Philosophy & Technology (2024) 37:37
https://doi.org/10.1007/s13347-024-00717-z

RESEARCH ARTICLE

Why Care About Sustainable AI? Some Thoughts From The Debate on Meaning in Life

Markus Rüther[1]

Received: 15 July 2023 / Accepted: 9 February 2024 / Published online: 2 March 2024

Abstract
The focus of AI ethics has recently shifted towards the question of whether and how the use of AI technologies can promote sustainability. This new research question involves discerning the sustainability of AI itself and evaluating AI as a tool to achieve sustainable objectives. This article aims to examine the justifications that one might employ to advocate for promoting sustainable AI. Specifically, it concentrates on a dimension of often disregarded reasons — reasons of "meaning" or "meaningfulness" — as discussed more recently in the "meaning in life" literature of analytic ethics. To proceed, the article first elucidates the working definitions of "sustainable AI" and "meaning in life", while also setting the criteria for evaluating the plausibility of these reasons. Subsequently, it presents and scrutinises three arguments for the claim that one has reasons to care about sustainable AI from a perspective of meaning: the Meaning-conferring-action Argument, the Afterlife Argument, and the Harm Argument. In conclusion, this article asserts that only the Harm Argument presents a viable line of reasoning. However, it also outlines the presuppositions of this argument and the additional steps necessary to make it compelling.

Keywords Sustainability · Sustainable AI · Meaning in Life · Meaningfulness · Harm Argument

1 Introduction

The profound impact of artificial intelligence (AI) on our individual lives and societal structures is undeniable. Its transformative power extends from personal work environments and everyday living to major public sectors such as healthcare, education, transportation, government, and politics. As AI-based applications become increasingly prevalent, they naturally spark significant ethical debates. A range of

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

[1] Research Center Jülich GmbH, Jülich, Germany

critical topics has been the focus of these discussions, including ensuring AI transparency and explainability, preventing data biases and discrimination, safeguarding privacy, contemplating the implications of AI-induced decision-making, and ensuring the equitable distribution of AI's benefits across society.

However, more recently, a group of ethicists has also linked AI ethics to another subject that some refer to as a new wave in AI ethics. Specifically, this pertains to the field's self-reflection on its own sustainability and impact on other fellow humans and future generations.[1] This conversation is sometimes referred to as the debate on "sustainable AI". Numerous reasons have been proposed for this expansion. Primarily, these reasons revolve around individual well-being or moral obligations. It is argued that the care for sustainability is something that is good for one's own life or that we owe to others, for instance our children, future generations, the environment, or the planet as a whole. These reasons, of course, are comprehensive in order to justify our commitment to sustainability. But do they represent the only path to justification?

This article operates under the premise that there are additional relevant considerations when justifying our commitment to sustainability. To characterise these reasons, it is beneficial to reference a discussion in normative ethics, specifically the so-called "meaning in life" debate that has emerged over the past 15 to 20 years in analytic ethics. Within this discourse, one fundamental proposition is that reasons of meaning are an independent element in our ethical considerations, distinct from reasons of well-being and morality. Hence, following this debate, the working hypothesis is that these reasons of meaning could prove valuable if incorporated into the AI ethics discourse, particularly within the context of sustainable AI.

To substantiate this claim, this article will advance in the following steps: Initially, it will clarify several critical prerequisites, particularly the central concepts of "sustainable AI" and "meaning in life". These terms are not inherently self-explanatory. "Sustainable AI" is a complex term, as is "meaningfulness". Subsequently, building upon this foundational understanding, the article will probe the potential intersections of these two concepts. Specifically, it will investigate three primary reasons why the notion of meaning might instigate concern about sustainable AI. Throughout this exploration, the article will aim to present a balanced analysis of the persuasiveness of each reason, acknowledging that not all reasons carry equal weight. Some arguments may prove more compelling than others, and the goal is to identify the most striking among them. Finally, in the concluding stage, the findings will be synthesised, and potential directions for future research will be outlined.

2 Sustainability and Sustainable AI

The primary aim of this article is to explore the relationship between sustainable AI and meaningfulness. Understanding the basic terms under discussion is essential. This section will outline my understanding of sustainability, followed by an exploration of the concept of 'meaning in life' in the next section.

[1] Cf. Brevini (2020) and van Whynsberghe (2021), standing in for many others.

Notably, the concept of sustainability has emerged as one of the most salient concepts in various disciplines, such as philosophy, politics, and sociology. Historically, the concept and its key connotations are largely attributed to the influential 1987 Brundtland Report, which characterized sustainability as "fulfilling the needs of the present without compromising the ability of future generations to meet their own needs."[2] This definition prompts us to give equal consideration to the present and the future, thus casting sustainability as a "visionary and forward-looking paradigm."[3] Moreover, the usage of the term "sustainability" itself has evolved over time. Initially used in the Brundtland Report primarily to describe an aspect of intergenerational justice, others found this definition too narrow, emphasizing additional aspects of the concept.[4] Zwarthoed (2017), for instance, pointed out that we also have to integrate the idea that sustainability is about global justice. In this regard, the term is not solely directed towards future generations but also towards living people around the globe and our moral duties towards them. Seen in this light, one might interpret the UN's Sustainable Development Goals not just as recommendations to safeguard future generations, but also as advice to care about people being harmed through unsustainable endeavors in the present.[5]

Within the context of this broader understanding, defining sustainability more systematically can take several forms. Mensah (2019), for example, mentions that many favor a three-pillar approach integrating environmental, social, and economic considerations into our decision-making.[6] Meanwhile, others propose adding more elements to the mix, such as culture.[7] However, the subsequent problem of such a move seems obvious and has to do with the hierarchy of its various pillars. As Heilinger et al. (2023) point out, considering them as equally important gives rise to numerous problems when it comes to actual assessments or decisions because trade-offs will often be necessary. For instance, is the preservation of natural resources

[2] Brundtland Report (1987).

[3] Mensah (2019), 9.

[4] I would like to thank an anonymous reviewer for pointing me towards this more encompassing understanding of sustainability.

