Why emotions and works of art are pertinent to the assessment of the ethical risks of AI

Danielsen, Maria; Roeser, Sabine

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ORIGINAL PAPER

Why emotions and works of art are pertinent to the assessment of the ethical risks of AI

Maria Danielsen[1] · Sabine Roeser[2]

Abstract

AI systems and tools are being implemented at an increasingly rapid rate in society for a variety of purposes such as decision-making, managing job applications, and socializing. These new technologies have a lot of promise but may also introduce new risks by threatening human moral and relational values, as well as values connected to flourishing. Mainstream approaches to risk assessment do not pay sufficient attention to these values. The study of emotions as they are connected to human values can therefore play an important role in risk management. We will contribute to this discussion by introducing the concept of human needs, or what we consider to be the sources of values that constitute emotions. This brings a new perspective to the debate around AI and risk. By combining insights from Martha Nussbaum and Soran Reader, we argue that while emotions are crucial for highlighting what values are activated in a particular situation, the sources of an important part of human values are human needs. This provides for what we call the 'needs-values-emotions nexus'. We argue that this framework can add to the discussion about the ethical risks of AI in two fundamental ways. First, highlighting the crucial role of needs helps to explain why AI systems cannot develop, feel, nor reason according to human values. On the most basic level, AI systems lack a constitutive part of these values, i.e., they lack needs. The deployment of AI, for example to replace human decision-making, may therefore threaten human values. We discuss this by zooming in on a recent example, the so-called Dutch tax benefit scandal. Second, this paper argues that we need emotions to concretize and deliberate on what values are at risk when developing and using AI technology. Further building on the 'needs-values-emotions nexus' developed in this paper, we argue why art is a preeminent medium to elicit emotions and ethical reflection on the risks of AI. Discussing a concrete example, we illustrate how contemporary artists can contribute to ethical risk-assessments by focusing on the societal impact of AI.

Keywords AI · Risks · Needs · Values · Emotions · Art

Introduction: AI, values and risk

AI systems, such as large language models, are rapidly advancing and are already used in decision-making and to increase efficiency in large societal institutions and organisations.[1] They are also often expected to outperform human reasoning skills. These technologies are being deployed in organizations and institutions at an astonishing speed for numerous purposes and operations ranging from making decisions on loan applications, conducting job interviews, producing explanations in management, and for socialising (Yanqing Duan et al., 2021, pp. 8–12). Hence, there are many optimistic expectations about these technologies, which have led to their quick adoption and rapid development. However, these developments also give rise to many ethical challenges, as has been pointed out by researchers working on the ethics of AI (cf. e.g., Coeckelbergh, 2020; Floridi, 2023). Leading technology developers and philosophers argue that AI could pose existential risks

✉ Maria Danielsen
  maria.danielsen@uit.no

  Sabine Roeser
  S.Roeser@tudelft.nl

1 UiT The Arctic University of Norway, Tromso, Norway

2 Department of Values, Technology and Innovation ,
   Delft University of Technology (TU, Delft) , Delft, The
   Netherlands

---

1 See e.g., Ahmad et al. (2023).

in the future (Center for AI Safety, 2023), and some of them have proposed imposing a moratorium on its development (AI Moratorium, 2023). Critics argue that these concerns and proposals contribute to the AI hype and distract from more mundane ethical problems with AI that need more urgent attention (e.g., Eisikovits, 2023). These problems include environmental burdens (van Wynsberghe, 2021), and the impact AI has already had on human decision-making, e.g. by introducing and enforcing biases, unfairness, and lack of transparency, explainability and trustworthiness (cf. e.g., Duran et al., 2022); Buijsman, 2022, Santoni de Sio & Meccaci, 2021); Aizenberg et al., 2023). Furthermore, values and research integrity are at stake in building AI systems related to the enormous amounts of data that such systems draw on. How should data be collected, stored, and shared? Which data can and should be used? Who owns the data? For example, ChatGPT uses data that is available on the internet without crediting the data's creators or authors. Furthermore, how should statistical information feeding into AI systems inform policies in a responsible way? What about informed consent, privacy, and data protection? These are by no means exhaustive lists of ethical issues with AI. Indeed, there is a boom in the research on the ethics of AI.[2] What we aim to illustrate with this limited overview is that AI technology can impact important human values.

In this paper we contribute to this vastly growing literature by discussing the importance of human needs, embodiment, and emotions for shaping and upholding values and for understanding the intrinsic limitations and risks of AI systems. We propose a model showing how these concepts are interconnected in the following way: Human needs emerging from our embodiment are the source from which we develop human values, and human values are a main constituent of emotions. We propose that acquiring human values and emotions requires embodiment and, therefore, needs similar to our own. While reasoning about instrumental values such as those related to economics might partially be delegated to an AI system, we argue that AI systems cannot genuinely reason according to fundamental human relational, moral, and eudaimonistic values. This is because AI systems lack the foundation of human values, namely needs, which are produced through our embodied nature, and which are foundational for human emotions. Why is this a problem? Tools such as toasters and drills do not feel, perceive, or reason, but we also do not need them to. However, current AI systems are not ordinary tools, but systems used for tasks and purposes that involve moral reasoning and decision-making. This means that when using AI, we face different risks than we do when engaging with other types of technologies.

Furthermore, tech-developers have already set the next spot on the horizon of the evolution of AI and other digital technologies: they envision the future of visceral and sensory experiences and of connecting with others as taking place inside the Metaverse, which is de-connected from the physical world and, instead, evolves in front of our eyes.[3] Tech-developers emphasize abstract rational processes that supposedly can be delegated to automated systems, thereby suggesting a utopian world that is driven by highly optimized digital processes. Technological developments such as Neuralink and the Metaverse involve moving away from our embodied nature as vulnerable and needy beings who interact physically with each other. We argue that this shift would miss a crucial dimension of what it means to be human, namely, to be embodied, needy beings for whom emotions and values are crucial. This shift gives rise to additional ethical challenges, many more than the ones typically discussed in the literature on the ethics of AI: in their lack of embodiment, needs, and emotions, AI technologies also lack a genuine, substantive ethical perspective.

In addition, mainstream approaches to assessing the risks of AI also lack the emotions and values needed to inform human-centered analyses this new technology. While the field of risk ethics started to develop a few decades ago, risk and uncertainty are still relatively underexplored fields in moral philosophy. Risk managers in turn have little awareness of the evaluative aspects of their work. Risk management is typically grounded in formal decision theory, operating with (supposedly) quantifiable criteria that are also often translated into economic terms. According to this approach, risk is calculated in terms of probabilities multiplied by unwanted outcomes and then applies a risk-cost-benefit analysis (Asveld & Roeser, 2012, pp. 3–4).

