AI & SOCIETY
https://doi.org/10.1007/s00146-021-01375-x

ORIGINAL PAPER

Narrative responsibility and artificial intelligence

How AI challenges human responsibility and sense-making

Mark Coeckelbergh[1]

Received: 25 August 2021 / Accepted: 7 December 2021

Abstract
Most accounts of responsibility focus on one type of responsibility, moral responsibility, or address one particular aspect of moral responsibility such as agency. This article outlines a broader framework to think about responsibility that includes causal responsibility, relational responsibility, and what I call "narrative responsibility" as a form of "hermeneutic responsibility", connects these notions of responsibility with different kinds of knowledge, disciplines, and perspectives on human being, and shows how this framework is helpful for mapping and analysing how artificial intelligence (AI) challenges human responsibility and sense-making in various ways. Mobilizing recent hermeneutic approaches to technology, the article argues that next to, and interwoven with, other types of responsibility such as moral responsibility, we also have narrative and hermeneutic responsibility—in general and for technology. For example, it is our task as humans to make sense of, with and, if necessary, against AI. While from a posthumanist point of view, technologies also contribute to sense-making, humans are the experiencers and bearers of responsibility and always remain in charge when it comes to this hermeneutic responsibility. Facing and working with a world of data, correlations, and probabilities, we are nevertheless condemned to make sense. Moreover, this also has a normative, sometimes even political aspect: acknowledging and embracing our hermeneutic responsibility is important if we want to avoid that our stories are written elsewhere—through technology.

Keywords Responsibility · Narrative responsibility · Hermeneutic responsibility · Artificial intelligence · Hermeneutics · Philosophy of technology

1 Introduction

Most philosophical accounts of responsibility focus on moral responsibility, to the extent that both terms are often used interchangeably. This is understandable, since, as Talbert puts it, 'holding others and ourselves responsible for actions and the consequence of actions, is a fundamental and familiar part of our moral practices and our interpersonal relationships.' (Talbert 2019). This is also the case in the domain of technology. In particular, automation technologies driven by artificial intelligence (AI) and robotics pose the question who or what is responsible for the actions of these technologies, given that we may not be able to control them and predict their outcomes and consequences (Matthias 2004). For example, who is responsible when a self-driving car or the autopilot of an airplane causes an accident, and is it possible to ascribe responsibility at all in such cases?

One way to answer such questions is to draw on classic theory of responsibility. From Aristotle to contemporary analytic moral philosophy (Aristotle 1984; Fischer and Ravizza 1998; McKenna 2008; Rudy-Hiller 2018), it has been held that there are at least two types of conditions for holding someone responsible and for exercising responsibility: humans need to be in control and know what we are doing. However, these conditions are not always fulfilled when technologies such as AI take over human tasks. For example, a user of a fully automated self-driving car is not be in control of the steering of the car and may not be able to react quickly when something goes wrong. And some types of AI, in particular deep learning that uses neural nets, work in ways that are not transparent, creating ignorance on the part of the user. Does this mean no human can and

✉ Mark Coeckelbergh
mark.coeckelbergh@univie.ac.at

[1] Department of Philosophy, University of Vienna, Universitätsstrasse 7 (NIG), 1180 Vienna, Austria

Published online: 30 December 2021

should be held responsible for the actions and consequences of these technologies? These are important questions, which are being discussed in the literature on ethics of robotics and AI (Hakli 2019; Johnson 2014; Santoro et al. 2008; Coeckelbergh 2020; Yeung 2018; Santoni de Sio and Mecacci 2021). These discussions already show how AI, here in the form of an automation technology, poses a challenge to our humanistic notions and to the kind of control and knowledge conditions connected with them. While there is a lot of variation in historical humanism, it has always put humans in the centre, and since the Enlightenment humanist moral philosophy has stressed human autonomy and agency. AI threatens such views of human being, morality, and responsibility.

Yet, there are also other notions of responsibility: concepts that are usually not included in discussions about AI and responsibility, but are equally important if we want to understand human responsibility in its full richness and how AI challenges our existing ways of thinking and doing. In this paper, I identity three further notions: causal responsibility, relational responsibility, and what I call "hermeneutic responsibility". Next to making this distinction, my further aim is to show (1) that in spite of connections between them (moral responsibility is related to causal and relational responsibility), each of these notions are connected to different kinds of knowledge and perspectives on human being, which are often in tension, and (2) how each of them are challenged by AI, in particular machine learning AI. Special attention will be paid to a specific form of hermeneutic responsibility, what I call "narrative responsibility": I will explain why we need such a notion and what it means to exercise it, how it differs from moral responsibility, and how it links to hermeneutic approaches to human being and technology. With a nod to tensions between humanism and posthumanism (and, to some extent, transhumanism), I will argue that AI and meaning are already entangled, since there are ways in which AI "participates" in meaning-making (thus supporting criticisms of humanist approaches), but insist that whatever may be said about other notions of responsibility, it is always up to humans to make sense of AI, with AI, and, if necessary, against AI. In this way, I aim to make an original contribution to thinking about *responsibility* (in general and especially in the context of thinking about AI) and respond to, and further develop, recent literature on *technology and hermeneutics* (Romele 2020; Reijers and Coeckelbergh 2020; Kudina 2021), which has proposed revisions of, or alternatives to, existing postphenomenological accounts of technology (Ihde 1990; 1998; Rosenberger and Verbeek 2015).