[5] But why be concerned only about humans? In this regard, environmental thinkers have often criticized the Brundtland Report for being too anthropocentric, presupposing humans as the sole subjects affected by unsustainable outcomes. Plumwood (1993), for instance, points to the language of the report, which uses terms such as "asset," "capital," and "resources" in connection with natural objects and systems. This terminology can lead to their being regarded merely as instruments for human usage, not recognizing their intrinsic value. Others have noted that the report actually endorses a broader moral perspective on the status of, and our relationship to, nature and non-human species. Brennan and Norva (2022) argue that this is evidenced by statements such as "the case for the conservation of nature should not rest only on development goals. It is part of our moral obligation to other living beings and future generations" (Brundtland Report 1987, Chapter 2, Paragraph 55), which they interpret as presenting a non-anthropocentric conception of sustainability, opening it up to include other sentient beings. This line of reasoning gains more attention in the field recently and is supported by several authors (see for instance also Visseren-Hamakers (2020), Sebo at el. (2022), and Bossert & Hagendorf (2023)). Although, I assume that there is something about it, I do not want to take stance in this paper. Here, I choose to remain neutral on this issue and adhere to the conventional reading and interpretation of sustainability as primarily concerned with humans.

[6] Cf. UCLG (2010).

[7] Cf. UCLG (2010).

or environmental systems more important than, or can it be discounted against, achievements regarding social justice, such as a reduction in poverty? At this point, it seems that normative presuppositions about the normative valence of the pillars enter directly into the conceptual clarification of the term. We must acknowledge that sustainability is a theory-laden term that cannot be defined without relying on ethical considerations in the first place. This does not mean that we cannot use the three-pillar approach to paint a more nuanced picture of sustainability. However, it does mean that we should keep its value-ladenness in mind when using the concept and not gloss over it easily.

Given the wide application of the concept, it is no surprise that sustainability is a key concern when discussing new technologies, such as AI. For instance, the European Commission's High-Level Expert Group on Artificial Intelligence claims that "AI systems can help to facilitate the achievement of the UN's Sustainable Development Goals, such as promoting gender balance and tackling climate change, rationalizing our use of natural resources, enhancing our health, mobility and production processes […]".[8] Yet, this perspective is not exhaustive. Van Whynsberghe (2021), for instance, invites us to contemplate the sustainability of AI itself, asking whether AI technologies are resource-efficient and do not produce unnecessary toxic waste.[9] This viewpoint shifts the focus from leveraging AI for sustainable ends to scrutinising AI technologies through the lens of sustainable criteria. Echoing this sentiment, the European Commission's European Group on Ethics in Science and New Technologies (EGE) contends that AI technologies should align with our "responsibility to ensure the basic preconditions for life on our planet, continued prosperity for mankind and preservation of a good environment for future generations".[10]

Considering the high regard for sustainability, it can be entirely expected that many reasons have been given in order to justify the promotion of sustainability, particularly in the sector of AI. Interestingly, though, most of the reasons fall in either of two main ethical dimensions: well-being and moral duties. Sustainability, it is argued, is something that we should promote either out of reasons that refer to aspects that are good for us or of reasons that refer to our moral obligations towards fellow human beings, such as those in other parts of the world, and those living in the future. For instance, some argue that not caring or act against the interest of these humans would interfere with our own happiness since it contradicts deep preferences such as the "love for humanity".[11] Others claim that such a care is part of having a virtuous character[12] or that one might have independent eudaimonistic reasons for it.[13] Those reasons can all be interpreted as referring to aspects of one's

[8] High-Level Expert Group (2019).

[9] An application of this thought, attempting to concretely measure the carbon footprint, can be found in Strubell (2019) and Henderson et al. (2020).

[10] EGE (2018).

[11] This is, for instance, one proposition that Samuel Scheffler made in his two latest books. Cf. Scheffler (2011, 2018).

[12] Cf. Jamieson (2007).

[13] Cf. Gardiner (2011), Thompson (2012), Williston (2015).

own well-being. In contrast, others insist that we do have individual obligations to care about currently living and future humans because it is an imperfect duty[14] or because not doing so or acting against it harms others,[15] increases the probability of harms,[16] or shows a failure in exercising a well-grounded duty of due care.[17] Those reasons can be understood as reasons that belong to the realm of moral duties and morality.

Undeniably, considerations of well-being and morality are sound justifications for the pursuit of sustainability. This remains true whether we view AI as an instrumental tool or as an entity subject to our evaluations. However, the compass guiding us towards sustainable AI may encompass more than just the familiar territories of the two standard dimensions. Particularly in the recent past, it has been argued that there are compelling reasons beyond these conventional confines. This leads us to an exploration of a dimension sometimes called "meaning" or "meaningfulness" that has gained some prominence in analytical ethics, but it is largely uncharted territory when it comes to ethical explorations on sustainability. In this article, it is assumed that this is a desideratum, and that we might have good reasons to link both discourses together. However, before delving into this connection any further, we need to develop a more refined understanding of the concept of "meaning". This endeavour will be the focus of the forthcoming section.

3 A Further Dimension: Meaningfulness

In this article, I propose that exploring the issues of sustainable AI from the ethical perspective of "meaningfulness" would be a promising approach. But what exactly does "meaningfulness" imply, and what are the implications of adopting such a perspective? In fact, it can be stated that this and analogous questions have garnered considerable attention over the past two decades in analytic ethics.[18] Commonly referred to as the discourse on "meaning in life", this topic has been subject to extensive exploration by scholars such as Susan Wolf (2010) and Thaddeus Metz (2013), who have produced highly influential works in this area, significantly enhancing the philosophical comprehension of this issue. Furthermore, in the recent past, the notion of meaning has been increasingly applied to various areas of ethics, including medical contexts involving questions of life and death,[19] animal ethics and the status debate,[20] climate ethics and the ethical responsibility to future generations,[21] or ethical considerations around the new emerging technologies of AI.[22]

[14] Cf. Baatz (2014).
[15] Cf. Nolt (2011).
[16] Cf. Hiller (2011).
[17] Cf. Vance (2016).
[18] For a newer overview of the field see Metz (2022).
[19] See for an overview Metz (2022).
[20] See Purves and Delon (2018), Monsó et al. (2018).
[21] See Campbell and Nyholm (2015), Di Paola and Nyholm (2023).
[22] Cf. Nyholm and Rüther (2023).