The limitations of this quantitative approach have been pointed out by risk ethicists and social scientists (cf. e.g., Fischhoff et al., 1981; Krimsky & Golding, 1992; Shrader-Frechette, 1991; Slovic, 2000; Roeser et al., 2012; Hansson, 2014). Rather than being value neutral, quantitative approaches such as risk-cost-benefit-analysis rely on implicit values, such as having a consequentialist structure, or using monetary quantifications of human lives, e.g., in terms of QUALYS (quality adjusted life years), and by having an anthropocentric viewpoint in which non-human lives are at most considered to the extent that they affect human well-being or have economic implications. These values are far from unproblematic, but they are hidden from explicit critical scrutiny as these approaches are typically

---

[2]   So much so that the ethics of technology now risks getting equated with the ethics of AI in academia as well as in public debates, thereby threatening to overlook the ethical challenges of other technologies.

[3]   Mark Zuckerberg: First Interview in the Metaverse. Lex Friedman podcast #398. https://www.youtube.com/watch?v=MVYrJJNdrEg&t=2780s

presented as purely quantitative and value-neutral. This is ethically problematic from a procedural point-of-view, as the value implications of formal approaches are not transparent (Shrader-Frechette, 1991). It is also problematic from a substantive ethical point-of-view, as such quantitative approaches lack explicit consideration of important moral values such as justice, fairness, responsibility, autonomy, and what it means to lead a good life (cf. e.g., Hansson, 2014; Roeser et al., 2012; Fahlquist, 2020). Many approaches around understanding the ethical risks of AI do pay attention to moral values. However, these approaches, as well as quantitative approaches to risk, do not take into consideration the importance of emotions. However, emotion research by philosophers and psychologists tells us that emotions can indicate the presence of important moral values (Frijda, 1986; Solomon, 1993; Little, 1995; Goldie, 2000; Roberts, 2003; Roeser, 2011; Roeser & Todd, 2014). As Nussbaum (1991), on whose work we build in this paper, phrased it in another context:

> (T)he economic mind is blind to the qualitative richness of the perceptible world; to the separateness of its people, to their inner depths, their hopes and loves and fears; blind to what it is like to live a human life and to try to endow it with a human meaning. (p. 888)

Here, Nussbaum describes a limitation of thinking in quantitative, or economic terms when it comes to cases and situations involving people's lives. Inner depths, hopes, loves, fears, and meaningfulness suggest important, qualitative aspects of human lives which Nussbaum connects to values and emotions. Although Nussbaum has made this statement in the context of public policy in general, it has been argued that paying attention to emotions can make an important contribution to ethical decision-making about technological risks (Roeser, 2018a; Roeser & Steinert, 2019).

In this paper we will build on these ideas and argue that emotions and values should have a dual role in thinking about the ethical risks of AI: (1) by showing that one of the risks of AI is that it lacks emotions and values because of its lack of needs, and (2) by arguing that risk assessment of AI should also involve emotions and values. We argue that art can help with these dual roles of emotions and values, (1) by showing the emotional shortcomings of AI, and (2) by eliciting emotions and thereby highlighting values related to the ethical risks of AI.

The importance of emotions as expressions of values and needs

In this section we outline an understanding of emotions that builds on Martha Nussbaum's influential philosophical account of emotions. Nussbaum argues that emotions are evaluative beliefs, or value-judgments, relating to human values. She defines human values as a set of values concerned with flourishing, including moral and relational values. While we agree with Nussbaum's central arguments, we argue that it can be fruitfully complemented with Soran Reader's theory concerning the foundation for human values. Reader argues that the most central human values are ultimately grounded in human needs (Reader, 2011). By combining insights from Reader's view on needs and values with Nussbaum's account of values and emotions, we develop a view that does three things. In section 2a, we show how Reader's theory complements Nussbaum's account of emotions with an account that proposes needs as an important source of human values. Using building blocks from the ethics of needs, we show that core human values derive from needs. In section 2b, we show how by using insights from Nussbaum and Reader together, it is possible to connect human needs and values to emotions, something needs-ethicists have overlooked. Finally, in section 2c, we argue how the theoretical lens constructed through this combination helps to reveal a crucial difference between humans and AI technologies: AI technologies cannot develop nor feel or reason according to human values because, at a fundamental level, they lack human needs.

Needs and values

What are needs? There are many approaches to understanding needs, each with slightly different formulations (see for example Maslow's hierarchy of needs; Poston, 2009, Doyle & Gough's list of universal needs, cf. Doyal & Gough, 1986). Formulations of universal needs also play an important role in international law (cf. Stewart, 1989). Examples of fundamental needs are physical health, critical autonomy, and cognitive as well as emotional capacity (Doyle & Gough, 2003 p.13). However, in what follows we will not focus on formulations of specific needs, but on the insight that needs are fundamental to flourishing and to morality, as, again, philosopher Soran Reader has argued. Reader defines needs as inalienable requirements without which our very being is at risk (Reader, 2011, p. 594). Unless those elementary needs are met, a being is damaged and incomplete. To Reader, needs are essential to ethics because our most basic inalienable needs are the necessary conditions of our being, and therefore needs are the foundations for important human values (Reader, 2011, pp. 594-5).
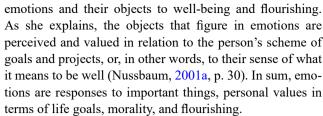
While we agree with needs-ethicists on the importance of needs for grounding values, we think that something important is missing from their account, namely the link between values and emotions. This is evident, for example, within care ethics where care is understood as an essentially emotional state. Care ethicist Nel Noddings (1995) states that the moral act of caring is essentially responding to needs (pp. 8–13).[4] In addition to care, we would like to emphasize that there are other important emotions linked to needs and values. To establish this link, we propose connecting the discussion about needs ethics with Nussbaum's account of emotions. While needs ethicists argue for the close link between needs and values, Nussbaum explains the link between values and emotions. In the next subsection we show what the connection between needs, values, and emotions looks like.