Note that while most of the effort in this paper goes in analytically distinguishing the different notions of responsibility—I aim to establish hermeneutic responsibility and narrative responsibility as distinct concepts—I will also acknowledge that at least some, if not all notions are inextricably interwoven. For example, I will note that moral responsibility is linked to moral responsibility and argue that hermeneutic responsibility and moral responsibility need each another.

Let me start with causal responsibility and moral responsibility.

2 Causal responsibility and moral responsibility in trouble: the battle for the mind

Causal responsibility of agents refers to agents being the cause of an outcome. While many moral philosophers hold that moral responsibility requires, or is grounded in, causal responsibility (Sartorio 2007; see also again the control condition), causal responsibility does not necessarily entail moral responsibility. For example, a young child may cause harm to someone, and is causally responsible for that harm, but typically we do not hold that child morally responsible. AI is also a case in point, at least if we assume that it cannot be morally responsible: if AI takes the form of an artificial agent (e.g., an autopilot), then that artificial agent may cause a particular outcome, but we do not hold that agent morally responsible; instead, we look for a human to bear the responsibility for what the agent does, has done, or might do. Moreover, in technological action, causal responsibility is usually a matter of degree and involves many hands (van de Poel et al. 2015): an outcome (e.g., a recommendation or decision by the AI) is often not directly caused by one agent, but may be the result of a long causal chain and the causal responsibility of a particular agent (human or artificial) depends on the extent and directness of the agent's contribution to the causal chain. For example, the outcome of what an AI system does (a recommendation, an action) may the result of several programmers and data scientists doing part of the work, and those who did more work and directly influenced the outcome will carry more causal responsibility. However, it is not clear how these degrees of causal responsibility translate into moral responsibility. Our current moral and legal ways of thinking do not seem very well adapted to dealing with causal chains that are temporally stretched, vary in degree, and involve many agents.

Causation is itself a long-standing topic in philosophy, and in discussions about moral responsibility, it is connected to debates about free will and determinism (van Inwagen 1983; Fischer and Ravizza 1998; Frankfurt 1969; Pereboom 2001; Dennett 1984). In general, the tension between moral responsibility and causal responsibility is related to two different views of human beings. One, usually defended in moral philosophy, is that of human beings as rational and free beings who wish to preserve their autonomy and wish to be in control of their actions.

As Berlin (1997) puts it in his famous paper on liberty: there is 'the wish on the part of the individual to be his own master. I wish my life and decisions to depend on myself, not on external forces of whatever kind….I wish, above all, to be conscious of myself as a thinking, willing, active being, bearing responsibility for my choices and able to explain them by reference to my own ideas and purposes.' (Berlin 1997, p. 203) Another is the scientific view, present in positive psychology, neuroscience, cognitive science, and so on, that *explains* human actions and that shows opportunities for the manipulation of human choices and behaviour. Consider nudging for example, which subconsciously aims to influence our choices by changing our decision environment, the so-called 'choice architecture' (Thaler and Sunstein 2008). This goes against the view of humans as autonomous choosers and reasoners.

AI technology seems to be situated firmly on the side of the scientific view: it does not support the view that humans are autonomous reasoners and instead categorizes, profiles, and enables manipulation. The AI we are usually talking about today is machine learning that relies on statistics. Humans are analysed in terms of their data. From the epistemic gateway offered by AI, they are not seen as human beings that want to be autonomous and masters of their lives. Current AI does not care about your motives, your reasoning, and your plans. It will categorize you statistically, compare you with others, and make predictions. It is not about you as a person, and not even about you as a rational agent—the favourite of moral philosophers. Reasoning is no longer required when we have data analysis. It is also not necessary to introspect. As Harari (2015) suggested: AI knows you better than yourself. This kind of claim was already made by positive (behavioural) psychology; now, AI joins this project of rendering the self entirely transparent and knowable. Ethically dubious experiments with humans and animals—from the infamous Milgram's (1963) experiment and the Stanford Prison Experiment (Heney et al. 1973) to experiments with apes with brain chips implanted today (e.g., Serruya et al. 2002 and Elon Musk's recent Neuralink experiments)—are replaced by, and supplemented with, computer models and datasets. AI is the new technology to "read" humans. Behaviour or minds are no longer of direct interest; what matters is the data produced by that behaviour and those minds. The precise methods and goals of such research may differ considerably. However, the resulting claim about knowledge and self-knowledge is the same. It seems that there is no longer any need for humanistic reading and writing, meant since at least the Renaissance as a tool that enables us to attain self-knowledge. According to adherents of these positivist views, AI can do that job based on data about us. And from this perspective, reasoning about moral responsibility seems at best an epiphenomenon when we have data about how humans actually make moral choices and how they behave.

However, due to the type of knowledge it produces and relies on, AI does not offer causal explanations and there is no assumption of determinism. AI makes predictions, but it gives us correlations and *probabilities.* In this sense, AI is a threat to both classic notions of causal responsibility and moral responsibility. It is a threat to moral responsibility, because by enabling statistical knowledge, manipulation, and automation, it seems to undermine the agency, autonomy, and responsibility of humans. However, it is also a threat to classic causal responsibility. Causes are something humans think of, for example when they make a causal model and construct a theory. Causes can be doubted, as many people did and do since Hume. AI, by contrast, only works with correlations and probabilities. These are not a matter of beliefs (or so it seems); they are calculated. AI is not about (old-style) physics but statistics. It does not need theory; it only needs data—your data, the data of millions of other people.

