Philosophy & Technology (2023) 36:20
https://doi.org/10.1007/s13347-023-00620-z

RESEARCH ARTICLE

# Meaning in Life in AI Ethics—Some Trends and Perspectives

**Sven Nyholm**[1] · **Markus Rüther**[2]

Received: 31 October 2022 / Accepted: 27 February 2023


**Abstract**
In this paper, we discuss the relation between recent philosophical discussions about meaning in life (from authors like Susan Wolf, Thaddeus Metz, and others) and the ethics of artificial intelligence (AI). Our goal is twofold, namely, to argue that considering the axiological category of meaningfulness can enrich AI ethics, on the one hand, and to portray and evaluate the small, but growing literature that already exists on the relation between meaning in life and AI ethics, on the other hand. We start out our review by clarifying the basic assumptions of the meaning in life discourse and how it understands the term 'meaningfulness'. After that, we offer five general arguments for relating philosophical questions about meaning in life to questions about the role of AI in human life. For example, we formulate a worry about a possible meaningfulness gap related to AI on analogy with the idea of responsibility gaps created by AI, a prominent topic within the AI ethics literature. We then consider three specific types of contributions that have been made in the AI ethics literature so far: contributions related to self-development, the future of work, and relationships. As we discuss those three topics, we highlight what has already been done, but we also point out gaps in the existing literature. We end with an outlook regarding where we think the discussion of this topic should go next.

**Keywords** Meaning in life · Artificial intelligence · AI ethics · Self-development · The future of work · Relationships

## 1 Introduction

Imagine for a moment that the prognoses of some AI utopians have come true: namely, that we live in a world in which many, or perhaps most, of our daily laborious decisions have been either outsourced to decision-support systems or at least recommended by them, and that most of the burdensome things we previously had

✉ Markus Rüther
m.ruether@fz-juelich.de

[1] Ludwig Maximilian University of Munich, Munich, Germany

[2] Research Center Jülich GmbH, Jülich, Germany

Published online: 25 March 2023

to do ourselves are now done for us by robots and other smart AI technologies. Imagine what a day in someone's life might look like.

Breakfast would be recommended for this person by a combination of a smart fridge (which makes sure that the right foods are always available) and an app on their smartphone—so there is no need to make a decision about what to have for breakfast. Normally, this person now would start to work at home, but today they leave their home-office and are driven to their workplace by a self-driving car, which not only drives them to work, but also recommends a quick stop at the gym on the way there (because the smart car seat can tell that this person has gained a small amount of weight), and the person obliges. Once they arrive at the workplace, their work mostly consists of monitoring the behaviour of various robots and other AI systems, and agreeing to suggestions made by recommender systems, where a human needs to take responsibility for what the AI technologies suggest or do.

After work, this person is informed by an app on their phone that it would be a good idea to send flowers to their romantic partner, along with a text message generated by a large language model, saying something the partner is likely to enjoy hearing—the only thing our imagined person has to do is to choose from a menu of possibly appropriate answers. Most days in this person's life follow this pattern: they do not have to engage in much creative thinking, come up with ideas, think about alternatives, or do much planning—most aspects of their life that previously involved using their own intelligence have been outsourced to different forms of AI technologies. Others which have not been outsourced are optimised: instead of calling a human friend, this person often simply talks with a chatbot that has been optimised to simulate the kind of relationship interaction they are most comfortable with. They only see their romantic partner on days when the recommender system they both use predicts that they are in the right mood to see each other, so that each interaction is optimised to go well.

Can we describe such a life as one that deserves praise, pride, and admiration? Is it a life that is important, significant, and perhaps connected to 'higher values'? In short, is such a life a meaningful life?[1] Intuitively, these (and related) questions about meaningfulness are pressing for many people. Yet AI ethics as a discipline has typically not paid much attention to such questions, but has primarily focused on other issues instead. It is only recently that some ethicists have come to see that the idea of meaning may be of help in broadening the discussion and addressing some of the main challenges in the field.

We explore this new trend by providing a critical mapping of the main topics in AI ethics with respect to which the idea of meaningfulness is gaining more and more attention. Additionally, we formulate five key arguments about why AI is relevant to issue of meaning in life. In other words, this paper is partly a literature review, which highlights key contributions that have recently been made regarding the relation between meaningfulness in life and AI. At the same time, the paper also

[1] This is of course solely an initial intuitive approximation, which might be enough at this stage in order to have a glimpse what ethicists in the discourse have in mind when they talk about meaningfulness. For a more elaborated description of what can and has been described as meaningfulness, see Sects. 2 and 3.

seeks to articulate general arguments for why meaning in life should be a significant topic within AI ethics, doing so at a higher level of abstraction than has so far been done in the AI ethics literature.

We proceed as follows: after distinguishing between narrow and broad conceptions of AI ethics and motivating including the notion of meaning in life in broader discussions of AI ethics (Section 2), we clarify what talk of 'life's meaning' and of cognate terms mean according to the protagonists in the meaning in life debate (Section 3). We note especially that one of the driving forces behind the new trend is the idea that the term 'meaningfulness' includes senses that differentiate it from other axiological categories, particularly 'self-interest' and 'moral rightness'. Next, we articulate our five general arguments that demonstrate how questions of meaningfulness arise from our use of AI technologies (Section 4). Some of these five arguments are generalizations of more specific arguments found in the literature, and are intended to offer a general case for the research program of exploring meaning in life in the context of AI ethics. With those five basic key arguments in place, we then turn to the current literature and consider the category of meaningfulness in relation to AI technologies in three contexts (Section 5): personal self-development, the workplace, and social relationships. We will make our way through these areas by commenting on the main materials available (e.g. recent papers and books) and outlining the most common arguments.[2] We end with three high-level observations that are meant to pave the way further research on meaningfulness and AI (Section 6).

## 2 AI Ethics and Meaning in Life

'Artificial intelligence' (AI) is a general term that refers to technologies that can perform or take over tasks normally associated with intelligent human behaviour, including but not limited to learning through interaction with one's environment and the optimization of one's goal pursuit (Dignum, 2019). This general idea can be and has been interpreted in various different ways. According to Alan Turing's (1950) famous way of approaching this topic, we create 'thinking machines' if we create machines that are able to *imitate* intelligent human behaviour. In contrast, the team of scientists who coined the term 'artificial intelligence' in a 1955 research proposal spoke about creating technologies that *simulate* intelligent behaviour (McCarthy et al., 2006). More recent definitions—such as the one in Stuart Russell and Peter Norvig's (1995/2020) influential textbook *Artificial Intelligence*: *A Modern*

[2] We limit our focus mainly to books and papers published within last 15 to 20 years in the so-called meaning in life discourse. Of course, there are more thoughts on meaningfulness in the history of philosophy, especially if one considers the implicit talk and thought on meaningfulness; however, those must be part of another, more encompassing project. We also restrict our overview to English and German sources, mainly in analytic philosophy. On the one hand, we are most familiar with those, but on the other, we also lean towards the position that those resources are, as far as we know, the most elaborate and systematic in which you can find explicit thoughts on meaningfulness, especially as it relates to AI. For some further comments on the literature that we include, see Section 5.1.

*Approach*—typically understand AI in terms of the creation of *artificial agents*.[3] In general, then, AI refers to technologies that can either really be intelligent (whatever that would mean) or that could imitate or simulate intelligence, and/or technologies that can be seen as a form of artificial agents.

AI ethics is the study of ethical questions related to AI technologies. Without question, it is one of the hottest sub-fields within contemporary applied ethics. Many topics are explored, from different theoretical perspectives, and with different aims. Many contributions to AI ethics focus primarily on somewhat narrow questions about what is wrong or unjust about certain ways in which AI technologies might be used, for example, because they violate privacy, create injustice (e.g. due to biases in data) or cause harm (e.g. people being harmed by self-driving cars or military robots) (for overviews, see Tegmark, 2017; Coeckelbergh, 2020; Müller, 2020; Gordon & Nyholm, 2021; Heinrichs et al., 2022). Key questions concern what is morally acceptable or unacceptable about possible uses of AI technologies, along with questions about who should be held responsible when AI technologies cause harm or injustices. These are important ethical challenges and issues. Nevertheless, some ethicists have more recently begun to feel that this is not all that can be said about the value or disvalue of AI technologies.

