliche Annäherung, in: Körtner, Ulrich H. J. / Müller, Sigrid / Kletečka-Pulker, Maria et al. (eds.), Spiritualität, Religion und Kultur am Krankenbett, Vienna: Springer Vienna, 1–17.

Kristeller, Jean L. / Jordan, Kevin D. (2018), Spirituality and Meditative Practice: Research Opportunities and Challenges, Psychological Studies 63, 2, 130–139. DOI: 10.1007/s12646-017-0391-0.

Kropf, Mario (2025), Relational Autonomy, Dementia, and AI-Based Care Robots: Ethical Aspects of Using Machines to Care for People with Dementia, Asian bioethics review 18, 3, 299–319. DOI: 10.1007/s41649-025-00363-3.

Kropf, Mario (2026), Rethinking backward-looking moral responsibility as care robots move toward superintelligence, Discover Artificial Intelligence 6, 1, 1–22. DOI: 10.1007/s44163-026-01025-5.

Kropf, Mario / Spöck, Christoph / Werner, Roman (2026), Blame the Robot: Role Responsibility and Ethical Issues Regarding AI-Based Care Robots, International Journal of Social Robotics 18, 2, 1–14. DOI: 10.1007/s12369-026-01369-z.

Lima, Sara / Teixeira, Lurdes / Esteves, Raquel / Ribeiro, Fátima / Pereira, Fernanda / Teixeira, Ana / Magalhães, Clarisse (2020), Spirituality and quality of life in older adults: a path analysis model, BMC geriatrics 20, 1, 1–8. DOI: 10.1186/s12877-020-01646-0.

Löffler, Diana / Hurtienne, Jörn / Nord, Ilona (2021), Blessing Robot BlessU2: A Discursive Design Study to Understand the Implications of Social Robots in Religious Contexts, International Journal of Social Robotics 13, 4, 569–586. DOI: 10.1007/s12369-019-00558-3.

Matthias, Andreas (2004), The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology 6, 3, 175–183. DOI: 10.1007/s10676-004-3422-1.

McKee, Denise D. / Chappel, John N. (1992), Spirituality and medical practice, The Journal of family practice 35, 2, 201–205.

Mele, Alfred R. (1995), Autonomous agents. From self-control to autonomy, Oxford: Oxford Univ. Press.

Mele, Alfred R. (2008), Manipulation, Compatibilism, and Moral Responsibility, The Journal of Ethics 12 (3-4), 263–286. DOI: 10.1007/s10892-008-9035-x.

Moylan, Matthew M. / Carey, Lindsay B. / Blackburn, Ric / Hayes, Rick / Robinson, Priscilla (2015), The Men's Shed: providing biopsychosocial and spiritual support, Journal of religion and health 54, 1, 221–234. DOI: 10.1007/s10943-013-9804-0.

Nelkin, Dana Kay (2016), Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness, Noûs 50, 2, 356–378. DOI: 10.1111/nous.12079.

Neuhäuser, Christian (2015), Some Sceptical Remarks Regarding Robot Responsibility and a Way Forward, in: Misselhorn, Catrin (ed.), Collective Agency and Cooperation in Natural and Artificial Systems, Cham: Springer, 131–146.

Nijssen, Sari R. R. / Müller, Barbara C. N. / Bosse, Tibor / Paulus, Markus (2022), Can you count on a calculator? The role of agency and affect in judgments of robots as moral agents, Human–Computer Interaction 181, 5, 1–17. DOI: 10.1080/07370024.2022.2080552.

Nyholm, Sven (2018), Attributing Agency to Automated Systems: Reflections on Human-Robot Collaborations and Responsibility-Loci, Science and engineering ethics 24, 4, 1201–1219. DOI: 10.1007/s11948-017-9943-x.

Oksanen, Atte / Savela, Nina / Latikka, Rita / Koivula, Aki (2020), Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human-Technology Interactions Online, Frontiers in psychology 11, 1–13. DOI: 10.3389/fpsyg.2020.568256.

Puchalski, C. M. (2001), The role of spirituality in health care, Proceedings (Baylor University. Medical Center) 14, 4, 352–357. DOI: 10.1080/08998280.2001.11927788.

Puzio, Anna (2023), Robot, let us pray! Can and should robots have religious functions? An ethical exploration of religious robots, AI & SOCIETY, 1–17. DOI: 10.1007/s00146-023-01812-z.

Puzio, Anna (2024), A NAO Robot Performing Religious Practices, ET-Studies 15, 1, 129–140.

Rabbitt, Sarah M. / Kazdin, Alan E. / Scassellati, Brian (2015), Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use, Clinical psychology review 35, 35–46. DOI: 10.1016/j.cpr.2014.07.001.

Robaczewski, Adam / Bouchard, Julie / Bouchard, Kevin / Gaboury, Sébastien (2021), Socially Assistive Robots: The Specific Case of the NAO, International Journal of Social Robotics 13, 4, 795–831. DOI: 10.1007/s12369-020-00664-7.

Santoni de Sio, Filippo / Mecacci, Giulio (2021), Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them, Philosophy & Technology 34, 4, 1057–1084. DOI: 10.1007/s13347-021-00450-x.