Given the status of the debate, one may assume to find many different perspectives, approaches, and views on the issue of meaningfulness. This is, of course, true in certain respects. However, when we consider the basic assumptions of the field, there are some that are widely shared. In the following, three of those assumptions that are considered in this article and that form the cornerstone of the subsequent sections are worth mentioning.

First, on a semantic level, the field essentially understands "meaning" as something good for its own sake that can be exemplified by a human's life or some aspect of their life to a variable degree. For many authors, that personal meaning is opposed to a purpose that has been or could be conferred on humanity by something external to it, such as God, as conceived 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, and the meaning "of" life, a cosmic end that might be ascribed to the human race or the physical universe as a whole.[23]

Second, on a normative level, the essential claim of the field is that "meaning" is neither identical nor fully subsumable under the standard axiological parameters. Many, for instance, contrast meaning with narrow self-interest or morality.[24] This means that, quintessentially, for a person to acquire meaning in their life, they must focus not solely on themselves, or at least not on their subjective well-being or what they morally owe to others, but instead orient their life *in some respects*[25] "outwardly" to some "higher values".

Third, there also seems to be a tendency to agree on typical activities that can confer meaning, such as rearing children, being in a romantic relationship, volunteering for a charity, demonstrating a refined skill to others, advancing knowledge through science, or creating works of art.[26] Given such examples, there have been attempts to subsume sources of meaning under the label of the good, the true, and the beautiful.[27] This does not imply an orientation towards platonic universals. The classic triad solely marks the realms in which meaning is supposed to take place and gives the different sources of meaning a more systematic shape, namely that they can be subordinated in the areas of altruistic (the good), scientific (the true), and artistic excellence (the beautiful). Systematically, those realms can and have been explored in different directions. Metaethically, for instance, one might ask about the semantic extension and meaning of "meaning" or its ontological status.[28] Normatively speaking, one might ask about the salient feature that makes "the good, the true, and the beautiful" a suitable candidate to confer meaning to one's life and how this feature might justify the difference between meaning, on the one hand, and well-being and morality, on the other.[29]

[23] See, for example, Seachris (2013), 3–4.

[24] This is the paradigmatic line of reasoning in Wolf (2010) and Metz (2013).

[25] The qualification "in some respects" is essential because most proponents hold that well-being or morality has at least some bearing on meaning while presuming that meaning and well-being or morality are not the same. See for an overview of the field Metz (2022).

[26] See the introduction in Landau (2022a).

[27] For further elaboration, see Metz (2011, 2013).

[28] For the different meta-ethical questions that are addressed in the debate, see Metz (2022).

[29] For paradigmatic works that address almost all normative questions, see Metz (2013) and the contributions in Landau (2022a).

Much more can certainly be said about this discourse and its research questions. However, for the purposes of this article, it is more critical to emphasise the key points that will be built upon in the subsequent sections. In line with the majority of discourse participants, it is assumed that "meaningfulness" signifies a value in a person's life that is at least partly autonomous and non-instrumental. This makes life more desirable, varies in degrees, and involves an orientation that extends beyond oneself. Furthermore, it is assumed that this orientation can be further explored, particularly within the realms of the good, the true, and the beautiful.

4 Meaningfulness in the Realm of Sustainable AI

4.1 Preliminaries

Having previously examined and defined the concepts of "sustainable AI" and "meaning in life" in the preceding sections, I now turn the attention towards bridging these two concepts. In the following, the primary query is thus: Can we identify reasons of meaning that substantiate our concern for sustainable AI? Note that this question can be understood in different ways. In order to understand the approach of this article, three things are worth mentioning.

Primarily, this article interprets the posed question in the light of pro tanto reasons, not all things considered reasons. All things considered reasons are derived from a deliberation process and are decisive after balancing the pros and cons. In contrast, pro tanto reasons may not be decisive and hence may not qualify as all things considered reasons. Instead, they carry some normative weight, while being neutral about their specific normative force and whether they override all other reasons.

Secondly, this question will be explored within the framework of universal reasons. These reasons are understood as being applicable to all humans, as opposed to relative reasons, which apply to specific variables such as a particular person, family, society, or culture. However, pursuing universal reasons does not necessarily endorse a universalism about reasons of meaning; rather, it signifies a research perspective that focuses on reasons with normative significance for everyone.

Lastly, the investigation proceeds from the standpoint of an overarching consensus. The aim is to identify reasons that can find widespread agreement across various substantive theories of meaning. This does not necessitate the acceptance of specific commitments but rather suggests that, ideally, we can pinpoint reasons that do not rest on a particular substantive theory of meaningfulness and can, therefore, resonate with a range of different approaches.

4.2 The Care for Sustainable AI as a Meaning-Conferring Action

The aim of this article is to investigate the various ways in which someone might present universal pro tanto reasons for the promotion of sustainable AI, particularly from the perspective of meaning. One strategy to construct a compelling argument in this regard is to highlight the meaning-conferring nature of certain actions. We

might call the argument, based on this strategy, the Meaning-conferring-action Argument.

The rationale for this argument could be constructed as follows: We begin with the assumption that a significant way to attain meaningfulness in one's life is by showing care for others. This premise is reinforced by reflecting on the lives of individuals who are frequently cited in discourse as definitive examples of meaningful lives. Figures such as Nelson Mandela, Mother Teresa, and Rosa Parks, each of whom dedicated themselves to assisting others in distinct ways, serve as paradigms in the realm of the "good".[30] Given this assumption, we might continue with the argument by proposing that there is a striking parallel to the promotion of sustainable AI: just as these individuals advocated for the welfare of others, promoting others currently living or coming into existence in the future represents a similar act of caring. Therefore, we can deduce that just as Mandela, Mother Teresa, and Marie Curie infused their lives with meaning through their acts of caring, so too does someone who works to enhance the sustainability of AI technologies or uses these technologies to foster a more sustainable future.