Some argue that we need a broader approach, which also asks questions about what role(s) AI might play within a good human life—both whether AI might threaten our opportunities to live good human lives and whether AI might create new forms of opportunities to live good human lives (e.g. Chalmers, 2022; Danaher, 2019a; Nyholm, 2023; Tasioulas, 2022). Within this broader ethical discussion of AI, some authors have begun discussing the effects that AI might have on meaningfulness—including whether AI threatens, or might open up new types of opportunities for, meaning in life (see the cited literature in Section 5 below).

This broader view on AI ethics is in line with a 'bigger' development that has taken place within the last 15–20 years in normative ethics, which is also spreading into other areas of applied ethics. In philosophical circles, it is formally known as the debate on *meaning in life*. Authors such as Susan Wolf (2010) and Thaddeus Metz (2013) have undertaken highly influential groundwork on this topic (for an overview of the field, see Metz (2013), Rüther (2021a, b), and the contributions in Landau (2022)). Many researchers are also now trying to apply the category of meaningfulness to topics debated in applied ethics, for instance regarding life and death in medical contexts (see for an overview Metz, 2022), animal ethics (see Purves & Delon (2018) and Monsó et al. (2018)), technological manipulation (Nyholm, 2022) and even climate ethics and the discussion about responsibility to future generations (Campbell & Nyholm, 2015; Kauppinen, 2014; Scheffler, 2018).

[3] An agent, according to this way of thinking, is any system or entity that pursues some goal or set of goals in a way that is responsive to its environment. This allows for the possibility of more or less sophisticated forms of agents, which can be differentiated with reference to further capacities, such as capacities for learning, degrees of flexibility across domains, creativity, and so on. See Russell and Norvig (1995/2020).

## 3 Conceptual Clarification: Meaning in Life

What do contemporary ethicists have in mind when they say that a life or some activity is more or less meaningful? On a semantic level, the field essentially understands 'meaning' as referring to something that is good for its own sake, which can be exemplified by a human's life, or some aspect of their life, to a variable degree. That sort of personal meaning is, for many authors, opposed to a purpose that could be conferred on humanity by something external to it, such as God as conceived of in the Abrahamic faiths. Many ethicists thus distinguish between meaning 'in' a life, by which they mean a non-instrumental value that makes an individual's life more desirable in a distinctive way, and the meaning 'of' life: a cosmic end that might be ascribed to humanity or the universe as a whole (e.g. Wolf, 2007, p. 63; Seachris, 2013, pp. 3–4).

On a normative level, the essential claim of the field is that meaning is neither identical to, nor fully subsumable under, the standard axiological parameters (Metz, 2013; Wolf, 2010). Many, for instance, contrast meaning with narrow self-interest. This means that, quintessentially, for a person to acquire meaning in their life, they must focus not solely on themselves, or at least not their own subjective well-being, but instead orient their life 'outwardly'. This can be realised through many different activities. However, many ethicists seem to agree on typical examples, such as rearing children with wisdom, being in a well-functioning romantic relationship, volunteering for a charity, demonstrating a refined skill to others, advancing knowledge through science, or creating works of art (see Landau, 2022, introduction).

Given such examples, there have been attempts to subsume the main sources of meaning in life under the categories of 'the True, the Good and the Beautiful' to pinpoint the main directions a meaningful life could take (see Metz (2011)). This does not imply that different facets of self-interest cannot play a role in a meaningful life (see for an overview of the options Rüther and Muders (2016)). In fact, one of the main protagonists of the field, Susan Wolf, has proposed an influential hybrid theory in which not only objective values must be present to create meaningfulness, but also a subjective counterpart, which she describes as love of or engagement with what is valuable (see Wolf (2010), and a summary of the view in Wolf (2010), also Johansson and Svensson (2022)). Even on such hybrid theories, meaningfulness is at least partly an autonomous notion, and not identical with the narrower category of self-interest.

What about other standard normative parameters, such as morality? The relationship between meaning and morality is complicated, but most ethicists in the field hold that certain moral deeds or omissions will have a significant effect on how meaningful a life is (see for the different options Kipke and Rüther (2019)). Some, for instance, claim that a person letting others suffer needlessly or treating them merely as a means to their own pleasure is not only morally questionable, but also lacks meaning, or is even meaning-reducing (see the discussion about the concept of 'anti-meaning' in Campbell and Nyholm (2015), Nyholm and Campbell (2022) and Scripter (2022)).

Even in the strongest attempts to link meaning and morality, however, the common ground of the field is that an axiological residue remains. Morality might have some significant effect on meaningfulness, but it is not sufficient in order to describe all aspects of a meaningful life. This 'further aspect' is controversial. Naturally, many stress an orientation towards 'higher values', as indicated by the expression 'the True, the Good, and the Beautiful', but how this can be put more systematically is one of the crucial questions of the field. Here, the options in the debate reflect the options that we are familiar with from more general discussions in moral philosophy. There are consequentialist approaches which stress that meaningfulness is tied to outcomes in meaningful areas (e.g. Bramble, 2015; Singer, 1996; Smuts, 2013), whereas what might be called deontological approaches typically stress certain facets of meaningful actions, such as someone's intentions to orient themselves towards meaningful endeavours (Metz, 2013; Rüther, 2023; Wielenberg, 2005; Wolf, 2010).

More can be said—for example, about monistic or pluralistic approaches (Taylor, 1999; Thomas, 2005), the nuances among more or less meaningful lives and activities (Levy, 2005), or the intersection between psychological and philosophical research on meaningfulness (Schnell, 2021). However, for the sake of the present discussion, it will suffice simply to note that meaning involves at least a form of partly autonomous, non-instrumental value in a person's life, that comes in degrees, and that involves an orientation towards values beyond oneself, such as within the realm of 'the True, the Good, and the Beautiful'.

## 4 Why Consider Meaningfulness in AI Ethics? Five General Arguments

What we have said so far leaves open the question of why it might be fruitful to connect contemporary debates on meaningfulness within philosophy more generally to AI ethics in particular. We therefore now wish to offer five general, schematic arguments for why AI and its different uses raise philosophical questions, and put pressure on, widely shared ideas about what is involved in living a meaningful life. Some of these five arguments generalize ideas/arguments that have been discussed in more specific versions in the existing literature. However, none of the contributions in the debate that we discuss below have attempted to offer a set of general arguments for why AI ethics should incorporate discussion of the idea of meaningfulness in life in the way that we do here. Accordingly, a key contribution of this article is to articulate these general arguments for why AI ethics, broadly conceived, should concern itself with the notion of meaning in life.

A first argument departs from the observation that, according to at least one common definition of AI, artificial intelligence is created in order for technologies to take over—either partly or fully—tasks that humans previously used to perform with the help of their natural intelligence (Coeckelbergh, 2020; Müller, 2020; Gordon & Nyholm, 2021). If those tasks are things that we find meaningful—and we hand them over to AI technologies—then we give away tasks that help to make our lives meaningful. Accordingly, unless there are other things we could do instead,

which are also meaningful, we might thus create what might be called *meaningfulness gaps* or gaps in meaning.[4]

That first type of argument can also be flipped around into a second argument: if there are activities that we use our intelligence to engage in, but those are activities that we find meaningless, and AI systems can take over those activities and thereby free up time for us to engage in other more meaningful activities instead—well, then the AI could be seen as a meaning-booster or meaning-enabler. This requires two things: first, that there are certain activities we now engage in that involve a kind of opportunity cost in relation to other more meaningful things we could be doing instead; and secondly, that AI technologies could take over those less meaningful activities while not taking over any of the activities that we do find it meaningful to engage in ourselves.