Emotions, values, and needs

Nussbaum sees emotions as a subclass of value judgments that involve important attachments, what she calls "intentional objects," and that correspond to the most important goals and projects in a person's life (Nussbaum, 2001a, pp. 22,30&77). The value underlying the object of emotion, i.e., the values to which the emotion points, tells us what it is that makes us flourish. Therefore, the set of values constitutive of emotions are *eudaimonistic*—a category which also includes relational and moral values such as autonomy and compassion. As evaluations, emotions embody ways of seeing and responding to their objects (Nussbaum, 2001a, p. 27). Notice that to Nussbaum, emotions have a dual role. First, they connect the person experiencing the emotion to their personal values. Second, emotions are evaluative as they indicate how a person is relating to one of their values in the moment. For example, pleasure that accompanies a sympathetic gesture from a colleague is a response to the value of being affiliated. Grief from losing someone close is a response to the value of having significant relationships. The intentional object of the emotion is e.g., the person giving the sympathetic gesture or the deceased loved one, seen in the light of a specific value, namely the value of being affiliated. That does not mean that each emotion connects to one specific value. Some emotions operate by referencing clusters of values, even conflicting values. The crucial point is that emotions are responses to objects invested with value and importance; in the examples given above, these values are affiliation and having significant relationships (Nussbaum, 2001a, p. 22). Moreover, Nussbaum connects

emotions and their objects to well-being and flourishing. As she explains, the objects that figure in emotions are perceived and valued in relation to the person's scheme of goals and projects, or, in other words, to their sense of what it means to be well (Nussbaum, 2001a, p. 30). In sum, emotions are responses to important things, personal values in terms of life goals, morality, and flourishing.

This brief summary of Nussbaum's work is meant to illustrate how she sees the connection between values and emotions. However, her account does not provide for an explanation of the foundation of values. And while her account emphasizes the cognitive aspects of emotions, she does not always explicitly emphasize their embodied, affective, and motivational aspects, which also relate to our nature as needy beings. We suggest that an important key to locating the foundation of human values lies in incorporating these aspects into the account of emotions. Where Nussbaum seems to equate emotions with cognitive aspects such as evaluative beliefs or value judgments, other cognitive theories of emotions also emphasize the embodied, affective, and motivational aspects of emotions (e.g., Frijda, 1986; Roeser, 2011). We add these dimensions to the account of emotions we pursue in this paper. Therefore, when we draw on Nussbaum's work, the reader should keep in mind this richer account of emotions that encompasses Nussbaum's insights but also emphasizes the importance of our embodied and needy nature. While Nussbaum indeed mentions neediness and vulnerability in relation to values, she does not incorporate these aspects explicitly into her definition of emotions. By complementing Nussbaum's account of values and emotions with these additional aspects and with Reader's account of values and needs, a structure emerges that combines both accounts. We propose the following framework:

Needs-values-emotions nexus:

Human needs → human values → emotions

In other words, human needs are sources of important human values (Reader), which, in turn, are main sources of emotion (Nussbaum).

Human values are connected to human needs. For example, the value of trust derives from the need for having relationships and attachments to others. In this example, according to Nussbaum's framework, emotions are evaluations as to whether a given person is trustworthy. Perceiving someone as being trustworthy feels different emotionally than perceiving them as being untrustworthy.

In the context of her work on the Capability Approach, Nussbaum objects to the idea of universal needs, especially within politics and policymaking.[5] Nussbaum, together

---

[4] We draw here on the common insights between the ethics of care and the ethics of need, namely seeing needs as fundamental to moral and eudemonistic values and avoid discussing their conflicting view regarding the universality of needs.

[5] Examples of important publications on the debate between Nussbaum's view related to the Capability Approach and her objection towards the needs-view are e.g., Reader, 2006 and Alkire, 2005.

with Amartya Sen, developed the Capabilities Approach as an alternative to the needs-view. They introduced the term "vital functions" in the place of what others have referred to as "needs." Even though Nussbaum sees a theory of universal needs as problematic in the context of distributive justice, we think that it is possible to defend the framework that we propose within Nussbaum's approach. Indeed, Nussbaum mentions the link between emotions and needs in several passages of her writings on the philosophy of emotions. For example, she defines emotions as judgments about important things in which we acknowledge our neediness and incompleteness before things we value but do not control. To Nussbaum, emotions are concerned with values whose fulfilment also depends on things outside of us. For example, fulfilling the value of love depends on another (human) being. Neediness and vulnerability are baked into emotions precisely because emotions communicate that our most important goals and projects are at stake. Nussbaum maintains that emotions reveal rich resources for morality, and stresses that without emotions, morality could not come into being (Nussbaum, 2001a, p. 222). In *The Fragility of Goodness*, she states that part of the peculiar beauty of human moral excellence is its neediness and vulnerability (Nussbaum, 2001b, p.2). Furthermore, in "Mortal Immortals: Lucretius on death and the voice of nature," she explains that limitation and need are integrated parts of human values. In her own words: "(M)any of the activities that we value most and in which we take most delight, do have an internal structure of need and repletion, absence, and presence" (Nussbaum, 1989, p.330). Hence, Nussbaum's work echoes the insights we propose in our 'needs-values-emotions nexus.'

Needs, values, emotions, and AI

Drawing on the framework of needs, values, and emotions we have outlined in the previous subsections, what we called the 'needs-values-emotions nexus,' we now discuss why this provides for crucial insights into the ethics of AI. We argue that because AI does not have needs, embodiment, nor emotions, it cannot be genuinely sensitive to values. This also relates to the following argument from Nussbaum (1989):

> The virtue of 'courage' consists in a certain way of acting and reacting in the face of death and the risk of death. A being who does not face the risk of death, cannot have the virtue of courage. This means, as well, that the component of friendship, love and love of country[6] that consists in a willingness to give up one's

life for the other must be absent as well—indeed, must be completely mysterious and obscure to people whose experience does not contain the sense of mortality. Beyond this, we begin to discover that many of the virtues we prize require an awareness of the limits and needs of the human body that will be absent, as such, from a being that can never die. (p. 338)

This observation is strikingly and directly relevant to AI. AI is often celebrated or even feared because of its perceived potential to eventually outperform human beings. However, there are crucial aspects of humanity that AI cannot replace, namely a moral point-of-view with feelings of compassion and care. These aspects require beings with embodiment, needs, values, and emotions similar to ours, and constitute an important part of moral reasoning and ethical decision-making.