At this point, several questions naturally present themselves. One might initially question whether the parallel between the paradigmatic cases of meaningful lives and promoting sustainability holds. Are there no significant differences between, say, the act of fighting for justice like Mandela and being a developer of a sustainable AI software? In the following, I want to leave those points aside and grant the claim that the care for meaningful AI can be a meaning-conferring endeavour as much as the caring of Mandela, Mother Teresa, and Marie Curie does. Instead, I want to focus on an additional assumption of the argument, namely that the importance, or the normative weight, of caring about sustainable AI is universally identical for all persons. In fact, the argument seems to suggest that there are equally pro tanto reasons for everyone to promote sustainability of AI and use AI for sustainability. However, this assumption can be doubted with good reasons. One reason for concerns lies in the role that relative factors play in the constitution of the normative weight of meaningfulness. Those factors can be interpreted differently, but two relevant factors that might be worth exploring are the role of circumstances and the shape of one's personality.

[30] This reflects on the methodological assumption on the meaning in literature, namely that a normative theory of meaningfulness can be discerned by reflecting on negative and positive examples. A classic negative case is the life of Sisyphus, who is condemned to eternal, meaningless toil: each day, he must roll a boulder up a hill only to have it roll back down when he approaches the top. In contrast, the positive examples are numerous and are mostly systemised under the classic triad "the true, the good, and the beautiful". The examples include individuals such as Marie Curie, Albert Einstein, and Charles Darwin in the realm of the true; Martin Luther King Jr., Mother Teresa, and Rosa Parks in the realm of the good; and Simone de Beauvoir, Vincent van Gogh, and Michelangelo in the realm of the beautiful. It is worth noting that the demanding lifestyles of these individuals are not viewed as the only or even the best way to attain meaning. Instead, they serve as clear-cut instances of meaningfulness that can be used to test our intuitions about what makes life meaningful. While there are undoubtedly meaningless lives, such as those of Sisyphus, meaningful lives exist, such as those of King, Curie, and de Beauvoir. The question then becomes: What features of these lives make them such paradigmatic cases?

To begin with the circumstances, one might argue that some situations call for other aspects of meaning that deserve a higher priority than sustainable AI. Think about a healthcare worker during a pandemic, a scientist with the opportunity to obtain a huge grant, or an artist with a once-in-a-lifetime gallery offer. All three seem to have good reasons to prioritise their other meaningful endeavours over the care for sustainable AI. This might change in the future when the pandemic ends or the opportunities for grants and gallery offers vanish. In these situations, the normative weight of the reasons might be different, and so the priorities are supposed to change again.

The same might be true for individual factors, namely those that determine the personality and its shape. Setting aside the issue of how the shaping process can be described, it might be reasonable to assume that individuals are different and that this factor also influences the normative weight of reasons. Some individuals seem to match up better with the profile of a healthcare worker; some are a better fit for scientific endeavours; some are better equipped to work as an artist. In all these cases, the normative weight of reasons is not independent of who they are, and this factor also may lead to prioritising those projects over the care for sustainable AI. Of course, this may change. If, for instance, the person changes, the weight of reasons and the priorities may change as well.

Now, we might consider taking this argument even further. The previous thoughts are based on the rationale that the normative weight of caring for sustainable AI varies since it depends on relative factors such as circumstances and the shape of the personality. However, there might also be a rationale that those relative factors can show that sometimes there might be no compelling reason to care about sustainable AI at all.

A compelling example might involve individuals who lack access to AI technologies and therefore have no opportunity to promote sustainable AI. This could apply to people living in resource-constrained environments, such as those in economically disadvantaged circumstances or in geographically isolated locations. For these individuals, immediate daily tasks may revolve around essential needs, such as food, shelter, and basic healthcare. Their energy and resources are devoted to securing their own survival and that of their families, leaving little room for consideration or promotion of sustainable AI. In such circumstances, it seems reasonable to assert that people do not have even pro tanto reasons to concern themselves with sustainable AI.

Also, one might elaborate on this point by considering individuals whose personalities are shaped in such a way that they dedicate their lives to a single, meaning-conferring project. Consider the healthcare worker, the scientist, or the artist who devotes her body and soul to a lifelong task, whether it is to help the needy, achieve scientific milestones, or create artistic masterpieces. All three have good reasons to do what they do, but given who they are, they have no reason to concern themselves with sustainable AI. Here, a counterargument might advocate a "ground-floor approach" asserting that a minimal level of concern for sustainable AI is integral to a meaningful life. However, this approach is not reasonable.[31] It would imply that

[31] Yet, the idea of assuming a "ground floor" could be plausible in other contexts. For instance, Landau (2022b) provides a compelling argument that life can sometimes be overloaded with meaningfulness to the point where it conflicts with crucial aspects of well-being and morality. If this assertion holds, it provides a justification for the idea that a minimum standard, a "ground floor", should exist in these dimensions and should not be violated or diminished.

lives centred on other pursuits lack meaning. Consider, for instance, that the above-mentioned individuals achieved significant accomplishments during their lives, including earning a Nobel prize or other merits for their endeavours, but with little concern for sustainable AI. Equating the meaningfulness of these lives to that of tragic figures such as Sisyphus seems highly counterintuitive. These high-performers might, at most, experience a slight decrease in the meaningfulness of life, but arguing that their life is devoid of meaning is an overreach.

Where does this leave us? The initial claim of the Meaning-conferring-action Argument was that everyone might have pro tanto reasons to care about sustainable AI because it is a meaning-conferring endeavour. However, the previous thoughts imply that the normative weight attributed to caring for sustainable AI is not fixed and the same under all circumstances and for all people. There are no equally compelling pro tanto reasons to care for sustainable AI in all situations for everyone. Beyond that, there might also be circumstances and individuals for whom and in which there are no reasons at all. If that is true, though, the main claim of the argument that everyone has a pro tanto reason cannot be correct.

4.3 The Relevance of the Afterlife for Meaningfulness

The Meaning-conferring-action Argument for caring about sustainable AI was constructed around the significance of caring, as it bestows meaning. The second argument, however, is slightly different, as it does not rely on the normative force of a specific action or type of action, but rather on a particular state of affairs in the future — specifically, the trajectory of the world following our demise, otherwise known as the afterlife. Given its object, in the following, we call it the Afterlife Argument.