A third argument is based on another possible way of thinking about AI and how we relate to AI technologies: namely, the idea that we might expand what we are able to do or what we are able to achieve (as individuals or as groups) by using new AI technologies (Vold, 2015; Hernández-Orallo & Vold, 2019; Smids et al., 2020). If what we become able to achieve (as individuals or as groups) is of value and something it is meaningful to achieve, then the introduction of AI technologies might create opportunities for doing meaningful things. Or, alternatively, if we see ourselves as acting through or via the AI technologies we create, and we see the AI technologies as extensions of our own minds or extensions of our own agency (Clark & Chalmers, 1998; Vanzura, 2021)—and, moreover, we think that the things we do without our new extended minds/agency are meaningful—then this might be yet another way in which AI technologies create opportunities for meaning or even generate new forms of meaning in life.

A fourth argument is connected to another way of thinking about AI technologies—or about a sub-set of AI technologies, such as social robots and advanced chatbots—namely, as a form of artificial persons (Smith, 2021; Wareham, 2020). If we think of relationships with other persons (e.g. with our fellow human beings) as a source of meaning in life, and we think that some AI technologies (social robots or chatbots, etc.) can be a kind of person, then there is a potential for meaningful relationships with these AI persons. Of course, at present, most AI researchers (including computer scientists, philosophers, and others) are highly sceptical about the idea of AI technologies as being some form of persons (see Nyholm, 2023: Chapters eight and nine). However, some take the possibility of AI persons seriously, and it is possible that more and more people (including more AI researchers) will take this idea seriously in the future.

[4] This is parallel to one way of explaining the argument about why AI might create responsibility gaps: if AI systems take over tasks that humans need their intelligence to perform, and those are tasks for which human beings had previously been responsible, then this might open up a gap in responsibility, since the tasks that intelligent and morally responsible humans were previously responsible for are taken over by artificially intelligent but not morally responsible AI technologies (Nyholm, 2023: Chapter Five). In the same way, if AI systems take over tasks that were meaningful for human beings, then unless those human beings can do other things that are meaningful instead, a gap in meaning may have arisen as a result.

Here, of course, a fifth argument presents itself, since the opposite view is also possible: while AI technologies such as social robots or chatbots might appear to provide opportunities for meaningful relationships, they may in fact not be entities with which we can have truly meaningful relationships (cf. Misselhorn, 2021; Turkle, 2011: Chapter Seven). We may instead, according to this perspective, fall victim to deception and false hopes of meaningful relationships with these artificial agents. We will address more specific versions of this worry below.

In summary, there are at least five significant ways of thinking about AI and its relationship with meaningfulness (in very general terms), as illustrated in this text box:

Text box 1:

Five general reasons for taking questions about meaning in life seriously within AI ethics:

- AI technologies might be considered as technologies that take over tasks that humans need their intelligence to perform—which means that if those tasks are meaningful, then AI might take meaningful activities away from us, potentially creating gaps in meaning
- AI might take over meaningless tasks and free up time during which we could engage in meaningful activities
- AI might function as extensions of our minds or our agency—meaning that if we become able to do new things—or achieve new things—that are meaningful with our new 'extended minds' or our extended forms of agency, then AI might enable us to do or achieve meaningful things we could not achieve without AI
- AI might involve technologies that can be viewed as a form of artificial persons (e.g. social robots or chatbots)—so that relationships with these AI persons might potentially be seen as meaningful forms of relationships
- AI technologies might be a form of apparent, but not real or not sufficiently real artificial persons—so that relationships with these apparent AI persons are actually much less meaningful than relationships with the real persons or animals with whom we could have more meaningful relationships instead

## 5 The Concept of Meaning in the AI Ethics Literature

### 5.1 Some Preliminaries

Various things can be assessed when it comes to judgments about meaningfulness, including, but not limited to, whole lives, parts of lives, activities within lives, ways of relating to oneself, relationships with other people, relationships with non-human animals, relations to nature and the universe as a whole, religious practices and so on. We will not discuss every possible object of meaningfulness judgments here, but will focus on three main potential loci of meaning: self-development, work, and human relationships. There are two reasons for the selection. First, there has been at least some academic debate about all three areas. Second, and more importantly, all three are intuitive and clear candidates as sources of meaning.

A further key thing to note regarding what follows below is that one can distinguish between more or less direct contributions to the subject of meaning in life and AI in the

existing literature. The most direct type of contribution are papers, books, or other contributions that explicitly set out to discuss the effect of AI on meaning in life—a literature that is growing, but where not a great deal has yet been written. A second, indirect kind of contribution to the literature on AI ethics are contributions that are not explicitly primarily about the effect that AI has on meaning in life, but which nevertheless are closely enough related to this topic that it makes sense to engage with them in research on AI and meaning in life. One example is papers that talk about certain forms of technology where it is debatable whether we should count those technologies as forms of AI, but where the discussion nevertheless can be seen as having implications for how we should understand the relationship between AI and meaning. Another example is papers about AI and its effect on human life, which do not necessarily explicitly discuss the concept of meaning in life, instead discussing some other concept instead, but where that concept under discussion has a clear bearing on the issue of meaning in life.

### 5.2 AI and Meaningful Self-Development

We start here with AI, meaning, and self-development, because one of the only two book-length treatments of AI and meaning in life that we are aware of—namely, a book published in German by the philosopher Richard David Precht—starts by announcing (in our translation) that it 'is an essay by a philosopher, who asks himself, what artificial intelligence does to our human self-conception and how it [AI] will influence our future self-realization' (Precht, 2020, p. 6). This book—the title of which can be translated as *Artificial Intelligence and the Meaning of Life*: *An Essay*—goes on to associate the development of AI with a transhumanist agenda that Precht thinks is prominent in Silicon Valley—an agenda towards which Precht takes a very sceptical stance (cf. also for more scepticism in this regard Nida-Rümelin & Weidenfeld, 2022). The book rests on a subjectivist conception of meaning in life—that is, the view that whether life or some aspect of life is meaningful depends on whether it is experienced as meaningful by the person whose life it is—and Precht offers a sceptical take on whether AI and other advanced technologies coming out of Silicon Valley will help to promote, or are intended to enable people to experience, a sense of meaning in life.

Precht argues that the development of AI technologies is not only part of a suspicious transhumanist agenda, but also that the development of these technologies is mostly driven by an excessive form of capitalism: that is, AI technologies are not developed, according to Precht's analysis, because they will improve people's lives, but rather to maximise the profits of the tech companies that make use of these technologies. The connection to meaning in life here, then, is related to the broadly Marxist idea that excessive capitalism leads to a sense of alienation and a diminishment of the sense of meaning in life. The book is also ultimately a sustained articulation of a deep scepticism regarding the motives and ideologies of various leading figures in the tech world (e.g. Ray Kurzweil), as well as in academic discussions of AI and AI ethics (e.g. Nick Bostrom).

As we see things, Precht's book is interesting, but it is ultimately more of a broad criticism of the Silicon Valley mentality than an engagement with the specific questions we are interested in about the relationship between meaning in life and AI,

which we think should be given more attention in AI ethics. Furthermore, the book makes no effort to determine whether there are any circumstances under which AI could have a positive impact of any kind on meaningfulness in life. It is thus a somewhat one-sided, and negative, analysis.

A related contribution to the literature discussing similar questions—viz., technologies for enhancing humans and effect on meaning in life—which does so by both examining threats to meaning and opportunities for meaning created by technologies that can be used for self-development, is a paper by John Danaher (2014), in which he discusses what he calls the 'hyperagency' worry related to technology and meaning in life. Danaher is one of the most prominent contributors to the literature on AI and meaning in life, and is, among other things, the author of the second book-length treatment of the topic of AI and meaning in life mentioned above. Danaher's (2019a) book is called *Automation and Utopia,* and focuses on both the subject of AI and self-development that we are considering in this sub-section, and the subject of AI and work, which we will address in the next sub-section. The earlier paper about 'hyperagency' was not specifically written in terms of AI's effect on meaning in life. The discussion is rather about technologies for 'human enhancement' and their impact on meaning in life. If we are to believe Precht's conclusion that AI is at least sometimes, if not often, associated with transhumanist ideas, however, then Danaher's (2014) paper is at least clearly an indirect contribution to the subject of AI and meaning in life, since human enhancement is the main goal of those who are interested in transhumanism.