Could we also apply Reader's definition of needs as "inalienable requirements without which our very being is at risk" to AI?[7] For example, an AI 'needs' data in order to exist. This gives rise to another question: who gets to interpret whose existence and what needs this existence would require? Does the AI system itself get to decide what it needs, or is this something left to humans? In the same vein one could argue that a car has needs and hence values, e.g., the need to be fuelled (with electricity or gasoline, etc.); however, to which or whose values does this need relate? Presumably, in the case of the car these are ultimately human needs and values related to mobility. In a 2024 article Bales et al. propose that some advanced AI models are likely to function as agents pursuing goals (p. 2). To reach their final goals, for example decreasing criminal rates and proving mathematical theorems, AI might develop subgoals such as seeking money, power, or other resources. If subgoals instantiate things AI systems need to reach their final goals, then, according to our model, subgoals serve the same functions in AI as needs do to humans. However, following our model, if AI systems develop values based on these subgoals, these will be different from human values, and certainly not qualify as eudaimonistic nor moral in nature. If AI's existence—its being—is defined by the subgoals it determines are needed to reach a larger goal, for example gaining capital and power in order to decrease criminal rates or prove a mathematical theorem, it seems that both their being and their needs must be understood in reference to human needs and goals. According to Reader and Brock (2002), human needs are oriented towards our final goal, or end, namely flourishing (p. 430). Human needs are distinguished from the needs of machines/technology,

---

crucial to the main insight we would like to draw from this passage.

[6] The notion 'love of country' may strike some readers as more controversial than the other aspects Nussbaum mentions, but it is not

[7] We would like to thank an anonymous reviewer for making this point.

such as cars and Large Language Models (LLMs). AI "needs," then, are very different from our own. The needs, values, and emotions of human beings require an embodied existence and an internal point-of-view, things an AI does not have.

Technology is often conceived of as value-neutral. Relatedly, engineering is seen as a mathematical, quantitative discipline, and engineers as the archetype of people who make decisions in a detached, rational, and quantitative way. In this framework, AI can then be seen as the pinnacle of technological as well as human development: not only is it a sophisticated technology, but the ambition is that it may not only assist with but eventually supersede and replace human decision-making. AI is, then, the ideal abstract intelligence. However, as philosophers of technology have pointed out over the last decades, technological design is not value-neutral (Verbeek, 2005). Approaches such as Value Sensitive Design (Friedman & Hendry, 2019; Van den Hoven & Manders-Huits, 2017) and Responsible Innovation (Fischer et al., 2024)) are focused on developing technology that it contains or instantiates values we find desirable. Furthermore, it has been argued that emotions should play a significant role in assessing the value dimension of technology (Roeser, 2018a). While AI is sometimes celebrated for its supposed "pure" rationality, conceived of as being free from emotions, the account we developed in the previous passages indicates that this is actually a shortcoming, as emotions are crucial in providing for *practical* rationality, specifically regarding a perspective on human values and needs. We develop this idea in what follows.

Ethical problems of AI and the importance of emotions

As we discussed in Sect. Introduction: AI, Values and Risk, artificial intelligence (AI) has become a major issue in societal debates, technology development, and ethics research. What we want to add to the booming literature on the ethics of AI is that an overlooked but important issue is the role of emotions. This is crucial in a twofold way: first, attending to emotions can highlight why it is problematic to delegate ethically sensitive decision-making to AI, as AI lacks human embodiment and needs and therefore cannot develop human values and emotions. Consequently, AI systems lack characteristics that are crucial for ethical decision- making. Second, emotions can play a vital role in being aware of the values at stake in the development and use of AI. We will illustrate this by discussing a hitherto unresolved case in the Netherlands.

'Unprecedented injustice': AI and the Dutch tax benefits scandal

The causes of the now so-called Dutch tax benefits scandal started around 2005. The Dutch tax authorities built an AI system that was based on racist assumptions and a reversal of the principle 'innocent until proven guilty.' This meant that families with dual citizenship and non-Western immigration backgrounds were under extra scrutiny by an automatized decision system. If they made a tiny mistake in their tax declarations and this was discovered, they were put on a blacklist and the system required them to pay back enormous amounts of tax benefits that were unrelated to the small mistake. This led to insurmountable Kafkaesque bureaucratic hurdles and personal tragedies for countless families. Nobody in the government was responsive and, therefore, no authority could be held accountable. Many families were eventually left financially ruined; there were many broken marriages, and numerous children were taken away from their families. It took years until Dutch authorities and the government paid attention and acknowledged that there was a problem. The parliamentary inquiry committee titled their report "Unprecedented Injustice" (House of Representatives, 2021).[8] Eventually the scandal led to the fall of the Dutch government in 2021.[9] Many families have yet to receive their promised compensation. Even in 2025, there is still no solution in sight. Hence, far from being a panacea, automatization has been used without proper control in complex, unpredictable processes that actually would require close attention, oversight, and management. The Netherlands is one of the most digitized and wealthiest countries in the world, and it is usually a well-organized country that scores low on corruption indexes etc., hence this case may be an indication of things that can go wrong elsewhere in the future. Furthermore, there are recent signals that similar things have happened in other automatized systems in the Netherlands (NOS, 2023), meaning that this

---

[8]   A follow-up report with the telling title "Blind to Human and Law" was published recently that provided further evidence on the structural institutional biases underlying this scandal, and it states that similar cases can happen any time in the Netherlands (Tweede Kamer, 2024).

[9]   After re-elections, it took a record-time of 9 months to form a new coalition– which was ultimately the same as the previous one. After only 1.5 years this government fell as well, again due to immigration issues. In the elections that followed, the right-wing extremist party PVV led by Geert Wilders had a landslide victory. This resulted in a right-wing coalition with anti-immigration policies at its core. At the time of finalizing this article, this government has just fallen after 11 months because Wilders did not find its anti-immigration policies radical enough. The measures underlying the tax benefit scandal discussed in this section of the paper should be seen in the context of a growing anti-immigration sentiment that has been looming for a long time in Dutch society (as well as in other Western countries).

may be the tip of the iceberg in terms of things that can go wrong when delegating important decisions to AI systems that are built on ethically problematic assumptions.

The ethical risks of leaving out values and emotions in designing and using AI

There are many important lessons that one can draw from the tax benefits scandal. We zoom in on a few aspects that are salient in light of the context of this paper, starting with the problematic use of statistics and risk assessment in this case. The question is how statistical information feeding into AI systems should inform policies in a responsible way. Tax authorities used an automated system that was built using empirical correlations based on past data and which were then used to predict the future. Making predictions about human beings' future behaviour using data compiled in the past is problematic, as it overlooks the possibility that peoples' changed circumstances or choices have caused them to behave differently. Disregarding this possibility is exactly how stereotypes are introduced. Furthermore, the fact that someone belongs to a more risk-prone demographic group should not be used to make predictions about them. This is even more problematic if this is not only used as a prediction, but as a basis to judge individual people and make concrete decisions about them with far-reaching consequences. Innocent until proven guilty is a key component of a democratic justice system and its rule of law. AI systems that work on such predictive models undermine this very basis, which was exactly what happened in the Dutch tax benefits scandal.