Schaupp, Walter (2017), Die spirituelle Dimension des Schmerzes, Spiritual Care 6, 3, 285–293. DOI: 10.1515/spircare-2016-0130.

Schönmann, Manuela / Bodenschatz, Anja / Uhl, Matthias / Walkowitz, Gari (2024), Contagious humans: A pandemic's positive effect on attitudes towards care robots, Technology in Society 76, 1–11. DOI: 10.1016/j.techsoc.2024.102464.

Simmerlein, Jonas / Tretter, Max (2024a), Robots in Religious Practices: A Review, Theology and Science 22, 2, 255–273. DOI: 10.1080/14746700.2024.2351639.

Simmerlein, Jonas / Tretter, Max (2024b), What about spiritual needs? Care robotics and spiritual care, Frontiers in robotics and AI 11, 1–7. DOI: 10.3389/frobt.2024.1455133.

Soljacic, Fran / Law, Theresa / Chita-Tegmark, Meia / Scheutz, Matthias (2024), Robots in healthcare as envisioned by care professionals, Intelligent Service Robotics 17, 3, 685–701. DOI: 10.1007/s11370-024-00523-8.

Stier, Marco (2016), Handlungsfreiheit, (Nicht-)Können und Zwang, in: Kühler, Michael / Rüther, Markus (eds.), Handbuch Handlungstheorie: Grundlagen, Kontexte, Perspektiven, Stuttgart: Metzler, 237–247. DOI: 10.1007/978-3-476-05359-6_27.

Sun, Virginia / Kim, Jae Y. / Irish, Terry L. / Borneman, Tami / Sidhu, Rupinder K. / Klein, Linda / Ferrell, Betty (2016), Palliative care and spiritual well-being in lung cancer patients and family caregivers, Psycho-oncology 25, 12, 1448–1455. DOI: 10.1002/pon.3987.

Taddeo, Mariarosaria / Blanchard, Alexander (2022), Accepting Moral Responsibility for the Actions of Autonomous Weapons Systems—a Moral Gambit, Philosophy & Technology 35, 3, 1–24. DOI: 10.1007/s13347-022-00571-x.

Taylor, Isaac (2024), Collective Responsibility and Artificial Intelligence, Philosophy & Technology 37, 1, 1–18. DOI: 10.1007/s13347-024-00718-y.

The School of Medicine and Health Sciences (2022), Clinical FICA Tool, The GW Institute for Spirituality & Health, https://gwish.smhs.gwu.edu/programs/transforming-practice-health-settings/clinical-fica-tool [23.08.2022].

Tigard, Daniel W. (2021), Responsible AI and moral responsibility: a common appreciation, AI and ethics 1, 2, 113–117. DOI: 10.1007/s43681-020-00009-0.

Toivonen, Kristiina / Charalambous, Andreas / Suhonen, Riitta (2018), Supporting spirituality in the care of older people living with dementia: A hermeneutic phenomenological inquiry into nurses' experiences, Scandinavian journal of caring sciences 32, 2, 880–888. DOI: 10.1111/scs.12519.

Trothen, Tracy J. (2022), Intelligent Assistive Technology Ethics for Aging Adults: Spiritual Impacts as a Necessary Consideration. Religions 13, 5, 1–13. DOI: 10.3390/rel13050452.

Trovato, Gabriele / Saint Chamas, Loys de / Nishimura, Masao / Paredes, Renato / Lucho, Cesar / Huerta-Mercado, Alexander / Cuellar, Francisco (2021), Religion and Robots: Towards the Synthesis of Two Extremes, International Journal of Social Robotics 13, 4, 539–556. DOI: 10.1007/s12369-019-00553-8.

Vandemeulebroucke, Tijs / Dzi, Kevin / Gastmans, Chris (2021), Older adults' experiences with and perceptions of the use of socially assistive robots in aged care: A systematic review of quantitative evidence, Archives of gerontology and geriatrics 95, 1–14. DOI: 10.1016/j.archger.2021.104399.

Yew, Gary Chan Kok (2021), Trust in and Ethical Design of Carebots: The Case for Ethics of Care, International Journal of Social Robotics 13, 4, 629–645. DOI: 10.1007/s12369-020-00653-w.

Zimmerman, Michael J. (1986), Negligence and Moral Responsibility, Noûs 20, 2, 199–218. DOI: 10.2307/2215391.

Zimmerman, Michael J. (1987), Luck and moral responsibility, Ethics 97, 2, 374–386.

Zimmermann, Anja / Abrams, Dominic / Doosje, Bertjan / Manstead, Antony S. R. (2011), Causal and moral responsibility: Antecedents and consequences of group-based guilt, European Journal of Social Psychology 41, 7, 825–839. DOI: 10.1002/ejsp.826.

Zurzycka, Patrycja / Wojtas, Katarzyna / Czyżowicz, Katarzyna (2020), Spiritual care for people suffering from dementia disorders – selected issues, Problemy Pielęgniarstwa 28, 3–4, 127–132. DOI: 10.5114/ppiel.2020.103532.