The Afterlife Argument could be framed in this manner: The meaningfulness of our own lives partly hinges on the afterlife, for example, on the economic, social, and environmental circumstances of people how outlive us or – looking further ahead – on future generations. If these people are worse off or potentially faced with extinction, perhaps due to unsustainable AI, the meaning of our own lives may be diminished or even rendered meaningless.[32] Consequently, we have compelling reasons to care about sustainability, especially sustainable AI.

Given the structuring of the Afterlife Argument, it is evident that it draws inspiration from the more recent works of Samuel Scheffler (2011, 2018). Scheffler uses the term "the afterlife" to denote what transpires for people who continue living after our own deaths. He also employs several thought experiments to propel what he labels the "afterlife conjecture", pertaining to the significance of the afterlife for our own lives. For instance, he encourages contemplation of scenarios where an asteroid might obliterate the earth's population or where the world's population might become infertile in the not-so-distant future, thus leading to our soon-after demise.

[32] It is important to note that additional claims are necessary to substantiate this argument. Most notably, a significant claim is required regarding the normative relevance of people who outlive us. Why is there life relevant for our life to be meaningful? I will revisit this point later, also discussing how this requirement poses challenges in formulating universal reasons for caring about sustainable AI.

As Scheffler posits, such scenarios would instil terror and despair in us. Intuitively, we cannot remain indifferent about what happens to others after our own deaths; instead, it appears we have reasons to care about those people, even if we might already be gone by the time they come into existence.

Scheffler does not employ the language of meaningfulness when discussing the afterlife. He primarily speaks about what holds value for us, what we care about, or what matters to us. His ideas seem more linked to the dimension of individual well-being than to what most participants in the meaning-in-life literature understand as meaningfulness. Nonetheless, I believe it is worthwhile to at least acknowledge the fundamental rationale that the afterlife matters, perhaps not solely for well-being, but also for leading a meaningful life.[33] For if we do, we can examine the previously given argument a bit further to determine if it holds true. Does the afterlife matter for meaningfulness after all?

Perhaps the most promising way to establish the connection is by arguing that the afterlife, or aspects thereof, constitutes an essential part of meaningfulness itself, or at the very least, it is integral to something else that imparts meaningfulness. For instance, some argue that meaning in life is partly constituted by creating lasting impacts on the world[34] or by reaching certain achievements that endure over time.[35] Others contend that the possibility of future generations carrying on one's project confers meaning,[36] or that meaning arises when one's life fits into a narrative story, which includes the afterlife.[37] These theories of meaningfulness each warrant their own debate. Here, I want to focus solely on a common trait. This is that all those theories assume the afterlife matters for meaningfulness because its status is relevant for, as one might put it metaphorically, fitting into a larger picture — whether through one's impact, achievements, generational projects, or a created narrative story. If there is no afterlife, or if future generations are worse off, it may become questionable whether these criteria for meaningfulness can be realised. How can we perceive ourselves as part of a larger schema if there is no schema at all or if it is degraded to a point that makes it difficult to positively relate to it? Given this rationale, we have compelling reasons to care about the afterlife, as it ensures our ability to maintain this connection, which in itself is a constitutive part of meaningfulness. Here, at least two distinct concerns may arise.

Firstly, one might challenge the diverse interpretations of what constitutes meaningfulness. For instance, questions could be raised about whether making lasting impacts, achieving enduring feats, fostering projects that span generations, or crafting a shared narrative are indeed the most crucial aspects — whether these are necessary or even sufficient for a life to be deemed meaningful. Some might argue that these proposed features are contributive and marginal at best, which, in turn,

[33] See also Nyholm (2021) for another attempt to make the "afterlife conjecture" applicable, though not specifically to sustainable AI, but to general concerns about the future.

[34] Cf. Smuts (2013) and Bramble (2015).

[35] Cf. Bradford (2016).

[36] Cf. Scheffler (2011, 2018).

[37] Cf. Kauppinen (2011) and Nyholm (2021).

complicates understanding why these should constitute a significant rationale for caring about future generations.[38]

Secondly, even if one deems a specific theory of meaningfulness as credible, questions could still be raised — as already hinted in the arguments against the Meaning-conferring-action Argument — regarding whether such a theory can genuinely provide a universally applicable reason. These doubts may stem from the potential belief that reasons for assigning meaning are relative, hinging on factors such as personal circumstances and personality shape. If this claim holds true, though, there may not be any universal reasons that would compel everyone, in all situations, to care about our present human fellows and future generations.

Both concerns are reasonable. However, I wish to set them aside, primarily because they presuppose more in-depth exploration of the different theories of meaningfulness that must be undertaken elsewhere. Instead, I want to emphasise another issue tied to the general research perspective of this article. As previously mentioned, this article seeks universal pro tanto reasons that can be accepted by numerous contributors in the meaning-of-life literature. An effective way to ensure this is to avoid making substantial claims about meaningfulness and instead rely on rationales that most contributors can accept. Examining the afterlife conjecture in this light, it is clear that it posits significant claims about meaningfulness. Furthermore, as demonstrated earlier, the argument only holds if one adheres to a specific theory of meaningfulness. Again, I do not aim to refute these theories. I merely want to emphasise the fact that the afterlife conjuncture only makes sense if one already maintains a substantial claim about what gives life meaning. Therefore, if we persist with the task of discerning reasons that many theories can accept, we might be better off if we explore other rationales.

4.4 Creating Harm and Anti-Meaningful Actions (or Omissions)

Thus far, this article has investigated two arguments positing that caring for sustainable AI might be rational from the perspective of leading a meaningful life. I have contended that both arguments fall short in providing persuasive reasons to care about sustainable AI. Nonetheless, I believe there's a beacon of hope and a pathway to construct a more compelling argument. This potential argument, which I refer to as the Harm Argument, has numerous foundations in the literature that can serve as a starting point to flesh it out more thoroughly. In the following, I will try to develop the bare bones of the argument.