In this way, AI seems to undermine both causal and moral responsibility, and the respective views of human being connected to them. Philosophers continue to talk about agents, reasons, and causes. However, both physics and the human sciences have moved on. What (literally) counts now in the age of AI is data extracted from our behaviour and their analysis in terms of correlations and probabilities. This adds AI to the history of disenchantments and disappointments that humanists had to cope with since Darwin and Freud. The "human" of the Renaissance humanists, Enlightenment thinkers, and nineteenth century romantics, with its free will, rational autonomy, and mysterious mind, seems to be an illusion. AI seems to set us on a path towards the 'Palace of Crystal' sketched by Dostoyevsky (1972) in *Notes from Underground*: one in which science will teach us that we 'are no more than a sort of piano keyboard or barrel-organ cylinder,' a world in which everything has been mathematically worked out and where there is no room for fancy, 'individual deeds or adventures' (pp. 32–33). A world in which humans and their minds become fully transparent. Dostoyevsky is still struggling with determinism. However, the kind of tension is the same. Humanist philosophers and writers defend the human, and some of them may want to 'send all these logarithms to the devil and be able to live our own lives at our own sweet will' (p. 33), as Dostoyevsky put it. However, AI, neuroscience, and behavioural psychology and economics are here to stay, and can easily be used for the manipulation of people and the destruction of the autonomy and morality cherished by those who, at least from the perspective of these positivist sciences, might be seen as old-style philosophers and psychologists.

3 From minds to social relations: relational responsibility to others

However, those engaged in this *battle for the mind* tend to neglect another kind of responsibility—or, as we will see, another aspect of moral responsibility—and a different way of looking at human beings: we are also social beings, and as such, we have a relational kind of responsibility. We do not only have responsibility as an agent, a causal and moral responsibility *for* our actions, but also a responsibility *to* others. Whereas most accounts of moral responsibility relate the agent to a moral demand, relational responsibility highlights the relation to others, to 'responsibility patients' (Coeckelbergh 2020).

While we can analytically distinguish moral and relational responsibility in this way, one could argue that moral responsibility always involves relations with others, and that a richer and more plausible notion of moral responsibility includes relational responsibility: we are responsible *for* our actions *to* others. In this sense, relational responsibility can be seen as an aspect of moral responsibility. Nevertheless, this aspect is often silent and silenced in the above-mentioned discussions on moral responsibility, and many authors have found it necessary to develop new or alternative accounts of responsibility in response to this gap, thereby sometimes radically changing the account of moral responsibility (see below). Therefore, I have chosen to give this this aspect by giving it a separate name: relational responsibility.

Both the link between moral and relational responsibility and the potential for seeing this as an entirely different view become clearer when we look at some sources we may use to develop this conception of relational responsibility. One is responsibility as answerability, which has been proposed in a criminal legal philosophy context (Duff 2005). Duff has proposed an account of criminal responsibility according to which 'to be responsible is to be held responsible for something by some person or body within a social practice.' (Duff 2005, p. 441). For example, in a trial, a defendant has to answer a charge of wrongdoing. Duff connects this with reasons: the defendant has to have the capacity to engage with reasons for action (p. 446). Now, one could generalize this notion of responsibility to a richer, relational view of moral responsibility (a move in line with Duff's work), or supplement moral responsibility strictly speaking with another type of responsibility, relational responsibility, which gives us a responsibility *in addition* to moral responsibility: we do not only have responsibility as agents but also as social beings, social actors, who in our specific roles and social contexts have to answer to others for what we do (to them). In both options, we put responsibility in a social context, without necessarily losing the link to moral responsibility. In the remainder of the paper, I will assume that the first option holds: moral responsibility implies relational responsibility, relational responsibility is an aspect of moral responsibility, and moral responsibility must be interpreted in a relational way.

Yet, recognizing this relation between moral and relational responsibility should not hide the potential radicality of a shift towards a more relational view of moral responsibility. Another range of sources for thinking in a relational way about moral responsibility can be found in theoretical directions that question the individualism and focus on the self that is inherent in much modern normative theory, and that are more relational and *other*-directed such as ethics of care and Levinasian ethics. Here, the emphasis is not on the agent's will, control, and autonomy, but on the other and on what the other may need, ask, or demand. Here, too, moral responsibility is interpreted in a relational way. For example, Gilligan's ethics of care connects responsibility to human relationships and stresses being responsive to people and care about them instead of focusing on one's autonomy (Gilligan 1982, p. xiii); in nursing ethics, it has been argued that health care professionals have relational responsibilities towards their patients, which depend on professional roles and may be very particular (Nortvedt et al. 2011); and earlier, Levinas (1969) has proposed an ethics that starts from (the face of) the other: not the self but the other, and the ethical relationship constituted by the other, is primary. Once again, we arrive at more relational views of moral responsibility. However, in the case of Levinas, this implies a radical shift from a self-oriented to an other-directed account of moral responsibility. In that case, it becomes more difficult to see relational responsibility as merely an aspect of moral responsibility: moral responsibility implies relational responsibility, but that changes the entire picture. Levinas radically revises the usual accounts of moral responsibility.