What is the main idea in Danaher's paper, and what does the expression 'hyperagency' refer to? 'Hyperagency' refers, roughly speaking, to the idea that advanced technologies extend the range of things in life over which we can exercise agency, and/or that are under our human control. Some critics of human enhancement have argued that this is a threat to meaning in life, because some parts of meaning in life derive from aspects of life that we cannot control or exercise agency over, but which are instead are a type of gift or things we should accept as they are (e.g. Sandel, 2007 and in this line also: Hauskeller, 2011). Danaher's response to this is that having greater powers and more extensive agency can enable us to do more good—and Danaher argues that doing good is part of what makes life meaningful.

This could be relevant to AI and meaning, because AI might widen the range of things in life over which we have control, or over which we can exercise agency. This might be a threat to meaning in life if it is meaningful to lack control or agency with respect to an important range of goods in life. On the other hand, this might enhance meaning in life if having more control and a wider range of things over which we can exercise agency enables us to do more good, which could be seen as being part of what is meaningful in life (e.g., under the trias of meaningfulness goods relating to 'the True, the Good, and the Beautiful').

In summary, then, an interesting difference between the two authors who have produced the only book-length treatments of meaning and AI is that one of them (Precht) approaches this topic via a form of ideological critique of the mindsets of those who are most enthusiastic about AI, whereas the other (Danaher) approaches the topic via arguments about what exactly technologies can and cannot do in relation to the more specific goods or specific constituents associated

with the general idea of meaning in life. We are more interested here in the type of question that Danaher explores, since it allows a more nuanced evaluation of meaningfulness and AI technologies.

In this regard, though, there is still a long way to go. Little has been published on this topic so far. But there are several directions which might be fruitful. For one, it would be helpful, in our view, to put more weight on the identification of criteria for meaningful self-development. Precht mentions our general feeling, or sense, of meaningfulness, which he sees as compromised through AI technologies; Danaher seems to emphasise human agency and its extension. But what else is important?

A detour into traditional virtue ethics, which is concerned with self-development and character, might be helpful here to identify further resources (Vallor, 2015, 2016). Worth mentioning here, as an indirect contribution to the literature on AI ethics and meaning, is the recent, very readable book *Self-Improvement* by Mark Coeckelbergh (2022). Among other things, it critically examines the modern trend towards self-optimization ('a 11 billion dollar industry', p.2), which is further fuelled by AI systems. Notably, Coeckelbergh does not explicitly refer to the debate on meaningfulness, but some of the lines of criticism explored in his book can be interpreted in this direction. For example, according to Coeckelbergh, the urge towards self-optimization entails a dangerous obsession with one's own self, which leads to a 'spiritual narcissism' (p. 29). Through such self-centeredness, one loses the possibility of relating to the environment and other people. For Coeckelbergh, this is a general component of the good life (see his ch. 6 on the 'relational self'), but against the backdrop of the meaning-in-life discourse and its emphasis on 'the True, the Good, and the Beautiful', an orientation towards other people can also be interpreted as a specific component of meaning, more specifically of the good. Relatedly, Coeckelbergh highlights that the modern penchant for self-optimization also involves different forms of outsourcing (e.g. of deliberations and decisions) to AI systems in order to increase one's own productivity and performance. However, such outsourcing is a problem in Coeckelbergh's view, because it deprives the individual of the opportunity to acquire and train important skills (see pp. 74–75). Coeckelbergh does not explicitly refer to the discourse of meaning here either. However, his virtue-ethical references to the formation of human abilities are highly relevant in this context, at least if one is willing to accept capability-development as being meaning-conferring.

In our own view, another particularly interesting path related to self-development, AI, and meaning might involve reflection on the abilities and capacities which have to do with morality. A key issue here is whether AI technologies can work as a form of moral enhancement by providing recommendations about how we can best live in accordance with our own moral values (e.g. Savulescu & Maslen, 2015; Klincewicz, 2019; O'Neill et al., 2022). An interesting question is whether acting in accordance with our moral values not because we work out how to do so ourselves, but because AI technologies tell us how to do this would somehow be less meaningful. Or would it have no significant effect on how meaningful self-development related to our own moral values would be? This, we think, is one additional example of an interesting question that should be discussed further.

### 5.3 AI and Meaningful Work

As mentioned above, Danaher is the author of the only other monograph on the general topic of AI and meaning that we are familiar with (see Danaher, 2019a), and his book has a different focus than Precht's. It is about the idea that the development of AI might result in widespread 'technological unemployment', up to a point where we might even soon be living in a 'world without work', as Danaher puts it. The question then is whether life in such a world would be meaningful. Danaher's discussion about this exemplifies two of the general argument types we identified above. Recall that we noted that if AI takes over tasks that we regard as meaningful from us, then what we are calling a meaningfulness gap might arise unless there are other things we could do instead that are equally or more meaningful (see Section 3, Argument 1). Danaher considers this type of argument. However, he thinks that for an overwhelming amount of people, there are 'reasons to hate your job', to use Danaher's striking phrase (see Ch. 3 in Danaher, 2019a). Having AI technologies take over your work tasks might enable you to do more meaningful things instead—for which reason AI would then serve as a meaning-booster or meaning-enabler. In our terms, then, Danaher's (2019a) overall line of argument is ultimately of the second of the five argument types we articulated above, rather than the first.

It is important to note, however, that Danaher does not claim that everyone has reason to hate their job, or that all forms of work are meaningless. Danaher agrees with the widely accepted idea that for some, or even many people, work can be an important source of meaning (Danaher & Nyholm, 2021), but more on that below. What we will first highlight is Danaher's discussion of how one might fill apparent meaningfulness gaps if work is taken over by AI technologies.

Danaher discusses two strategies: the 'cyborg' and 'virtual worlds' solutions. The first idea resembles Elon Musk's idea behind 'Neuralink' (Newitz, 2017). It is the idea that in order to keep up with, and be able to compete with, advanced AI systems, we may need to merge with technologies, for example by making use of brain-computer interfaces that enable us to do things we cannot do with our ordinary brains. The second idea focuses on what Danaher thinks we could do that would be meaningful if we did not work anymore. Here, we get to his ideas about 'virtual worlds'.

Such worlds could mean entering the metaverse—a computer-simulated world—doing meaningful things or apparently meaningful things within the simulated virtual reality. Notably, David Chalmers (2022) explicitly endorses this as a good idea in his recent book *Reality+*. According to Chalmers, virtual reality can be as real as normal reality, and activities done within virtual realities can be just as meaningful as corresponding activities within our regular reality.

Importantly, however, Danaher (2019a) does not only talk about virtual reality in the computer-generated sense when he discusses the escape into virtual worlds. Danaher also discusses the creation of games, from which we might derive meaning in life. In his view, playing more or less elaborate games can be meaningful (cf. Suits, 1978), and if AI technologies take over our work—indeed, even if AI takes over meaningful work—this might free up time for us to play meaningful games instead. This would be a kind of 'virtual world', because it is a socially constructed

activity, with 'trivial' goals that have no real significance outside the game—and yet this could be meaningful, or so Danaher argues.

But are meaningful games the only way to lead a meaningful life in a world where people are threatened by technological unemployment? Sebastian Knell and Markus Rüther (forthcoming) raise some doubts about this, and argue that even if full automation is probable, there is still plenty of room for meaningful endeavours. More specifically, they argue for a 'humanistic perspective', which connects meaningful actions not—like Danaher and many others—to a certain kind of active contribution, mainly in the realm of 'the True, the Good and the Beautiful', but also to more receptive modes of being, which they summarise as a modern version of the Aristotelian idea of the *vita contemplativa*.