The Harm Argument, in its most basic form, can be unfolded in four stages. First, it begins with a premise that many would find comprehensive: inflicting harm on others is a disvalue, and thus causing harm to other people, whether currently living or coming into existence in future, by endorsing unsustainable AI is equally a disvalue. Second, the argument broadens its perspective. It considers not only harmful actions taken through the use of unsustainable AI, but also the harm caused by

[38] For an overview of the different positions in the field see Metz (2022). My own view on this matter in fact contains some scepticism about the relevance of the mentioned aspects. Since my main point in this text does not pressupose my own take on what makes a life meaningful, I will not present and develop my position in detail. For further details see Rüther (2023).

the omissions when not using sustainable AI. Third, it proposes the applicability of this rationale to the concept of meaning. Why should the idea of harm be contained only within the framework of morality? One might argue that it should also factor into reflections on leading a meaningful life.[39] Fourth and finally, these elements are interconnected to reach the ultimate conclusion: causing harm to individuals through the promotion of unsustainable AI, or through omissions to support sustainable AI, diminishes life's meaningfulness. If this argument holds, it suggests that we have universal pro tanto reasons to care about sustainable AI. Acting against it through using unsustainable AI or not using sustainable AI inflicts harm, thereby negatively affecting the meaningfulness of one's life.

While navigating through the stages of the Harm Argument, numerous questions might arise. I will address some of these queries in the paragraphs below. However, at this juncture, it seems more promising to start with an appreciation of this argument's compelling nature, especially when compared to the two other arguments previously discussed.

In contrast to the Meaning-conferring-action Argument, the Harm Argument is not based on the premise that the act of supporting sustainable AI in itself imparts meaning. As shown, the challenge in supporting this claim stems from the assumption that it generates universally applicable pro tanto reasons since it appears more reasonable to suggest that such reasons are relative, influenced by specific circumstances and the individual involved. In contrast, the Harm Argument is better suited to offering universal pro tanto reasons. This is primarily because the concept of harm stands as a compelling foundation for universal claims. If there are universally applicable reasons at all, the avoidance of harm is undeniably an essential aspect.

In contrast to the Afterlife Argument, the Harm Argument does not necessitate a substantial theory of meaningfulness that might not find widespread agreement. The Harm Argument does not require extensive foundational work. While it does carry some presuppositions, as discussed later, the notion that inflicting harm on others has a significant impact on life's meaning is widely acknowledged within the literature concerning meaningfulness.

Therefore, if we are committed to finding widely shared, universal pro tanto reasons, the Harm Argument may provide the most feasible path for further investigation. However, it is critical to note that the argument is not self-evident in all its aspects, but also encompasses assumptions that call for further explanation. In what follows, I will pinpoint a few of these assumptions and endeavour to clarify how they might be understood, justified, or further developed in the future.

One of the key notions used in the argument is the concept of "harm", yet its definition and measurement are not clear. What do we understand by harm in the sphere of sustainable AI? How can we evaluate or quantify this harm? As Tadros (2014) notes, an effective argument requires a clear "currency" and "measure" of

[39] This point is pivotal yet not immediately obvious in the argument. To substantiate it, one must argue for the thesis that immoral actions can diminish meaningfulness. I will delve deeper into this claim later, as I lay the groundwork for future research on the harm argument. I extend my gratitude to an anonymous reviewer for emphasizing the importance of this aspect.

harm. We need to grasp not only what harm encompasses, but also its relevance in normative contexts. The literature on harm with regard to both issues is expansive, and while I do not intend to add to it, I would like to bring attention to two points that can help to refine the understanding of harm within the scope of sustainable AI or at least show the way for future work. For one, concerning the domains of harm, the earlier definition of sustainable AI can help to get one's feet on the ground. Here, sustainability was defined as addressing issues of intergenerational and global justice. The first aspect, clearly evident to many readers of the Brundtland Report, can now be understood as the prevention of harm to future generations. The second aspect pertains to the harm we currently inflict on individuals currently living. Furthermore, within these two domains, each can be further subdivided using a three-pillar approach that encompasses environmental, social, and economic factors. Upon closer examination, it may become apparent that additional pillars are worth considering. I have also noted in the terminology Section. 2 that this approach is not without its presuppositions. However, these three commonly acknowledged pillars provide a promising foundation for future research. Additionally, it may be valuable to distinguish between different types of harm. This could inform us about the potential severity of harmful actions towards other fellow humans, whether living today or in the future. To begin, it seems prudent to avoid extremes: On one hand, an unreflective utopianism that dismisses every act promoting unsustainable AI as harmless and without significant consequences for humans is ideologically misguided. On the other hand, labeling every such act as an instance of 'ultimate harm'—a term Persson and Savulescu (2012) use to describe threats to human existence—might be overly dramatic. However, as numerous researchers have emphasized, the impact of unsustainable AI technologies is profound, exceeding that of many other technologies.[40] Therefore, as many have already recognized, it is justifiable to allocate more resources to this area. Doing so will enable more nuanced assessments of harm in the future, likely involving case-by-case studies of specific unsustainable technologies.

Another foundational premise of the Harm Argument that warrants examination involves the idea that omissions can lead to harm, specifically by not promoting sustainable AI. This assumption might not be immediately intuitive, but it deserves careful exploration. A good starting point could be Quinn's (1989) observation that omissions are not actions, but inactions. Therefore, if someone fails to promote sustainable AI, he or she is not acting in a way that directly causes harm to other people currently living or coming into existence in the future; rather, he or she is failing to prevent that harm. But how can this be understood? One way might be to point to the status quo: if we carefully analyse the state of sustainable AI technologies and their contribution to sustainable goals, we must admit — as stated in the introductory section on sustainability — that they are in dire straits. Thus, if we neglect to promote sustainable AI, we indirectly allow harm to befall the people around us or future generations. Of course, further explanations are needed. The ethical literature on omission is tremendous. For instance, one might explore the normative relevance

[40] Cf. van Whynsberghe (2021) for further references to the existing literature.

of omissions in contrast to actions. Intuitively, there seems to be a difference, but in the ethical literature, this is a hotly debated topic and far from obvious.[41] Depending on the position one takes on this matter, this also might have implications for the relevance of omissions as part of the Harm Argument.