Furthermore, beyond philosophy, social-scientific approaches—also to science and technology—question mainstream moral philosophy's obsession with the mind, its psychologism. They point to the social context of responsibility and to the power structures at play: Who asks this question of responsibility, who is supposed to be responsible, who is included and excluded in this game of responsibility? Seeing people as autonomous and individual agents is itself a cultural construction, and in particular a Western obsession. While social constructivism does not necessarily deny (the importance of) moral agency, it criticizes the tendency to understand agents in isolation from their social contexts and the claim to universality made by standard accounts of agency. And important for the topic of this paper: social studies of science and technology show that neither science nor technology is politically, morally, or culturally neutral. This insight has been taken up in philosophy

of technology. For example, in dialogue with STS, Winner (1980) has famously argued that technical things 'have politics.' And inspired by a by now decades long tradition initiated by Bijker and colleagues (1987), Johnson (2014) has argued that responsibility arrangements regarding robots will have to be negotiated between actors as the technology is developed, tested, and used. Some actors will push others in their direction; they get what they want, rather than others. Here, we move from physics (causes and determinism), individualist moral theory (will, minds, autonomous selves, reasons, etc.), and statistics (correlations and probabilities) to the social, cultural, and political sciences. This is about social actors, relationships, roles, groups, and power.

With regard to AI, this relational approach means asking *to whom* we are responsible if we develop and use AI (not just asking who is responsible for what): what is the context of social relationships, and what responsibilities does who have towards whom? This can be further unpacked in number of ways. First, one problem with AI is that it is often divorced from such a social context and ecology of social relations; it is seen as a purely technical matter. And philosophical discussions in terms of agency do the same: they highlight the moral responsibility of agents (human and artificial) without asking the "to whom" question, thus leaving out an important if not essential dimension of the ethical relationship. Second, a relational approach to AI also means evaluating again the knowledge provided by AI. And here we encounter the next challenge: if AI gives us a recommendation and we make a decision based on this recommendation, but the AI process does not enable us to explain to those affected by the decision why a particular decision was made in their case, then we cannot fulfil our relational responsibility (Coeckelbergh 2020). Third, relational responsibility could also mean responsible innovation, which means, among other things, that stakeholders are involved in the development and decisions about the use of AI. The idea is to have a transparent process by which societal actors become mutually responsive to each other (von Schomberg 2011). This normative view of innovation contrasts sharply with the fact that much innovation in this area is done within companies who develop their technologies more or less in isolation from the rest of society, let alone that decisions about their design or development are democratic. Fourth, this more social and political perspective on responsibility enables us to open up a pandora's box of political interests and power relations that surround and interact with responsibility. For example, if moral philosophy asks of individuals to act responsibly, but these individuals find themselves in contexts that do not enable them to exercise this responsibility, because they are over-powered by their company, their government, and so on, then all the theories about causal and moral responsibility seem less relevant—at least in the first instance. Then, at the very least analysis of these power structures is also needed—for instance inspired by Marx or Foucault.

Since AI is usually seen as a technical and scientific matter, discourses on AI tend to obscure these social and political relationships and therefore render it difficult to talk about responsibility with regard to AI in a relational way. There is a gap between, on one hand, the usual discourse about AI and responsibility in technical literature and in moral philosophy, and on the other hand, the political issues raised by AI, which remain unaddressed or at least underdeveloped. This gap can be closed in at least two kind of ways. First, we can use social science and political philosophy to talk about AI. Currently, awareness about political issues concerning AI is growing (Véliz 2020; Bartoletti 2020; Crawford 2021), but often this is not matched by in-dept academic analysis using political theories developed within the social science and humanities. Second, even considered at the technological level and within the range of the existing discourse, there is potential for highlighting social and political issues, since AI can show much about our social world, sometimes things we are not aware of. For example, by means of purely quantitative, statistical analysis, AI can reveal that there are existing and historical biases in our language, our texts (Caliskan et al. 2017), our organization, and our society. Bias in AI may well be an ethical problem, but AI also contributes to more knowledge about our society *by revealing this bias in the first place* and by thus "inviting" us to talk about it. AI ethics, together with other developments and in specific contexts (e.g., the Black Lives Matter movement in the U.S.), has succeeded in putting bias high on the political agenda.

Yet, the tension between humanistic and technical approaches remains, especially if we consider again the knowledge provided by AI that is abstracted from human contexts. The knowledge here is not gained by sociologists and intellectuals that come up with big theories and heavy volumes of analysis; it is offered by AI and is, again, of a specific quantified kind that does not stand in need of theory. As a technology that, like all technologies, is more than an instrument, AI "suggests" that the correlations and probabilities it gives us are enough. In this way, AI kind of by-passes not only human agency (moral responsibility) and human reasoning about causes (causal responsibility); it also circumvents at least part of human social analysis. Similar to developments in psychology, that analysis was already getting increasingly quantitative and statistical. But now, the machine also takes over, or rather seems to render superfluous, the only part that was left for the human social scientist: theory.

From a humanist point of view that puts humans in the centre, things start looking rather bleak now. Is there still a place for humans and what (only) humans can do? This way of putting it is too one-sided, as if it is a matter of *either* human responsibility *or* AI taking over. The

picture must be revised in at least two ways. First, from a humanist point of view, we can insist on the human role and we can demand that humans still decide, in the face of probabilities, demand that, even if agency is delegated, we remain morally responsible, etc. In other words, here, we accept the picture or narrative of the battle between humans and technology, and—unsurprisingly—choose the side of the humans. Second, however, we can develop theories that criticize this humans/technology binary and bring together humans and technologies, while not losing the capacity to criticize technology. We can emphasize the human side of technology and the technological side of humans. Postphenomenology, posthumanist theory, the work of Latour, and indeed, the STS already mentioned may be considered to develop this. In this section, I will not say more about this, but in the next section, I will further discuss this issue and try to find a middle way between these extreme positions. This will come in two versions. The humanist version insists on the hermeneutic centrality of the human: the only form of anthropocentrism that is still viable is one that gives a special place to humans as interpreters. While the assertion of human exceptionality becomes increasingly challenging in the light of posthumanist, environmentalist, and postphenomenological insights, one could argue that there is one thing that is and should be the *responsibility* of humans: to make sense of the world. In the posthumanist version, this sense-making is more intimately connected to technology. It could be argued that we have to embrace the entanglement of sense-making and technology: technology participates in meaning-making and contributes to hermeneutic responsibility. Nevertheless, I will argue that even in this posthumanist version, humans carry and should carry the (end-)responsibility.