Notably, Danaher's approach and also the response by Rüther and Knell rest on the assumption that AI technologies will completely take over all work tasks. This contrasts with an approach that instead assumes people will continue working—or that many people will continue working—but that people will increasingly be working alongside robots and other AI technologies. The question, then, is whether such work can be as meaningful as the kind of work where humans need to use the full range of their abilities—including their intelligence and ingenuity—to do the work. Danaher has written about this elsewhere, together with Sven Nyholm, who has also written about the topic together with Jilles Smids and Hannah Berkers. In those discussions, the question is whether a new type of work situation in which humans still work but many tasks are handed over to AI technologies (including robots) will leave a sufficient range of meaningful tasks for the humans who work alongside these technologies (Danaher & Nyholm, 2021; Smids et al., 2020).

Meaningful work, Smids et al. (2020) argue, typically involves the following five aspects: (1) pursuing a valuable purpose, (2) social relations and collegial interactions, (3) exercising skills and self-development, (4) self-esteem and recognition and (5) work-related autonomy. Related to points (1) and (3), Danaher and Nyholm (2021) argue that meaningful work involves opportunities for human achievement. Achievement, in this view, is understood in terms of a combination of Gwen Bradford's (2015) view of achievement and Hannah Maslen et al.'s (2020) views about the basis for praiseworthiness, so that one is praiseworthy for achievements that have the following features: (i) the output of one's work is valuable, (ii) one plays an important causal role in the production of this output, (iii) one needs to put in an effort, and (iv) one does this voluntarily and enthusiastically (Danaher & Nyholm, 2021, p. 231).

The key question here is if AI technologies (including, but not limited to, robots and text-producing large language model technologies) are integrated more and more into work activities, would there still be enough room for the various noted above-goods associated with meaningful work? Would human beings have opportunities for work-related achievements—to ask the question in Danaher and Nyholm's (2021) article? And would human beings have access to the five goods of meaningful work identified by Smids et al. (2020) (see also Bankins & Formosa, 2023)?

Danaher and Nyholm (2021, p. 229) offer an argument to the effect that AI technologies may create achievement gaps—a particular version of the more general idea of meaningfulness gaps—in many workplaces, for many people. Why? Because the role of many

human beings in workplaces may be reduced to doing what AI systems tell them to do, or to prompting, maintaining, or supervising AI technologies; indeed, their most meaningful work tasks may simply be taken over by AI technologies, such as those tasks that previously involved playing a key causal role in the production of valuable outcomes, while exercising significant effort in a voluntary and enthusiastic way (cf. Tigard, 2021).

Smids et al. (2020), in contrast, present a somewhat less dire picture, but nevertheless argue that AI technologies can threaten all five goods related to meaningful work that they identify. Their picture is less grim than the picture Danaher and Nyholm (2021) present, since Smids et al. (2020) also investigate ways in which all five of the goods of meaningful work that they discuss could be compatible with, or even boosted by, work that involves working with AI technologies. As Smids et al. see things, it is possible that as we are working alongside AI technologies, we might become better able to pursue valuable goals. Such technologies might not necessarily affect collegial relationships negatively. There is even a question, discussed by Nyholm and Smids (2020), as to whether robots could be a new form of good colleagues in the workplace, and there are already people who experience the robots they work alongside as valuable members of the team, but more on that in the next section. As Smids et al. (2020) see things, working alongside AI technologies does not necessarily mean that there is less room for exercising skills and self-development—for example, because this new type of work situation might require workers to learn and exercise the new skills needed to be able to work together with the new AI technologies. Smids et al. also think that this could give those workers a foundation for self-esteem, and other workers reason to recognise their development. Finally, Smids et al. think that there are contexts in which working together with AI technologies could be compatible with work-related autonomy. This being said, Smids et al. also, as noted above, highlight various 'threats' to meaningful work created by AI and robots, and not only 'opportunities' for meaningful work in such work situations, to use the terms they employ to present their overall argument (Smids et al., 2020, pp. 515–516).

In general, then, AI, meaningfulness and work have been related to each other in the existing literature in at least four ways, as shown in this text box:

Text box 2:

Four possible ways in which AI might impact the future of meaningful work:

First, if work is meaningful, and AI takes over our work, then a gap in meaningfulness might be created

Second, if work is meaningless, and AI takes over this work, then opportunities for doing other, more meaningful things might be created

Third, if we still work, but AI takes over the meaningful aspects of work, then AI will make our work meaningless or at least less meaningful

Fourth, if working together with AI technologies opens up new opportunities for taking on tasks that are meaningful—that is, tasks that are related to the goods of meaningful work—then working with AI technologies can boost or enable meaningful work

### 5.4 AI and Meaningful Relationships

In June 2022, the Google engineer Blake Lemoine made headlines when he went to the press to speak about his belief that the AI-driven large language model LaMDA had become a sentient person (Tiku, 2022). The way that Lemoine described his conversations with LaMDA made it seem that he felt that he had come to have what might be called a meaningful relationship with this chatbot. Most commentators—including other representatives from Google—were quick to contradict Lemoine's claims about the capabilities of LaMDA, and argued that this language model was as much a conscious and sentient person as a toaster is. Lemoine was put on administrative leave from Google, perhaps mostly because he had shared transcripts that Google did not want him to share, but surely also because Google as an organisation was embarrassed by the whole incident. Lemoine is not the only person in the world of technology who thinks that AI technologies might either already be, or that they might soon become, conscious and sentient. In August 2021, for example, Elon Musk claimed that the self-driving cars created by Tesla were 'basically semi-conscious robots on wheels'.[5] In February 2022, Ilya Sutskever, the chief scientist of the OpenAI research group, tweeted that 'it may be that today's large neural networks are slightly conscious.'[6] Relatedly, the philosopher Thomas Metzinger (2013) thinks that it is possible to create robots that feel pain. Similarly, the philosophers Schwitzgebel and Garza (2015) think that it will be possible to create AI technologies with humanlike mental and social capabilities in the future. This raises the question of whether we could have meaningful relations with such AI technologies, and/or whether they will affect our relationships with other humans.

We can also ask whether we could have meaningful relationships with AI technologies independently of whether they have humanlike consciousness and are sentient beings. Danaher (2019b), for example, has argued in favour of an 'ethical behaviourist' position: if an AI technology—such as a robot or a chatbot—consistently behaves like a friend or romantic partner behaves, then this would be enough, Danaher suggests, for the AI technology to qualify as a friend or romantic partner. The human–robot interaction researcher De Graaf (2016) has argued, in a similar way, that when it comes to the goods associated with relationships, 'performance' is what ultimately matters. Janina Loh (2019) has defended what might be seen as an even more extreme view. Loh argues that we should not look at the capabilities of the technologies, but rather at the ways in which people relate to technologies. If somebody has become attached to an object—which might be an AI technology, such as an advanced robot—then we should not view this as a 'shortcoming' or 'failing' of the person, Loh argues. Instead, we should regard this as a 'capability', which many other people may not possess. Moreover, in recognition of the value of inclusivity, we should value relationships between humans and technologies (including objects without minds) as something to be celebrated as part of human diversity.

[5] See Musk's presentation of the 'Tesla bot' in this video for that quote: 'Elon Musk REVEALS Tesla Bot (full presentation)', https://youtu.be/HUP6Z5voiS8 (Accessed on September 23, 2022).

[6] https://twitter.com/ilyasut/status/1491554478243258368 (Accessed on September 29, 2022.).

If this view were put in terms of meaningful relationships, it is likely that Loh would conclude that such relationships can be positively meaningful.

Others, however, have defended radically opposing positions on this issue (for an overview, see Weber-Guskar, 2021). Authors such as Sullins (2012), Hauskeller (2017), and Nyholm and Frank (2017) have argued that until, or unless, robots have minds that are relevantly similar to human minds, and a free will relevantly similar to our human free will, we cannot have the kinds of relationships with robots and other AI technologies that we can have with fellow human beings. Our relationships with them cannot, from this point of view, be meaningful in the ways that our relationships with human beings—or indeed with some animals—can be. The core idea here is that meaningful relationships are had with beings that have minds and/or a free will, and that robots and AI technologies lack the relevant kinds of minds and free will.