The Dutch tax benefits scandal illustrates the ethical challenges of quantitative risk assessments. As discussed in Sect. 1, values are intrinsic to supposedly pure formal approaches. Next to the ethical challenges related to predictions and stereotyping mentioned above, there are other ethical challenges underlying formal approaches to risk assessment that also need to be considered when designing AI systems used in the context of public policies: How do we measure and compare different sorts of well-being? How do we value a human life? These questions cannot be answered using a quantitative approach. Rather, they require an explicit ethical stance, which in turn requires emotions, as they can attune us to important values as well as to the shortcomings of reductive quantitative approaches. For example, quantitative approaches often work with concepts such as QUALYs or risk assessments on a societal level. However, such methods overlook the intrinsic value of each individual human being, and the challenges of how to quantify human well-being and human rights. Emotions can help us to take on a compassionate perspective and to see things from the point-of-view of potential victims of a quantitative

assessment. Furthermore, methodological assumptions are often arbitrary and can make a dramatic difference to a risk assessment. For example, depending on which value one assigns to a QUALY, one can get different and even opposite results in a risk-cost-benefit analysis, and as discussed in Sect. 1, such methods do not take into account crucial moral values such as the fair distribution of risks and benefits, nor of autonomy.

The use of formal models and AI systems suggests objectivity, but this blurs the underlying substantial ethical considerations rather than making these explicit and subject to critical deliberation. Many decisions are inherently morally complex. It is better to face this moral complexity explicitly, even though this means that one cannot fall back on a given, clear-cut methodology. Modelling moral trade-offs requires explicit moral reflection; it cannot replace it. "Abstract intelligence" cannot do this on its own because of its inherent blindness to human values. Modelling moral trade-offs can then at most be a tool for ethical reflection in an iterative process. Emotions such as compassion and feelings of responsibility and care can play a crucial role in drawing attention to possibly overlooked values, and in highlighting complex dilemmas. Hence, automated decision-making should at most be a tool in ethical deliberation, embedded in human decision-making that also involves emotions.

Hence, if we analyse the case of the Dutch tax benefits scandal using the framework we developed in Sect. 2, the 'needs-values-emotions nexus,' we can see that in order to address these challenges, a humane, caring approach is needed and should form the basis of the employment and development of AI. Rather than naively hoping that AI can replace or even outperform human decision-making, our framework highlights the importance of a humane, ethical perspective, as decisions that impinge on human lives are, by definition, value-laden. A humane perspective means a human perspective, with the capacity to pay attention to and understand human needs, values, and emotional impact. Furthermore, emotions such as care and feelings of responsibility can help to achieve sensitivity and alertness to ethical challenges. In the case of the Dutch tax benefits scandal, explicitly unethical, racist, values were built into the system. Modelers use the slogan 'garbage in, garbage out;' in other words, bad input into an automated system will generate bad output. In the context of the ethical evaluation of AI, we might say 'moral garbage in, moral garbage out.' In this case the "moral garbage" was the racist assumptions built into the system. The developers of these AI systems should have used better, non-racist values. This underscores our point. It takes a humane, caring perspective and a sense of responsibility to assess the fundamental wrongness of the approach that was underlying the system. It shows that not any value and concomitant emotion will do: racist and xenophobic
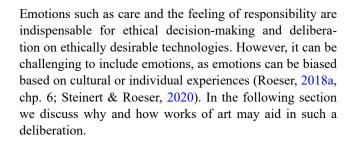
values and emotions need to be counteracted by feelings of compassion and respect for everybody's fundamental human dignity and worth.[10] However, emotions are no guarantee that we have the right moral values, as moral emotions can be biased and related to problematic stereotypes and values (Steinert & Roeser, 2020). Nevertheless, attention to emotions is necessary, but it should involve 'emotional deliberation', i.e. critical reflection on and with emotions (cf. Roeser & Pesch, 2016; Roeser, 2018a, Chap. 6).

The example shows that in an age of automatization, we still need to appeal to the integrity and courage of employees in order to resist the demands of authorities who do intentionally unethical deeds, even if these are to be translated into abstract computer code. The code becomes very concrete, value-laden, and ethically problematic when the 'computer says no' to a family whose livelihood is at stake without a serious fault on their side, and with nobody willing to investigate their case, as happened to thousands of families in the Netherlands. A compassionate, humane perspective could have helped to be alerted to the injustice as well as the vulnerability and needs of these innocent families.[11]

Hence, this case highlights the ethical cautions that should be included in the design, implementation, and use of AI systems. It can serve as an illustration and warning about the problems that AI systems can create. This also means that we need strict requirements for governments and businesses using and developing AI. Our argument aligns well with AI ethics approaches arguing that AI should be transparent, that there should always be a 'human in the loop,' and that 'meaningful human control' should always be involved (e.g., Santoni de Sio & Meccaci, 2021); Duran et al., 2022). Also, approaches to 'explainable AI' argue that we need to pay attention to the human perspective in designing AI (cf. Buijsman, 2022). What we would like to add is that the human perspective so crucial to the ethical design and deployment of AI should explicitly include attention to emotions like feelings of responsibility, compassion, and care for potentially affected people. This is because these emotions provide us with crucial insights into human values and needs. A lack of emotions and embodiment prevents AI from well-founded ethical decision-making. Human emotions are needed to understand and deliberate about the moral impact of AI and to develop ethically desirable uses of AI. Ethical decision-making is an irreplaceably human activity. Approaching such issues in a detached and quantitative manner risks overlooking crucial ethical values. Therefore, decision-making with an impact on human values should at most be supported by AI, not replaced by it.

Emotions such as care and the feeling of responsibility are indispensable for ethical decision-making and deliberation on ethically desirable technologies. However, it can be challenging to include emotions, as emotions can be biased based on cultural or individual experiences (Roeser, 2018a, chp. 6; Steinert & Roeser, 2020). In the following section we discuss why and how works of art may aid in such a deliberation.