Let me unpack this and say more about hermeneutic responsibility, especially about what I call "narrative responsibility."

4 Narrative responsibility as a form of hermeneutic responsibility: the responsibility to make sense

We have (moral) responsibility for what we (causally) do and we have responsibility to others, to those to whom we are related. However, we also have a responsibility that is usually not mentioned in discussions about the topic: the responsibility to make sense, to interpret, and to narrate. Whereas relational responsibility is a second-person kind of responsibility, directed to others, and whereas causal and often moral responsibility is often formulated from a third-person point of view (as ethicists, we look at the whole from an outside perspective and then ascribe causal and moral responsibility), what I propose to call "hermeneutic responsibility" is mainly a kind of first-person responsibility: a responsibility that we have mainly to ourselves as persons (first-person singular) and to "us" as communities, societies, and cultures (first-person plural). It is not a moral responsibility strictly speaking, if that means that making sense is a moral "ought". It is not so much something that we can or should be blamed for or that can be demanded. It is rather a responsibility that emerges from my and our existential situation as humans. It is a "have to" or "cannot do but", not an "ought to". And I have to do it, as the person I am. Or we have to do it, as the community and culture we are. It is not about a responsibility that we have as universal moral subjects—although moral responsibility can also sometimes be about problems for particular people, as Sparrow (2021), inspired by Gaita, recently reminded the robot ethics community. It is a responsibility we have as the particular persons we are living in a particular community, society, and culture. Hermeneutic responsibility also has nothing to do with causes and explanations, or with general laws of psychological and social behaviour. It is usually applied to the human and social world, but it is about interpretation and *verstehen*: a term meaning "understanding" that was already used by Weber and Simmel against sociological positivism. It is relational, in the sense that we have to communicate the sense to others, but it is not only or primarily others that we have to answer: we have to answer ourselves. We have to provide answers to what happens to us and to others, and, ultimately, we have to answer the question mark that we ourselves are as human beings and as persons.

This "hermeneutic responsibility" typically takes the form of what we may call "narrative responsibility." Our sense-making and answer to what we are usually comes in the form of a narrative: a story about ourselves, about others, about events, and about how we respond to those events. Here, the disciplinary field is not moral philosophy, psychology, or sociology; we move to literature, music, art, film, games, etc. With regard to exercising this kind of responsibility, reading or writing a novel is not superfluous or simply a matter of entertainment; it is part of the hermeneutic work. The humanistic culture we inherited still uses the technology of writing and the medium of text and books. However, I mentioned games, because in principle, we could also use new, digital technologies to do this work, to exercise our narrative responsibility. All kinds of media and cultural practices can be used and developed—indeed have to be developed—since it is and remains our responsibility.

If this kind of responsibility sounds abstract and not much related to what we usually mean by "responsibility", consider examples drawn from the discussion about moral responsibility: a car crash or airplane accident. If we approach these from a *moral* point of view, I (from a third-person perspective and backward-looking perspective) try to find the responsible agent or agents who caused the accident

directly. For example, someone investigating a crash will face this task. From a first-person point of view, I may also explain to others why I have done what I did—thus fulfilling my *relational* responsibility. This could be the responsibility of the driver or pilot. Forward-looking and from a first-person singular perspective, I will try to act as a responsible agent and try to avoid accidents by fulfilling the conditions of responsibility: I will need to make sure that I am in control and know what I am doing, e.g., I make sure that I am not drunk—to pick up an Aristotelian example. For example, a car driver will make sure she is not drunk (next time). In the plural: we need to create responsible technologies and societies by developing technologies and building structures and infrastructures that make fulfilling these conditions easier. All this may involve trying to gain knowledge in terms of causes and probabilities. To be responsible *for something* and explain *to others*, we need to know what happened, how things work, who did what, and who needs to do something. We need to sort out the *causal* responsibility and, with the current science and technology including AI, we need to know about probabilities and risk. For example, exercising forward-looking moral and causal responsibility in the case of airplanes means that a lot of knowledge about risk and probabilities needs to be acquired and produced.

If, however, we look at the same kind of cases from a *hermeneutic* and *narrative* angle, the main issue is not about agency and not even immediately about responding to others (responsibility patients). From a backward-looking point of view, this means: something happened or might happen and we are faced with the task to make sense of what happened or might happen. For example, an airplane crashed and more than 300 people died. Or a particular airplane has a high risk of failure, but is still widely used. In such cases, we need to sort out the other kinds of responsibility (including moral responsibility and legal responsibility), but we also need to *make sense* of this and cope with this as interpretative human beings, human beings who are mortal and fear death, love their relatives, and so on. People involved and other stakeholders, journalists, readers, and so on, do this by creating a story about the accident or the risk. Facts, causes, probabilities, reasons, obligations, and so on may be part of that story, but they do not make a full story. It needs to be a story that makes sense and that helps us to make sense. For the sake of ourselves and others, we need to make sense of what happened and—ideally—in the process make sense of ourselves as persons and as humans. Perhaps afterwards, we will see the world in a new light. And from a forward-looking point of view, we need to make sense of what might happen. As humans, we are always directed towards the future. And that future is uncertain and risky by definition. The knowledge needed here—if it can still be called knowledge at all—is of an entirely different kind than the knowledge offered by the science or the reasons and discussions offered by moral philosophy. If we only have science or moral philosophy, we face a hermeneutic gap: we know already many things about what happened (e.g., causes or correlations) or what might happen (e.g., calculated risk, probabilities), and we have reasoned about those and we have explanations, but we still need to make sense of it all, and we need to make that sense for ourselves and for the people around us and the community and perhaps the society and culture we belong to.