A further assumption explicitly made in the argument is that inflicting harm is a disvalue that can be incorporated into the realm of meaningfulness. Typically, most people would agree that the disvalue of causing harm plays a crucial role in moral matters. Thus, including it within the sphere of meaningfulness might be surprising and necessitates further explanation. Several interpretations are possible here, largely depending on one's understanding of the two dimensions of "morality" and "meaning" and how one interrelates them.[42] This article cannot encompass and evaluate all possible interpretations. Therefore, I will mention only a common strategy that, in my view, is also a reasonable one. This approach asserts that causing harm is morally reprehensible, and this moral failure *also* impacts meaningfulness. In order to show this connection, many contributors refer to extreme examples of "moral monsters". Landau (2011), for instance, cites the life of Hitler as emblematic of an exceptionally immoral life that could — even through the most creative thought experiments — hardly be construed as meaningful. This appears to indicate that immoral actions must impact the concept of meaning. A decrease in morality — and many contributors would follow Landau on this point — corresponds to a decrease in meaningfulness. Therefore, if one inflicts harm through the use of unsustainable AI, it may be argued that those actions are immoral and therefore result in a loss of meaning, rendering the immoral life as less meaningful.

In connection with the assumption that inflicting harm influences meaningfulness, another premise inferred from previous considerations is integrated into the argument. Namely, it suggests an expansive understanding of meaningfulness, which encompasses not only meaningful and meaningfulness actions and omissions, but also those that diminish the level of meaningfulness in one's life. There are also actions and omissions that are, as some call it, "anti-meaningful"[43] or "negatively meaningful".[44] There is still an ongoing debate on this issue, particularly about its implications. For instance, if we consider actions that are anti-meaningful or negatively meaningful, and if a life consists, on balance, of more negative than positive actions, it could be described as anti-meaningful rather than merely meaningless. As such, in comparison, a meaningless life, like that of Sisyphus, might not be the worst possible scenario in terms of meaningfulness. At this point, it could be beneficial to delve deeper into this implication, especially to understand if and how it impacts the normative "weight" of reasons for caring about sustainable AI.[45]

[41] Cf. Woollard and Howard-Snyder (2022).

[42] For an overview of the different options and their challenges Kipke & Rüther (2019).

[43] Cf. Campbell and Nyholm (2015), Nyholm and Campbell (2022).

[44] Cf. Metz (2013), Landau (2011), Smuts (2013).

[45] A helpful reference might be the insightful considerations in Di Paola and Nyholm (2023). Although the authors explored the role of anti-meaning in the context of climate change, many ideas might also be transferable to the topic of sustainable AI.

5 Summary and Final Remarks

In this article, I explored the question "Why should we care about sustainable AI?" from a unique perspective, namely the perspective of leading a meaningful life, as recently discussed in the meaning-in-life literature.

Initially, I laid down the assumed understanding of sustainable AI and meaningfulness. I differentiated between the sustainability of AI and the use of AI for sustainability. I also emphasised that the term, at best, has shared connotations on a very abstract level, but that more concrete definitions were often underdeveloped or even conceptually questionable. This necessitated a stipulative definition for the article. In terms of meaningfulness, I referred to the literature and elaborated on four assumptions, which served as starting points. I defined meaningfulness as signifying a value in a person's life that is at least partly autonomous and non-instrumental. It increases life's desirability, comes in degrees, and involves an orientation beyond oneself. Furthermore, I suggested that this orientation can be further explored, for instance, in the realms of the good, the true, and the beautiful. After that, I set the criteria by which I could evaluate the reasons for promoting sustainable AI. I clarified that my research perspective focuses on universal pro tanto reasons, which could be widely shared.

Then, I turned to three different arguments that could potentially justify the claim that we should care about sustainable AI. The first argument, which I labelled as the Meaning-conferring-action Argument, proposed that caring for sustainable AI is reasonable because caring is a meaningful action. I argued that while it was plausible to assume that caring is meaning-conferring, the argument could not deliver universal pro tanto reasons. This was due to relative factors such as circumstances and personality shapes, which suggested that caring for sustainable AI did not formulate a reason for everyone in every situation.

The second argument, the Afterlife Argument, was heavily inspired by Samuel Scheffler's "afterlife conjecture". This argument suggested that there were reasons to care about sustainable AI, as the afterlife, meaning the future lives of others after our death, mattered for leading a meaningful life in the present. I contended that while the argument was likely built on solid intuitions that were hard to ignore, it needed some theoretical investments to be true. Specifically, the argument presupposed substantial understandings of meaningfulness that only attracted theories sharing the same outlook.

The third argument, the Harm Argument, was predicated on the rationale that inflicting harm on other people would lead not only to a moral loss but also to a loss of meaning in one's life. I argued that this argument had some intuitive appeal. However, I also highlighted some presuppositions of the argument that needed further elaboration. The argument was underdeveloped in terms of the "currency" and "measure" of harm, the role and relevance of omissions in causing harm, the place of harm in a theory of meaning, and the assumption that negative meaning or anti-meaning existed. All things considered, however, I argued that the Harm Argument was the best approach so far if one was in search of universal pro tanto reasons for caring about sustainable AI. Therefore, if one intends to explore the "Why care?" question further, it would be promising for future work to start from here.

Acknowledgements I would like to thank my colleagues in Bonn und Juelich for helpful comments at earlier stages of the manuscript. For more feedback I also like to thank Jan-Christoph Heillinger who helped to make the manuscript more readable and pressed me to be detailed and nuanced in our presentation of the different arguments and positions.

Authors contributions Not applicable.

Funding Open Access funding enabled and organized by Projekt DEAL.

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

Declarations

Ethics Approval Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare that they have no competing interests.

References

Baatz, C. (2014). Climate change and individual duties to reduce GHG emissions. *Ethics, Policy and Environment, 17*(1), 1–19.

Bossert, L. N., & Hagendorff, T. (2023). The ethics of sustainable AI: Why animals (should) matter for a sustainable use of AI. *Sustainable Development, 31*, 3459–3467.

Bradford, G. (2016). Achievement, Wellbeing, and Value. *Philosophy Compass, 11*(12), 795–803.

Bramble, B. (2015). Consequentialism About Meaning in Life. *Utilitas, 27*(4), 445–459.