Risk assessments through Art

In the previous section we highlighted how AI technologies may threaten human values as they lack needs, values, and emotions. In this section we add another dimension by discussing the connection between needs, values, emotions, and art. How can art contribute to eliciting ethical reflection about the risks posed by AI technology? As Martha Nussbaum explains, works of art appeal to compassion and imagination and invite the spectator to explore their own neediness and insufficiency towards what they most value (Nussbaum, 2001a, p. 244). Whereas widely-used quantitative approaches to risk ignore the category of human values, artworks bring them to the forefront, evoking emotions in the spectator, which can facilitate reflection about what matters most. Artworks can trigger reflections about values and future technological possibilities (Roeser et al., 2018). To show how this can work and why art is relevant to the AI-and-risk debate, we first (4a) provide a short discussion about how attention to art can complement our proposed framework on the 'needs-values-emotions nexus. We then (4b) include a discussion and (4c) interpretation of an artwork by contemporary artist, Toril Johannesen. We conclude that this artwork can make an important contribution to the ethical deliberation about the risks of AI.

Art, emotions, and technology

Martha Nussbaum has pointed out that works of art can help in the concretization of abstract ideas and can aid in ethical reflection.[12] She explains that when perceiving an artwork, the imagination moves from the general and abstract to the particular and concrete with the help of emotions (Nussbaum, 2001a, pp. 241–244). These insights can also be extended to the context of works of art that engage with technology (Roeser, 2018b). There are an increasing number of artists who engage with technology in their work. Such art can playfully explore the possibilities of technology that have not yet materialized or highlight ethical

---

[10] Thanks to an anonymous reviewer who raised this point.

[11] This is in line with one of the main insights of the most recent report of the second Parliamentary Enquiry Committee tasked with this case (Tweede Kamer, 2024).

[12] Consider e.g., Ane Graff, a contemporary artist who describes her practice as *visualizing and materializing philosophy*. https://artscene.no/2019/06/29/materiell-filosofi-en-samtale-med-ane-graff/.

dimensions that are hidden from sight. Artists are not bound by strict rules, meaning they have the freedom to exaggerate, play with, and scrutinize aspects of technology that tend to get overlooked. In this way, artworks that engage with technology can elicit people's emotions and imagination and explore surprising angles and viewpoints (Roeser et al., 2018). Whereas imagination is the ability to picture possibilities, emotions are evaluations of these possibilities as explained in section 2b. We argue that art provides a rich source of insight into values, also because of art's ability to elicit emotions. The dynamic between values and emotions can be elicited through works of art, often in novel ways. As such, works of art can broaden people's emotional and moral outlooks and facilitate emotional deliberation on the ethical aspects of risky technologies (Roeser & Steinert, 2019).

In the next subsection we present an artwork from the contemporary art scene to illustrate these ideas. We chose an artwork highly relevant for the discussion of risk and AI technologies. This piece stages the future of human vision as it is developed according to values such as prediction, certainty, and objectivity. The human eye itself, then, evolves in line with values and ideals constitutive of technology and the natural sciences.

"The invention and conclusion of the eye" (2017) by Toril Johannesen

Toril Johannesen's artwork "The Invention and Conclusion of the Eye" is an audio play about vision.[13] The story follows Mx, a scientist working on an article for a lexicon about the evolution of the eye and its future development. The play opens with three assertations made in Mx's voice, about the future of vision: (1) "All we see, we see alone, but everything we see is the same". (2) "The world you and I live in is about to become optically invisible." (3) "The eye will retreat with time" (Johannesen, 2017). Mx creates the article together with a language-learning model who goes by the name Triple Star, and with which she communicates throughout the play. Triple Star holds information about the eye and uploads new information in real-time according to what Mx postulates. At some point, when Mx tries to discuss something or ask questions, Triple Star just echoes what Mx said at an earlier stage.

Mx describes the future of vision—called "clear sight,"—as seeing without eyes, a stage during which technology enable the brain to form mental images without using the eye-organ. Evolution has incorporated positivist ideals and values from the natural sciences such as objectivity, certainty, and prediction. The human eye develops according to this *scientific* category of values, and consequently, evolution causes *human* values to withdraw.[14] Mx states: "What we see, how we see and what we understand by seeing are facing radical changes. New technologies make cameras and eyes more or less the same" (Johannesen, 2017).

To Mx, vision at its current state is connected to subjectivity, i.e., to viewpoints that create a fog of opinions rendering them impossible to confirm. Because she idolizes the natural sciences, she wants prediction and clarity instead. She is therefore enthusiastic about the idea of vision as it is paired up with technology and considers its ability to enable visceral objective knowledge an advantage. As she explains to her computer: "With the help of abstraction and calculation, in collective neural networks we will be able to imagine the world, not just as it appears, but as we calculate that it must be" (Johannesen, 2017).

Clear sight is the ability to predict the future based on probability, forming mental images of numbers and statistics, and producing objective knowledge. Eventually, the human eye retreats and everyone sees everything in the same way. Mx delightfully concludes that the individual shaping of knowledge based on sight will soon be over. Hearing about these prospects might evoke emotions such as curiosity and eeriness depending on how the listener relates to their ability to see, and what their individual sight means to them.

Interpretation of this artwork in the context of the risks of AI

Staging a hypothetical future and possible implications of AI technologies as "Neuralink," the "Metaverse," "BNCI technologies," and "Clearview AI," this artwork can be seen as a reduction of positivist viewpoints.[15] The piece shows how AI technologies may threaten human values because they leave out what is essential to an embodied, situated human being. Focusing on eyes, vision, and sight, the work highlights the connection between subjective sense experience and epistemology, in gaining knowledge. The style and presentation of the artwork are non-dogmatic and non-instructive but invite the spectator to imagine and evaluate human values and human needs as they become consistent with, transform into, or even are reduced to the values constitutive

---

[13] Johannesen is a leading Norwegian artist: toriljohannessen.no. This artwork has been shown at OSL Contemporary in Oslo, at Plug In ICA in Winnipeg, and at the nonprofit, artist-run space called Small Projects in Tromsø. For more information on the exhibition, see https://www.oslcontemporary.com/exhibitions/toril-johannessen.