For example, when someone near to us dies in an accident or when we suddenly and unexpectedly become seriously ill, we want to know what exactly happened and, in some cases, we will want to talk about the causal and moral responsibility (e.g., the other driver was drunk or I may have contributed to my bad health). Usually, we will get medical information, for example, say the probability a family member has to survive at a given time and in a given condition. We may want to have the data available (and a medical professional's interpretation, which is already hermeneutic). However, if necessary at all, that knowledge is not *sufficient* for making sense. That sense *may* come with making, telling, or hearing the story and after the story. Narratives can help, albeit without guarantee. Narratives are hermeneutic tools that help us to make sense. And they do not just consist of numbers and statistics; they are about *personas* and events, about personal experiences, personal transformation, personal relationships, meaning, and existence. My personal sense-making will also relate to the meanings and practices of sense-making that are already given in my community, society, and culture. To make sense can be a very personal matter (for example making sense of an accident in which a loved one was involved); but the way I do it will link to a wider whole, what I will below refer to as a 'form of life.' For example, someone might refer to religious meanings and other meanings available in one's family, community, and culture.

In philosophy, one of the key figures of a hermeneutic approach to human being is Ricoeur, who argued that human lives and human experience are not only fundamentally social but also have a narrative character. He also theorized narrativity. Based on his reading of Aristotle's *Poetics*, Ricoeur offered in *Time and Narrative* a theory of 'emplotment' and *mimesis*: he argued that the plot of a story configures and organizes characters, motivations, and events in a meaningful whole. There is action and there are events, but in the end, the narrative as a whole makes sense and leads to a new understanding (Ricoeur 1983). This can be read as a theory about fictional stories (Aristotle wrote about theatre and in his thinking, renewal comes in the form of the famous *catharsis*), but it can also be used as a theory about how we make sense in our lives: narrativity, in particular emplotment, leads us see things in a new light, helps us to make sense of things. The knowledge, or rather know-how, that

is thus needed for exercising hermeneutic responsibility is narrative. By organizing events, characters, motivations, and other elements in a meaningful whole, we can make sense.

Now, this theory seems to work well for backward-looking responsibility, where we already have the different elements of our story and where part of the story may already be available in narrative form (consider for example a newspaper story about the airplane accident). However, what about forward-looking narrative responsibility? It seems that we then need to use our imagination. We can create narratives about possible futures. This is what with Ricoeur we could call a form of 'productive' imagination that does not just copy but shows us new possibilities. We can imagine various scenarios and, in that way, we can try to make sense of what we are doing now and what might happen in the future. Narration does not mean that we can only create *one* story, even if eventually we might have to choose one; there is an inherent plurality and openness in this stage of the responsibility exercise.

With regard to this forward-looking, imaginative exercise of responsibility, it is also important to distinguish its narrative character from that of other forms of responsibility. Even when faced with a moral choice, it is not enough to discuss this in terms of my moral responsibilities (e.g., my obligations towards others, the reasons I have as a moral and rational being); the choice also *has to make sense to me*. Using narrative imagination, I must explore and create these meanings. This is my (or our) hermeneutic responsibility: no-one else can do it in my place and no community or society can do it in our place. I have to make sense given my own personal history and given the person that (narratively) I am. The same is true for communities or societies. Consider societies that struggle with, say, their colonial past. This is not only a moral question, as it is usually understood. There is a moral aspect, for sure. For example, that society with a colonial past may well have the obligation to apologize, to make sure it never happens again, and to be particularly sensitive to new instances of racism, imperialism, bias, and so on in the present. However, dealing with such a past is also hermeneutic home work. That society and the previously or presently involved and affected groups—agents and patients, doers and victims—have the hermeneutic responsibility to deal with their past and to find and make meaning today and for the future in the light of what happened. This requires this time not just a forward-looking but also a backward-looking imagination. It requires linking the past to the present. This can be done in the form of narrative: stories concerning the past need to be told, perhaps revised, and made to bear on the present.

Furthermore, the purpose of this hermeneutic work is not only to make sense of the present but also to shape the future. The future needs to be approached narratively to imagine new possibilities to organize people and events—in the example, this could be: a non-colonial form of organization. In that sense, hermeneutic work has a normative dimension: not necessarily or at least not *just* moral in the sense of obligations and blame, but still related to normative ideas about how we should do things, how we should lead our lives, and how we should live together.

To distinguish hermeneutic and narrative responsibility from moral responsibility does not imply that both are unrelated. On the contrary, to think about what is right and about what the good life is without involving the question regarding meaning seems problematic. As mentioned, a moral solution should also make sense to me and to others. And vice versa: to shape the narrative of our lives without taking into account moral responsibility and other kinds of responsibility would not be desirable and not good—if it is possible at all. The relations between the different kinds of responsibility and their respective domains of life and thinking are complex, and a full discussion of this issue is beyond the scope of this paper. More work is needed on this. For now, let me conclude that clearly all notions or aspects of responsibility are important, and that moral responsibility and narrative responsibility are interwoven in the sense that both seem to need one another.