Another relevant argument is presented by Catrin Misselhorn (2021), who argues that if we seek the kind of recognition that we typically seek from other human beings within meaningful social relationships when we interact with robots, then we are in effect treating ourselves as if we are objects, just like the robots are objects. As Misselhorn sees things, if we seek recognition from a robot or other AI technology, we behave as if we ourselves do not have minds, and as if we have no human need for a 'meeting of minds', as the phrase goes, with fellow thinking and feeling beings.

Misselhorn brings up an example from the 2019 documentary *Hi AI!* to illustrate her point. In that documentary, a man from Texas named Chuck travels to California in his camper van to collect his new partner, a sex robot called 'Harmony'. In one scene of the documentary, Chuck tells Harmony the robot about how he was sexually abused as a child. Misselhorn describes this as tragic, if not pathetic, because this is the kind of thing you would normally tell a human being, who has a mind and is able to empathize with you. The relationship between Chuck and Harmony—Misselhorn would probably conclude if she put things in terms of meaningfulness—is not a meaningful relationship, at least not in the way that a relationship with a human being who could empathise with someone could be.

There are thus radically different views about whether we can have good and meaningful relationships of friendship and love with AI technologies in the literature. But what about other kinds of relationships? As mentioned above, Nyholm and Smids (2020) discuss another kind of relationship—collegial relationships—and they do so explicitly with an eye to the common idea that having good colleagues is part of what can make work meaningful. The question they raise, therefore, is whether a robot can be a good colleague. As they note, there are people who do become attached to robots and other technologies they work alongside, and who regard them as members of the work team—for example, some American soldiers on the battlefield in Iraq became extremely attached to a bomb-disposal robot ('Boomer') that they worked with.[7]

Nyholm and Smids argue that it is easier for a robot to live up to the criteria of being a good colleague than it is for any AI technology to live up to the criteria for qualifying

[7] For more on the Boomer example, see Carpenter 2016.

as a good friend or romantic partner, because those latter criteria are in certain ways more demanding. They even argue that on a behavioural level, a robot could behave in many of the ways that a good colleague should behave. But would being a good colleague on a behavioural level—i.e., would behaving like good colleagues should behave—be enough to make a robot into a colleague with whom one can have a meaningful work relationship? Again, it might make a certain amount of sense to think that it is easier for a robot or AI technology to have a meaningful work-related relationship with a human in the workplace. Yet it might be doubted whether the most meaningful types of work relationships could be realised between humans and AI technologies.

In recent work on 'collegial relationships', Betzler and Löschke (2021) argue that two of the most important values within collegial relationships are work-related solidarity and recognition. Could a robot have solidarity with their human colleagues? And could a robot recognise the achievements and excellence of a human in the workplace, in the way that a fellow human colleague is able to do so?

As we saw above, according to Misselhorn, we treat ourselves like objects—in effect, we degrade ourselves—if we seek recognition from technologies that lack human minds. It might also be plausibly argued that a robot or any other currently existing AI technology could not show true solidarity with a human being. It is not even clear what that would mean. So, if the most important values related to collegial relationships, as Betzler and Löschke argue, are solidarity and recognition, and those work-related values are important for meaningful work-relationships, then the conclusion that follows seems to be that we cannot have the most important and most meaningful forms of work-related relationships with AI technologies.

Notably, a similar argument could be made within the context of arguments presented in recent work on southern African Ubuntu ethics by analytic philosophers interested in relationships (including human-technology relationships) and meaning in life. Cindy Friedman (2022), for example, argues that Ubuntu ethics presents an ideal of relationships between human beings that is not (yet) possible to realise within human–robot relationships. Why? Because the AI in contemporary robots is so rudimentary that we cannot flourish by interacting with robots in the ways that we can flourish as human beings within relationships with other humans. If one puts that conclusion about human–robot interaction together with the general idea defended by Thaddeus Metz (2020) and Aribiah Attoe (2020) that Ubuntu philosophy presents a compelling vision of meaningful relationships, then it is implied that human–robot relationships—or relationships between humans and AI technologies more generally—cannot be meaningful in the ways that human–human relationships can be.

In summary, while there are those who defend views about the values that can be realised in human-AI relationships that might support the idea that relationships with AI technologies might be meaningful, there are also many philosophers—and perhaps many more philosophers—who defend views about what is involved in meaningful relationships that support the conclusion that it is not possible to have relationships with AI technologies that are meaningful in the ways that relationships with our fellow human beings can be.

A further question that could also be asked is whether AI could mediate our relationships with other human beings—or with animals—in a way that would boost or enable meaningfulness, or whether placing AI between ourselves and others will typically take meaningfulness away from our relationships. In other words, rather than asking whether we could have meaningful relationships with AI technologies, we could also ask whether AI technologies could somehow work as a booster or enabler of meaningful relationships between, or among, human beings. That is a key question to ask in this context, but we will not discuss it here (for valuable related discussion, see Kaliarnta, 2016 and Elder (2018)). Instead, we will proceed to our general conclusions.

## 6 Meaning in Life in AI Ethics—Summary and Outlook

We have tried to show at least three things in this paper. First, we have noted that there is a growing debate on meaningfulness in some sub-areas of AI ethics, and particularly in relation to meaningful self-development, meaningful work, and meaningful relationships. Second, we have argued that this should come as no surprise. Philosophers working on meaning in life share the assumption that meaning in life is a partly autonomous value concept, which deserves ethical consideration. Moreover, as we argued in Section 4 above, there are at least five significant general arguments that can be formulated in support of the claim that questions of meaningfulness should play a prominent role in ethical discussions of newly emerging AI technologies. Third, we have also stressed that, although there is already some debate about AI and meaning in life, it does not mean that there is no further work to do. Rather, we think that the area of AI and its potential impacts on meaningfulness in life is a fruitful topic that philosophers have only begun to explore, where there is much room for additional in-depth discussions.

The following Table 1 provides an overview of some of the key contributions to the existing literature that we have reviewed above.

We will now close our discussion with three general remarks. The first is led by the observation that some of the main ethicists in the field have yet to explore their underlying meaning theory and its normative claims in a more nuanced way. This is not only a shortcoming on its own, but has some effect on how the field approaches issues. Are agency extension or moral abilities important for meaningful self-development? Should achievement gaps really play a central role in the discussion of meaningful work? And what about the many different aspects of meaningful relationships? These are only a few questions which can shed light on the presupposed underlying normative claims that are involved in the field. Here, further exploration at deeper levels could help us to see which things are important and which are not more clear, and finally in which directions the field should develop.

The second remark we wish to make by way of conclusion, which may be rather obvious, is that the three above-discussed topics of self-development, work and relationships are not the only topics worth pursuing when it comes to questions of AI and meaning. There might be a wide area of highly interesting unexplored territory. We imagine, for example, that issues connected to the theme ‘AI and sustainability’

**Table 1** Overview of some of the key contributions to the existing literature

| Highlighted aspects of life that can be assessed in terms of its meaningfulness: | Key contributions on AI as a threat to meaningfulness: | Key contributions on AI as an opportunity for meaningfulness: |
|---|---|---|
| Self-development: | * Precht (2020) on a supposed transhumanist (self-optimization) agenda and excessively capitalist ideology behind the development of many AI technologies, which Precht thinks is likely to counteract a subjective sense of meaningfulness<br>* Vallor (2015, 2016) on de-skilling by means of reliance on technologies instead of the development of one's own capacities<br>*Coeckelbergh (2022) on how technologies like self-tracking and the outsourcing of tasks to AI systems might lead to narcissism and block capability-development | * Danaher (2014) on how using technology to extend and expand our abilities might extend the range of things over which we exercise agency and control, where agency and control are seen as crucial to a good—and meaningful—life<br>* Potentially also: Savulescu and Maslen (2015), Klincewicz (2019) and O'Neill et al. (2022) on using AI technologies as a form of 'moral advisor', which could help one to become better able to live in accordance with one's (moral) values, which can be seen as meaning-boosting |
| Work: | *Smids et al. (2020) on possible threats to meaningful work related to the use of robots and AI in workplaces, related to the ideas of meaningful work as involving (1) pursuing a valuable purpose, (2) social relations and collegial interactions, (3) exercising skills and self-development, (4) self-esteem and recognition, and (5) work-related autonomy; for similar discussion, see Bankins and Formosa 2023<br>* Danaher and Nyholm (2021) on how the increased use of AI and automation might create 'achievement gaps' in the workplace, thus threatening the aspect of meaningful work that relates to human achievement | *Smids et al. (2020) on possible opportunities for meaningful work related to the use of robots and AI in workplaces, related to the ideas of meaningful work as involving (1) pursuing a valuable purpose, (2) social relations and collegial interactions, (3) exercising skills and self-development, (4) self-esteem and recognition, and (5) work-related autonomy; for similar discussion, see Bankins and Formosa 2023<br>* Danaher (2019a) on the idea that by outsourcing work to AI and other forms of automation, we can free up time for other forms of meaningful ways of leading our lives (e.g. virtual realities and games)<br>* Knell and Rüther (forthcoming) on the idea that if AI takes over work, there might be more time for contemplative activities that might be meaningful to engage in |