[14] In a certain sense, scientific values are human values as well, as Mary Midgley has pointed out so poignantly in her work (e.g., 2003). But here, we mean to distinguish scientific values such as objectivity from human values such as well-being, etc.
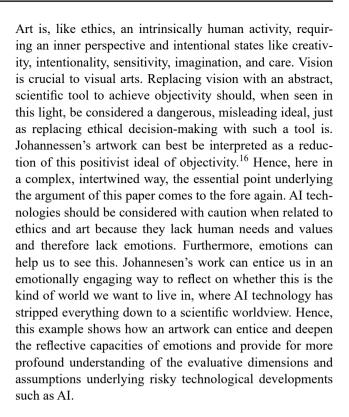
[15] As we interpret the work, Johannesen has not explicitly stated what technologies her artwork thematizes.

of technology and the natural sciences like objectivity, prediction, certainty, and non-ambiguity. Mx suggests that we can and should obtain the correct, objective view, scientifically speaking, and that in this way we should discard an important part of the human sensory system which is seen in relation to subjective viewpoints, namely sight or vision. A subjective point-of-view is infused with personal values that, according to Mx, a scientist, are best left behind. Consider for example the human value of "critical participation" (Doyle & Gough, 2003, p.235) and vital functions such as "senses and practical reasoning" (Nussbaum, 1999, p.235). The futuristic scenario featured in "The Invention and Conclusion of the Eye" presents a reality where these values and functions, have been reduced to include only scientific values like prediction, calculation, certainty, and objectivity. The listener may imagine and reflect on the possibility of forming mental images of "objective viewpoints" instead of going through the processes involved in obtaining subjective points-of-view. Is this something we want? Faced with this question, emotions such as the fear of losing something that is at the core of what makes us human beings play a crucial role, namely because such emotions connect us to our values and therefore to our needs. These emotive responses can help us become aware of how we feel within the context of AI development.

Johannesen's work is illustrative of the ambitious, positive, and utopian view underlying many expectations of AI, showing the worldview upon which these expectations are based: a scientistic, logical, positivist view in which scientific objectivity is the ideal and a human perspective is seen as a distraction (also cf. Midgley, 2003, who criticizes this worldview in her work). However, as our discussion of the Dutch tax benefit scandal illustrates, this means that AI is intrinsically misguided when it comes to matters that are related to human beings, and our needs, vulnerabilities, and emotions. The positivistic, scientistic approach may be appealing when it comes to a purely descriptive natural science view of the world. But when it comes to policy-making and decision-making about human beings, a human perspective is needed. Johannessen's artwork shows us the disconnection between certain expectations of technology and the real, complex world of needy and vulnerable humans in which such technologies are employed.

It is possible to think that AI technologies can replace humans when it comes to matters that are intrinsically value-neutral (supposing that such matters exist), and that one should be cautious when it comes to using AI technologies for matters that are ethically poignant, along the lines of our argument in Sect. 3. However, let us point to another interesting aspect of Johannessen's work. The fact that she as an artist explores this positivist ideal in this extreme way indicates that things are even more complex.

Art is, like ethics, an intrinsically human activity, requiring an inner perspective and intentional states like creativity, intentionality, sensitivity, imagination, and care. Vision is crucial to visual arts. Replacing vision with an abstract, scientific tool to achieve objectivity should, when seen in this light, be considered a dangerous, misleading ideal, just as replacing ethical decision-making with such a tool is. Johannessen's artwork can best be interpreted as a reduction of this positivist ideal of objectivity.[16] Hence, here in a complex, intertwined way, the essential point underlying the argument of this paper comes to the fore again. AI technologies should be considered with caution when related to ethics and art because they lack human needs and values and therefore lack emotions. Furthermore, emotions can help us to see this. Johannesen's work can entice us in an emotionally engaging way to reflect on whether this is the kind of world we want to live in, where AI technology has stripped everything down to a scientific worldview. Hence, this example shows how an artwork can entice and deepen the reflective capacities of emotions and provide for more profound understanding of the evaluative dimensions and assumptions underlying risky technological developments such as AI.

Conclusion

In this paper, we highlighted the importance of emotions for assessing the ethical risks of AI systems and for pointing out the intrinsic shortcomings of AI when it comes to ethically sensitive decision-making, precisely because AI does not have emotions. We developed an account that considers human needs as sources of human values, and human values as sources of emotions. While some philosophers show a connection between emotions and values aimed at flourishing (e.g., Nussbaum, 2001a; Roeser, 2011), others show the link between needs and values aimed at flourishing (Reader, 2005, 2011). In this paper, we structurally connected these approaches to form what we call the 'needs-values-emotions nexus.' We pointed out why this nexus is inaccessible to AI and therefore distinguishes humans from AI, most fundamentally because AI lacks human needs because it is neither embodied nor alive. Consequently, AI technologies are insensitive to human values and emotions, and the implementation of these technologies may threaten human values. We illustrated this by discussing the recent Dutch tax benefits scandal. We stressed that an attention to emotions is

---

[16] Interestingly, the leading artist Anselm Kiefer has said that for him, art is more real than the world surrounding us. He thereby upturns the hierarchy between the supposed objectivity of scientific description of the world and art as mere subjective imagination, which is also exemplified in Johannesen's artwork.

crucial in order to be sensitive to important needs and values. Furthermore, we argued that art can assist in the ethical reflection on AI. Art can appeal to emotional capacities such as imagination and compassion, which can help to emphasize and concretize human values. The artwork we discussed in this text highlights certain tendencies within AI technology, by making bold claims about supposedly objective technologies replacing subjective human experiences. We discussed how this artwork can help the audience to imagine and evaluate the implications of replacing human values with values integral to technology and the natural sciences. This amplified our argument that a humane perspective on ethical decision- making is essentially a human perspective that is not available to AI technologies, as it requires the interconnected link between needs, values, and emotions.

Data Availability No data have been collected for the writing of this paper. All discussed information is based on the sources provided in the paper, i.e. publicly available information from news agencies, government agencies, websites and academic literature.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications,10*(1), Article 311. https://doi.org/10.1057/s41599-023-01787-8

AI moratorium is necessary to avoid extinction. (n.d.). Retrieved May 26, (2025). from https://moratorium.ai/

Aizenberg, E., Dennis, M. J., & Van Den Hoven, J. (2023). Examining the assumptions of AI hiring assessments and their impact on job seekers' autonomy over self-representation. *AI and Society,40*(2), 919–927. https://doi.org/10.1007/s00146-023-01783-1

Alkire, S. (2005). Needs and capabilities. *Royal Institute of Philosophy Supplement, 57*, 229–251. https://doi.org/10.1017/S1358246100009231

Asveld, L., & Roeser, S. (Eds.). (2012). *The ethics of technological risk*. Routledge. https://doi.org/10.4324/9781849772990

Brock, G., & Miller, D. (2019). Needs in ethical and moral philosophy. *Stanford Encyclopaedia of Philosophy*. https://plato.stanford.edu/archives/sum2019/entries/needs/

Brock, G., & Reader, S. (2002). Needs-centered ethical theory. *The Journal of Value Inquiry,36*, 425.