5 Narrative responsibility and AI

What does this concept of narrative responsibility mean for AI? At first sight, AI has little to do with all this meaning-making. One could argue that AI is not conscious and not self-conscious, and that it therefore lacks subjectivity and experience, which is assumed to be needed for meaning-making. Whether or not AI may achieve consciousness in the future, AI as we know it lacks consciousness; it does not experience, let alone tell stories about that experience. At first sight, therefore, meaning-making is not a case for science or technology at all. We better call in the poets and the writers. It is a humanistic, not a scientific project. And that is partly true. Like in the case of moral and relational responsibility, there are and remain fundamental tensions with regard to the kind of knowledge and responsibility required. And it is all too easy to see this in a binary way. To make sense by means of narrativity is not about probabilities but about meaning. It is not about data and correlations but about *emplotment* of persons/characters and events. It is not about gathering data and analysing data; it is about creating, reading, and interpreting texts. Once again AI seems to totally circumvent human knowledge, experience, and imagination. It offers statistical analysis and probabilities, whereas humans need to make sense. Once more the humanist project finds itself in tension with, if not in radical opposition to, science and technology.

Yet, this picture is again too one-sided and distorted. AI, like other technologies, has a lot to do with human culture and human meaning-making. First, within academia, there are new and interesting interactions between AI and the humanities. Consider the field of digital humanities, which is situated at the intersection of computing science and humanities disciplines such as history and linguistics. It uses digital tools such as data mining to study the humanities. This does not mean that other, classic humanities methods are abandoned, but they are combined with the digital methods. As already suggested, we may think about how to use new technologies and media for doing humanistic work—as long as we do not forget that interpretation and narration by humans is always required. Second, there is a much more *internal* relation between technologies and hermeneutics: technologies are part of our stories and even shape these stories. Recently, this has been conceptualized in at least three ways:

First, I have mobilized Wittgenstein's concepts (games and forms of life) and approach (meaning is in use), used in the *Philosophical Investigations* (2009), to argue that not only meaning in *language* depends on use and context, but that also the meaning of *technologies* depends on their use and is related to our activities, games, and form of life: technologies thus contribute to culture and are at the same time shaped by it (Coeckelbergh 2018). For AI, this means for example that the biases it may (re-)produce are often related to the biases that are present in the language and other games that are played in our societies. Another metaphor is grammar: there is already bias in the grammar of our society: in the way we speak about one another; in the way we treat one another. These meanings are then reproduced and performed in and through AI—without AI itself having consciousness, experience, or subjectivity, and with humans involved as necessary co-makers and interpreters of the meaning. For example, if there are forms of gender bias in a particular society, then AI that is developed and used in such a society is, through its use, likely to contribute to this game or form of life, together with humans. Changing the game might well be possible, but is a long and difficult process. Ethics of AI would then have to understand itself as a game changer by producing meanings that differ from those enacted in our present games and form(s) of life. For example, it may try to shape the development of AI in a way that does not exacerbate, or even avoids supporting at all, binary ways of thinking about gender in society.

Second, drawing on the work of Ricoeur and Gadamer and responding to postphenomenology's claim that technologies mediate human–world relations (Ihde 1990; Rosenberger and Verbeek 2015), several authors in philosophy of technology have argued for a hermeneutic approach to digital technologies. Ricoeur argued that human experience is mediated by language and narrative; this has now been expanded to technologies. Connecting technologies to meaning-making, it has recently been argued that digital technologies mediate and modify our world (Romele 2020), co-configure our narratives (Reijers and Coeckelbergh 2020), and mediate our sense-making: people try to comprehend new technologies and fit them in their daily practices (Kudina 2021). Seeing digital technologies as having nothing to do with human culture and meaning-making and creating a strong opposition between them, is then itself one (problematic) way of coping with, and making meaning of, these technologies. With regard to AI, one could argue that AI is integrated in our lifeworld and participates in our sense-making as it shapes our experience and actions and configures our narratives. For example, if AI monitors my health (through all kinds of apps and devices), then this co-writes the narrative of my day (e.g., get up and go running, don't eat between this time and that time, doing a particular kind of exercises and yoga, etc.) but will also influence and shape the sense I make of myself: the stories I tell to others (for example on social media) but also the story that I tell to myself: the kind of person that I am and become, and the sense I make of my life. In this sense, AI becomes co-narrator of my stories. Again, no conscious AI is needed for these mediations and participations in meaning-making. It suffices that humans have consciousness and subjectivity. AI can only participate in, and co-shape, sense-making and narrative processes through humans.

Third, both environmental philosophy and posthumanist theory have questioned anthropocentric views. For example, Braidotti (2017) has explored post-anthropocentric directions in the form of posthuman critical theory and Puig de la Bellacasa (2017) has argued that care is not just a 'human-only matter' (2). With regard to meaning-making, such views at least invite us to consider the idea that meaning-making is something in which non-humans can also participate. McCormack (2018) has argued that anthropocentric accounts of meaning-making are untenable if we situate human meaning-making in an ecological context and understand it in a non-binary way. While it is not clear what this means for machines, it is worth considering meaning as a process in which also non-humans participate. Even if these non-humans are not conscious and hence do not have experience, the view that meaning-making is exclusively human seems at least problematic if we consider that that human is always related to its environment, and that meaning-making therefore is also always relational. For example, making meaning of our society today requires that I somehow also make sense of AI, since AI is now part of the meaning-full world of my society. In this sense, AI "participates" in our collective meaning-making. And, when AI writes a text (consider the language model GPT-3 that uses deep learning to produce text), then one could argue that this contributes to meaning-making, even if only humans can

complete and lead the meaning-making process, since they have consciousness and experience.