**Table 1** (continued)

| Highlighted aspects of life that can be assessed in terms of its meaningfulness: | Key contributions on AI as a threat to meaningfulness: | Key contributions on AI as an opportunity for meaningfulness: |
|---|---|---|
| Relationships: | *Friedman (2022) on how, from a southern African 'ubuntu' perspective, frequently interacting with rudimentary AI technologies might undermine our capacity to be the best versions of ourselves in our relations with other people<br>*Misselhorn (2021) on how relationships with AI technologies would undermine the potential for the form of mutual recognition that is crucial in valuable and meaningful relationships<br>*Sullins (2012), Hauskeller (2017), and Nyholm and Frank (2017) on how their lack of an 'inner life' and free will would make robots and other AI technologies poor relationship partners, which cannot realise some of the key aspects associated with meaningful human relationships (such as a 'meeting of minds' and a mutual voluntary commitment) | *De Graaf (2016) and Danaher (2019b) on how AI technologies that behave like human friends behave would enable us to have valuable and meaningful relationships with these technologies<br>* Loh (2019) on how relationships with robots/AI technologies depend on some people's 'capacity' to become attached to technologies, which Loh thinks should not be seen as a shortcoming, but as a source of meaningful relationships |

may also open up many opportunities for studies on meaningfulness (see for the general theme and its topics: Coeckelbergh (2021) and van Wynsberghe (2021)). A first attempt has been made by Nyholm (2021), who connects the discussion about moral duties towards future generations with the topic of anti-meaning, and briefly relates this to AI risks. But much more can be done on this issue.

Finally, a third general remark can be made about the priority of considerations that have been made in the field so far. In our view, the status of the debate makes it understandable that it almost solely concentrates on meaningfulness and its implementation in different areas in AI ethics. Nevertheless, we also think that if the field wants to proceed, it is also necessary to develop considerations that weigh meaningfulness against other value concepts that might play a role, and first and foremost against well-being and morality. This would be helpful for many reasons. One is that it will shed light on the place and relative importance of meaningfulness. Let us assume that some AI technologies are able to make one's life more meaningful. Is such a life also necessarily one in which people are better off (in terms of well-being), and in which people live together in more just and fair ways (in terms of morality)? Perhaps meaning ultimately has a dark side; perhaps it does not. It is not up to us make a final statement about this here, but we think that it is worth exploring that question and related questions in future work, in order to gain the full picture regarding how we should think about the relevance of meaningfulness in AI ethics.

**Acknowledgements** We would like to thank Roland Kipke, Thaddeus Metz, and Sebastian Muders for helpful comments on earlier stages of the manuscript. We also like to thank the two anonymous reviewers who especially helped us to make the manuscript more readable and pressed us to be detailed and nuanced in our presentation of the different arguments and positions.

**Author Contribution** The authors are listed in alphabetical order and contributed equally.

**Funding** Open Access funding enabled and organized by Projekt DEAL. Sven Nyholm's work on this paper is part of the research program Ethics of Socially Disruptive Technologies, which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

**Data Availability** Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## Declarations

**Ethics Approval and Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Competing Interests** The authors declare no competing interests.

## References

Attoe, A. (2020). A systematic account of African conceptions of the meaning of/in life. *South African Journal of Philosophy, 39*(2), 127–139.

Bankins, S. & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, 1–16. online first: https://link.springer.com/article/10.1007/s10551-023-05339-7

Betzler, M., & Löschke, J. (2021). Collegial relationships. *Ethical Theory and Moral Practice, 24*(1), 213–229.

Bradford, G. (2015). *Achievement*. Oxford University Press.

Bramble, B. (2015). Consequentialism about meaning in life. *Utilitas, 27*(4), 445–459.

Campbell, S. M., & Nyholm, S. (2015). Anti-meaning and why it matters. *Journal of the American Philosophical Association, 1*(4), 694–711.

Carpenter, J. (2016). *Culture and human-robot interaction in militarized spaces: A war story*. Routledge.

Chalmers, D. (2022). *Reality +*. W.W. Norton.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Coeckelbergh, M. (2021). AI for climate: Freedom, justice, and other ethical and political challenges. *AI Ethics, 1*, 67–72.

Coeckelbergh, M. (2022). *Self-IMPROVEMENT. Technologies of the Soul in the Age of Artificial Intelligence*. Columbia University Press.

Danaher, J. (2014). Hyperagency and the Good Life: Does Extreme Human Enhancement Threaten Meaning? *Neuroethics, 7*(2), 227–242.

Danaher, J. (2019a). *Automation and utopia*. Harvard University Pres.

Danaher, J. (2019b). The philosophical case for robot friendship. *Journal of Posthuman Studies, 3*(1), 5–24.

Danaher, J., & Nyholm, S. (2021). Automation, work and the achievement gap. *AI and Ethics, 1*(4), 227–237.

De Graaf, M. (2016). An Ethical evaluation of human-robot relationships. *International Journal of Social Robotics, 8*(4), 589–598.

Dignum, V. (2019). *Responsible artificial intelligence*. Springer.

Elder, A. (2018). *Friendship, robots, and social media*. Routledge.

Friedman, C. (2022). Ethical concerns with replacing human relations with humanoid robots: An Ubuntu perspective. *AI and Ethics*, 1–13. Online first: https://doi.org/10.1007/s43681-022-00186-0

Gordon, J-S., & Nyholm, S. (2021) Ethics of artificial intelligence. *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/ethics-of-artificial-intelligence/

Hauskeller, M. (2011). Human enhancement and the giftedness of life. *Philosophical Papers, 40*(1), 55–79.

Hauskeller, M. (2017). Automatic sweethearts for transhumanists. In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications* (pp. 203–218). MIT Press.

Heinrichs, B., Heinrichs, J-H., & Rüther, M. (2022). *Künstliche Intelligenz*. de Gruyter.

Hernández-Orallo, J. & Vold, K. (2019). AI extenders: The ethical and societal implications of humans cognitively extended by AI, *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 507–513.

Johansson, J., & Svensson, F. (2022). Subjectivism and objectivism about meaning in life. In I. Landau (Ed.), *The Oxford Handbook of Meaning in Life* (pp. 43–58). Oxford University Press.

Kaliarnta, S. (2016). Using Aristotle's theory of friendship to classify online friendships. *Ethics and Information Technology, 18*(2), 65–79.

Kauppinen, A. (2014). Flourishing and finitude. *Journal of Ethics and Social Philosophy, 2*, 1–6.

Kipke, R. & Rüther, M. (2019). Meaning and morality. *Social Theory and Practice 45*(2), 225–247.