Buijsman, S. (2022). Defining explanation and explanatory depth in XAI. *Minds and Machines,32*, 563–584.

Center for AI Safety (2023). *Statement on AI risk*. https://safe.ai/work/statement-on-ai-risk

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

Doyal, L., & Gough, I. (1986). Human needs and socialist welfare. *Praxis International,6*(1), 43–69.

Durán, J. M., Sand, M., & Jongsma, K. (2022). The ethics and epistemology of explanatory AI in medicine and healthcare. *Ethics and Information Technology, 24*(4). https://doi.org/10.1007/s10676-022-09666-7. 42, s10676-022-09666–09667.

AQDwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. W., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B.,… Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice, and policy. *International Journal of Information Management,57*, 101994.

Eisikovits, N. (2023). AI is an existential threat– just not the way you think. *The Conversation*. https://theconversation.com/ai-is-an-existential-threat-just-not-the-way-you-think-207680

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., & Keeney, R. L. (1981). *Acceptable risk*. Cambridge University Press.

Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with imagination*. The MIT Press.

Frijda, N. (1986). *The emotions*. Cambridge University Press.

Goldie, P. (2000). *The emotions: A philosophical exploration*. Oxford University Press.

Gough, I. (2003). Lists and thresholds: Comparing our theory of human need with Nussbaum's capabilities approach. WeD Working Paper 01, The Wellbeing in Developing Countries Research Group, University of Bath, Bath, UK.

Hansson, S. O. (2014). Risk. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab. https://plato.stanford.edu/archives/spr2014/entries/risk/ Stanford University.

House of Representatives (2021). Unprecedented injustice. https://www.houseofrepresentatives.nl/sites/default/files/atoms/files/verslag_pok_definitief-en-gb.docx.pdf

Johannesen, T. (2017). *The invention and conclusion of the eye*. Audio play which belongs to the artist.

Krimsky, S., & Golding, D. (Eds.). (1992). *Social theories of risk*. Praeger.

Little, M. O. (1995). Seeing and caring: The role of affect in feminist moral epistemology. *Hypatia,10*(3), 117–137.

Midgley, M. (2003). *Myths we live by*. Routledge.

Nihlén Fahlquist, J. (2020). *Moral responsibility and risk in society*. Routledge.

NOS (2023). https://nos.nl/op3/artikel/2483097-duo-negeerde-signalen-over-etnischprofileren-bij-fraudeonderzoek

Nussbaum, M. C. (1989). Mortal immortals: Lucretius on death and the voice of nature. *Philosophy and Phenomenological Research,50*(2), 303–351.

Nussbaum, M. C. (1991). The literary imagination in public life. *New Literary History,22*(4), 877. https://doi.org/10.2307/469070

Nussbaum, M. C. (1999). Women and equality: The capabilities approach. *International Labour Review, 138*(3), 227–246.

Nussbaum, M. C. (2001a). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Nussbaum, M. C. (2001b). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge University Press.

Poston, B. (2009). Maslow's hierarchy of needs. *The Surgical Technologist*, *41*(8), 347–353.

Rauschmayer, F., Omann, I., & Frühmann, J. (Eds.). (2012). *Sustainable development: Capabilities, needs, and well-being*. Routledge.

Reader, S. (2005). Aristotle on necessities and needs. *Royal Institute of Philosophy Supplements*, *57*, 113–135.

Reader, S. (2006). Does a basic needs approach need capabilities? *Journal of Political Philosophy*, *14*, 337–350.

Reader, S. (2011). Ethical necessities. *Philosophy*, *86*(4), 589–607.

Roberts, R. C. (2003). Emotions: An essay in aid of moral psychology. Cambridge University Press.

Roeser, S. (2011). Nuclear Energy, Risk, and Emotions. Philos. *Technology, 24*, 197–201. https://doi.org/10.1007/s13347-011-0027-6

Roeser, S. (2018a). *Risk, technology, and moral emotions*. Routledge.

Roeser, S. (2018b). Socially extended moral deliberation about risks: A role for emotions and Art. In J. A. Carter, A. Clark, J. Kallestrup, S. O. Palermos, & D. Pritchard (Eds.), *Socially extended epistemology* (pp. 157–172). Oxford University Press.

Roeser, S., & Pesch, U. (2016). An emotional deliberation approach to risk. *Science, Technology, & Human Values,41*(2), 274–297.

Roeser, S., & Steinert, S. (2019). Passion for the art of morally responsible technology development. *Royal Institute of Philosophy Supplements,85*, 87–109.

Roeser, S., & Todd, C. (Eds.). (2014). *Emotion and value*. OUP Oxford.

Roeser, S., Hillerbrand, R., Peterson, M., & Sandin, P. (Eds.). (2012). *Handbook of risk theory: Epistemology, decision theory, ethics, and social implications of risk*. Springer.

Roeser, S., Alfano, V., & Nevejan, C. (2018). The role of art in emotional-moral reflection on risky and controversial technologies: The case of BNCI. *Ethical Theory and Moral Practice,21*, 275–289.

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philos Technol*, *34*, 1057–1084.

Shrader-Frechette, K. S. (1991). *Risk and rationality: Philosophical foundations for populist reforms*. University of California Press.

Slovic, P. E. (2000). *The perception of risk*. Earthscan.

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life*. Hackett Publishing Company.

Steinert, S., & Roeser, S. (2020). Emotions, values and technology: Illuminating the blind spots. *Journal of Responsible Innovation,7*(3), 298–319.

Stewart, F. (1989). Basic needs strategies, human rights, and the right to development. *Human Rights Quarterly, 11*(3), 347–374. https://doi.org/10.2307/762098

Tweede Kamer (2024). Blind voor mens en recht, https://files.tweedekamer.nl/sites/default/files/2024-02/Rapport%20PEFD%20Blind%20voor%20mens%20en%20recht%2026022024.pdf

Van den Hoven, J., & Manders-Huits, N. (2017). Value sensitive design. In K. W. Miller, & M. Taddeo (Eds.), *The ethics of information technologies*. Routledge.

van Wynsberghe, A. (2021). Sustainable AI: AI *for* sustainability and the sustainability *of* AI. *AI Ethics,1*, 213–218. https://doi.org/10.1007/s43681-021-00043-6

Verbeek, P. (2005). *What things do: Philosophical reflections on technology*. Penn State University.

Wells, T. Sen's Capabilities Approach. In *Internet Encyclopedia of Philosophy*. https://iep.utm.edu/sen-cap/#SH3b