Brennan, Andrew and Norva Y. S. Lo, "Environmental Ethics", *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2022/entries/ethics-environmental/

Brevini, B. (2020). Black boxes, not green: Mythologizing artificial intelligence and omitting the environment. *Big Data & Society, 7*(2). https://doi.org/10.1177/2053951720935141

Campbell, S., & Nyholm, S. (2015). Anti-Meaning in Life and Why It Matters. *Journal of the American Philosophical Association, 1*(4), 694–711.

Di Paola, M., & Nyholm, S. (2023). Climate Change and Anti-Meaning. *Ethical Theory and Moral Practice, 26*(5), 1–16.

European Group on Ethics in Science and New Technologies (EGE). (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. European Commission.

Gardiner, S. (2011). *A Perfect Moral Storm: The Ethical Tragedy of Global Climate Change*. Oxford University Press.

Heilinger, J. C., Kempt, H., & Nagel, S. (2023). Beware of sustainable AI! uses and abuses of a worthy goal. AI ethics. https://doi.org/10.1007/s43681-023-00259-8

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. ArXiv:2002.05651.
High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. B-1049 Brussels.
Hiller, A. (2011). Climate Change and Individual Responsibility. *The Monist, 94*(3), 349–368.
Jamieson, D. (2007). When Utilitarians Should Be Virtue Theorists. *Utilitas, 19*(2), 160–183.
Kauppinen, A. (2011). Meaningfulness and Time. *Philosophy and Phenomenological Research, 84*(2), 345–377.
Kipke, R., & Ruther, M. (2019). Meaning and morality. Some considerations on a difficult relation. *Social Theory and Practice, 45*, 225–247.
Landau, I. (2011). Immorality and the Meaning of Life. *Journal of Value Inquiry, 45*, 309–317.
Landau, I. (Ed.). (2022a). *The Oxford Handbook of Meaning in Life*. Oxford University Press.
Landau, I. (2022b). Can there be overly meaningful lives? Conflicts between meaning in life and other values. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.946648
Mensah, J. (2019). Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature Review. *Cogent Social Sciences, 5*(1). https://doi.org/10.1080/23311886.2019.1653531
Metz, T. (2011). The Good, the True and the Beautiful: Toward a Unified Account of Great Meaning in Life. *Religious Studies, 47*(4), 389–409.
Metz, T. (2013). *Meaning in Life*. Oxford University Press.
Metz, T. (2022). The Meaning of Life. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). https://plato.stanford.edu/archives/win2022/entries/life-meaning/
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.
Nolt, J. (2011). How Harmful Are the Average American's Greenhouse Gas Emissions? *Ethics, Policy and Environment, 14*(1), 3–10.
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., & Campbell, S. (2022). Meaning and anti-meaning in life. In: I. Landau (Ed.), *In Handbook of Meaning in Life* (pp. 277–291). Oxford University Press.
Nyholm, S., & Rüther, M. (2023). Meaning in life in AI ethics—some trends and perspectives. *Philosophy & Technology, 36*, 1–24.
Persson, I., & Savulescu, J. (2012). *Unfit for the Future*. Oxford University Press.
Plumwood, V. (1993). *Feminism and the Mastery of Nature*. Routledge.
Purves, D., & Delon, N. (2018). Meaning in the Lives of Humans and Other Animals. *Philosophical Studies, 175*(2), 317–338.
Quinn, W. S. (1989). Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing. *Philosophical Review, 98*(3), 287–312.
Rüther, M. (2023). *Sinn im Leben. Eine ethische Theorie*. Suhrkamp.
Scheffler, S. (2011). *Death and the Afterlife*. Oxford University Press.
Scheffler, S. (2018). *Why Worry About Future Generations?* Oxford University Press.
Seachris, J. (2013). General introduction. In J. Seachris (Ed.), *Exploring the Meaning of Life: An Anthology and Guide* (pp. 1–21). Wiley Blackwell.
Sebo, J., Verkuijl, C., Hötzel, M., Achakulwisut, P., Lima, M. B., & Green, J. (2022). *Sustainable Development matters for animals too*. Governments have a responsibility to recognize that: *CABI One Health*. https://doi.org/10.1079/cabionehealth.2022.0002'
Smuts, A. (2013). The Good Cause Account of the Meaning of Life. *South Journal of Philosophy, 51*, 536–562.
Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *ArXiv:1906.02243*. https://doi.org/10.48550/arXiv.1906.02243
Tadros, V. (2014). What might have been. In J. Oberdiek (Ed.), *Philosophical Foundations of the Law of Torts* (pp. 171–192). Oxford University Press.
Thompson, A. (2012). The virtue of responsibility for the global climate. In A. Thompson & J. Bendik-Keymer (Eds.), *In Ethical Adaptation to Climate Change: Human Virtues of the Future* (pp. 1–24). MIT Press.
UCLG. (2010). *United Cities and Local Governments. Culture. Fourth Pillar of Sustainable Development*. UCLG.

United Nations World Commission on Environment and Development (1987) Our Common Future (Brundtland report). UN documents.
Vance, C. (2016). Climate Change, Individual Emissions, and Foreseeing Harm. *Journal of Moral Philosophy, 14*(5), 562–584.
Visseren-Hamakers, I. (2020). The 18th Sustainable Development Goal. *Earth System Governance, 3*, 100047. https://doi.org/10.1016/j.esg.2020.100047
van Whynsberghe, A. (2021). *Sustainable AI: AI for sustainability and the sustainability of AI.* AI and ethics, 213–218. https://doi.org/10.1007/s43681-021-00043-6
Williston, B. (2015). *The Anthropocene Project: Virtue in the Age of Clmate Change*. Oxford University Press.
Wolf, S. (2010). *Meaning in Life and Why it Matters*. Princeton University Press.
Woollard, F., & Howard-Snyder, F. (2022). Doing vs. allowing harm. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). https://plato.stanford.edu/archives/win2022/entries/doing-allowing/
Zwarthoed, D. (2017). Why Sustainability Principles Should Integrate Global Justice Concerns. *Ethics, Policy & Environment, 20*(3), 251–254.