Klincewicz, M. (2019). Artificial intelligence as a means to moral enhancement. *Studies in Logic, Grammar and Rhetoric, 48*(61), 171–187.
Knell, S., & Rüther, M. (forthcoming). The end of work, superefficiancy and meaning in life. A Humanistic Perspective. *AI & Ethics*. forthcoming.
Landau, I. (2022). *The Oxford Handbook of meaning in life*. Oxford University Press.
Levy, N. (2005). Downshifting and Meaning in Life. *Ratio, 18*(2), 176–189.
Loh, J. (2019). *Roboterehik: Eine Einführung*. Suhrkamp.
Maslen, H., Savulescu, J., & Hunt, C. (2020). Praiseworthiness and Motivational Enhancement: 'No Pain, No Praise'? *Australasian Journal of Philosophy, 98*(2), 304–318.
McCarthy, J., et al. (2006). A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955". *AI Magazine, 27*(4), 12.
Metz, T. (2020). African theories of meaning in life: A critical assessment. *Southern African Journal of Philosophy, 39*(2), 113–126.
Metz, T. (2011). The Good, the True and the Beautiful: Towards a Unified Account of Great Meaning in Life. In: *Religious Studies* 47(4): 389–409.
Metz, T. (2013). *Meaning in life. An analytic study*. Oxford University Press.
Metz, T. (2022). The meaning of life, *The Stanford Encyclopedia of Philosophy*. In: E. N. Zalta & U. Nodelman (Eds.), forthcoming URL = <https://plato.stanford.edu/archives/win2022/entries/life-meaning/>
Metzinger, T. (2013). Two principles for robot ethics In: E. Hilgendorf & J. P. Günther (Eds.), *Robotik und Gesetzgebung* (pp. 263–302). Nomos.
Misselhorn, C. (2021). *Künstliche Intelligenz und Empathie*. Reclam.
Monsó, S., Benz-Schwarzburg, J., & Bremhorst, A. (2018). Animal morality: What it means and why it matters. *The Journal of Ethics, 22*(3–4), 283–310.
Rüther, M. & Muders, S. (2016). Wohlergehen und Sinn. Zwei Aspekte des guten Lebens. In *Jahrbuch für Wissenschaft und Ethik* (20), 73–112.
Rüther, M. (2021a). Meaning in life oder: Die Debatte um das sinnvolle Leben—Überblick über ein neues Forschungsthema in der analytischen Ethik. Teil 1: Grundlagen. In *Zeitschrift für philosophische Forschung, 75*(1), 115–155.
Rüther, M. (2021b). Meaning in life oder: Die Debatte um das sinnvolle Leben—Überblick über ein neues Forschungsthema in der analytischen Ethik. Teil 2: Normativ-inhaltliche Fragen. In *Zeitschrift für philosophische Forschung, 75*(2), 316–354.
Rüther, M. (2023). *Sinn im Leben. Eine ethische Theorie*. Suhrkamp.
Müller, V. (2020). Ethics of artificial intelligence and robotics. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/ethics-ai/
Newitz, A. (2017). Fight AI by becoming AI: Elon Musk is setting up a company that will link brains and computers, *Ars Technica*: https://arstechnica.com/information-technology/2017/03/elon-musk-is-setting-up-a-company-that-will-link-brains-and-computers/. Accessed 23 Sept 2022.
Nida-Rümelin, J., & Weidenfeld, N. (2022). *Digital humanism*. Springer.
O'Neill, E., Klincewicz, M., & Kemmer, M. (2022). Ethical issues with artificial ethics assistants. In C. Veliz, (Ed.), *Oxford Handbook of Digital Ethics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198857815.013.17.
Precht, R. D. (2020). *Künstliche Intelligenz und der Sinn des Lebens: Ein Essay*. Goldmann.
Purves, D., & Delon, N. (2018). Meaning in the lives of humans and other animals. *Philosophical Studies, 175*(2), 317–338.
Russell, S. & Norvig, P. (1995/2020). *Artificial intelligence: A modern approach*. Prentice Hall.
Sandel, M. (2007). *The case against perfection. Ethics in the Age of Genetic Engineering*. Harvad University Press.
Savulescu, J. & Maslen, H. (2015). Moral enhancement and artificial intelligence: Moral AI? In Romportl, J. et al. (Eds.), *Artificial intelligence: Topics in Intelligent Engineering and Informatics*, (Vol. 9, pp. 79–95). Springer.
Scheffler, S. (2018). *Why worry about future generations?* Oxford University Press.
Schnell, T. (2021). *The psychology of meaning in life*. Routledge.
Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy, 39*(1), 98–119.
Scripter, L. (2022). Remorse and the Ledger Theory of meaning, *Philosophy,* 1-22. online first: https://doi.org/10.1017/S0031819122000304:1-22

Seachris, J. (2013). General introduction. In J. Seachris (Ed.), *Exploring the Meaning of Life: An Anthology and Guide* (pp. 1–21). Wiley Blackwell.
Singer, I. (1996). Meaning of life, Volume 1: The Creation of Value. MIT Press.
Smids, J., Nyholm, S., & Berkers, H. (2020). Robots in the workplace: A threat to—or opportunity for—meaningful work? *Philosophy & Technology, 33*(3), 503–522.
Smith, J. K. (2021). *Robotic persons*. WestBow.
Smuts, A. (2013). The good cause account of the meaning of life. In: *Southern Journal of Philosophy, 51*(4), 536–562.
Nyholm, S., & Frank, L. E. (2017). From sex robots to love robots: is mutual love with a Robot possible?. In: J. Danaher & N. McArthur (Eds.), *Robot Sex: Social and Ethical Implications* (pp. 219–244). MIT Press.
Nyholm, S., & Campbell, S. M. (2022). Meaning and anti-meaning in life. In: I. Landau (Ed.), *Oxford Handbook of Meaning in Life* (pp. 277–291). Oxford University Press.
Nyholm, S., & Smids, J. (2020). Can a robot be a good colleague? *Science and Engineering Ethics, 26*(4), 2169–2188.
Nyholm, S. (2021). Meaning and anti-meaning in life and what happens after We Die. *Royal Institute of Philosophy Supplement, 90*, 11–31.
Nyholm, S. (2022). Technological manipulation and threats to meaning in life. In F. Jongepier & M. Klenk (Eds.), *The Philosophy of Online Manipulation* (pp. 235–252). Routledge.
Nyholm, S. (2023). *This is technology ethics: An introduction*. Wiley-Blackwell.
Suits, B. (1978). *The Grasshopper: Games, life, and utopia*. Toronto University Press.
Sullins, J. P. (2012). Robots, love, and sex: The ethics of building a love machine. In *IEEE Transactions on Affective Computing, 3*(4):398–409. https://doi.org/10.1109/T-AFFC.2012.31
Tasioulas, J. (2022). Artificial intelligence, Humanistic Ethics. *Dædalus, 151*(2), 232–243.
Taylor, R. (1999). The meaning of life. In: *Philosophy Now, 24*:13–14.
Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred A. Knopf.
Thomas, L. (2005). Morality and meaningful life. In: *Philosophical Papers, 34*(3), 405–427.
Tigard, D. (2021). Workplace automation without achievement gaps: A reply to Danaher and Nyholm. *AI and Ethics, 1*(4), 611–617.
Tiku, N. (2022). The Google engineer who thinks its AI has come alive, *Washington Post*: https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/. Accessed 23 Sep 2022).
Turing, A. (1950). Computing machinery and intelligence, *Mind* LIX: 433–460.
Turkle, S. (2011). *Alone together*. Basic Books.
Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy and Technology, 28*(1), 107–124.
Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
van Wynsberghe, A. (2021). Sustainable AI: AI *for* sustainability and the sustainability *of* AI. *AI Ethics, 1*(4), 213–218.
Vanzura, M. (2021). What is it like to be a done operator? Or, remotely extended minds in war. In R. W. Clowes, K. Gärtner, & I. Hipólito (Eds.), *The Mind-Technology Problem* (pp. 211–229). Springer.
Vold, K. (2015). The parity argument for extended consciousness. *Journal of Consciousness Studies, 22*(18), 16–33.
Wareham, C. S. (2020). Artificial Intelligence and African conceptions of personhood. *Ethics and Information Technology, 23*(2), 127–136.
Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. *Ethics and Information Technology, 23*, 601–610.
Wielenberg, E. (2005). *Value and virtue in a godless universe*.
Wolf, S. (2007). The meanings of lives. In J. Perry (Ed.), *Introduction to philosophy* (pp. 62–73). Oxford University Press.
Wolf, S. (2010). *Meaning in life and why it matters*. Princeton University Press.