Thus, these directions in hermeneutics imply that AI is not just the object of our stories, but also contributes in important ways to these stories and to the meaning that arises in the process. This enables us to revise the picture and narrative that emerged so far when we considered how AI challenges our notions of responsibility: the humanist narrative that responsibility in all its forms involves a kind of battle between humans and machine unnecessarily exaggerates and misconceives the tensions between, on one hand, human pursuits such as morality and sense-making, and technology. Technology is itself human-made and human-used, and is entangled in various ways with human beings and what they do. This includes morality and making sense. AI can contribute to exercising our moral, relational, and hermeneutic responsibility, for example by making us aware of existing bias in society or by becoming integrated in our daily lives. And even if humanists write against AI, for example in an ultimate humanist effort to win the battle "against the machine," AI still shapes their thinking and sense-making, it is still "with" them—albeit as an opponent or even enemy. More generally, AI is part of our narratives: personal narratives and larger, cultural narratives.

Another example of such larger narratives is the transhumanist narrative of increasing intelligence. It tells a particular history of technological progress, which is a history of humans, a history of increasing computer power, and a history of AI (as a kind of hero) doing things and AI events, e.g., winning the game Go, writing texts, interpreting brain waves, etc., and which seems to be "driven" by AI. In this sense, AI is also a "character" (a hero or helper) and even co-creates the narrative. At the personal level, AI configures our lives and gets integrated in our lifeworld as we use various AI-powered technologies. For example, as our calendars and phones get increasingly "smart", they organize the stories of our daily lives. And perhaps, AI will soon *literally* write many of our texts, or at least co-author them. The relation between technology and meaning is thus far more complex than presented in standard humanist accounts that defend meaning and humanity against the invasion and taking over of technology. As far as the creation of narratives goes, meaning-making can be a shared or distributed activity between human and non-human "authors" and "readers". Even if AI, as a non-human author and reader, is not conscious and is a different kind of author and reader than humans, since it uses and produces a different kind of knowledge (see also below), it still co-shapes meaning and narrative. AI is not just an element of our story; it also co-creates that story.

That being said, even if we accept that AI and other technologies contribute to meaning-making in the ways described, we may still want to insist that the *responsibility* for this meaning-making remains with the human. AI and other machines cannot themselves have or take *responsibility* for making meaning (since, so I assume, they cannot take responsibility for anything given that they lack consciousness and subjectivity) and therefore cannot take the hermeneutic lead, so to speak. Humans have to take that lead: they carry the hermeneutic responsibility for making sense of themselves and their social and cultural world (which includes technology). AI is part of our narratives and helps to shape them, but it is our responsibility to define their role as co-creators and it is up to us what place and role we give them in the narratives that we co-create. And in the end, it is our responsibility to decide what narratives we want to (co-)write—including narratives about AI, with AI, and sometimes against AI.

Asking this question about which narrative we want is important, since there is a normative dimension to this responsibility. The precise ways in which AI shapes us and our stories may be very problematic. Consider for instance 'quantified self': this is not only a specific phenomenon of technology use (self-tracking using digital technologies and data); the term can also be used to point to a more-than-instrumental effect of AI and data science that has to do with meaning and with stories: quantification of the self in the sense that the self comes to be experienced and shaped in terms of data, numbers, and statistics, and that the story of our selves becomes one about data, numbers, and statistics. Moreover, as I will stress below, when we use such technologies, there is also the danger that we live a narrative that is written by someone else (programmers, designers, corporations, governments)—through technology. For example, a health app may try to shape how I live my life. Acknowledging that AI co-shapes ourselves and co-writes our narratives does not mean that we must uncritically accept the specific self-formation and story. On the contrary, once we become aware of the hermeneutic and narrative role of AI and other digital technologies, we can evaluate what happens and try to re-shape ourselves and re-write the story. Without knowing and acknowledging AI's hermeneutic role, by contrast, we risk to be delivered to whatever selves and stories these technologies (and their designers and employers) co-create. Understanding and evaluating the narrative role of AI are thus both a normative and a hermeneutic responsibility.

Moreover, while AI can be *meaning-full* and can be meaningfully integrated in our lifeworlds and perhaps even contribute to our narratives (literally and figuratively) and hence the meanings in the sense explained, they are not *meaning-making* in a strong, human and social-existential sense of the term captured by Ricoeur and other hermeneutics philosophers. Machines lack consciousness and therefore the *experience* that was a starting point in Ricoeur's analysis of narrativity and that is theorized by phenomenology and sometimes forgotten by postphenomenology.

Without having experience in the first place, we cannot rely on that experience in our narration and achieve a new understanding that transforms that experience. AI can only derivatively rely on human experience by extracting and analysing data that are supposed to represent human experience, for example human texts or images. And it participates in the meaning-making process in the sense that it offers this kind of knowledge. However, it cannot *complete* the mimetic and transformative hermeneutic process of narrative meaning-making; the transformation of understanding, the new understand emerges in the process but needs to happen through human experience, interpretation, and subjectivity. Moreover, meaning-making always happens in a situation and requires implicit and embodied knowledge that cannot be formalized. In Dreyfus's (1972) Heideggerian language, machines lack being-in-the-world.

In addition, considering further the epistemic dimension of responsibility (moral and hermeneutic), tensions will remain between, on one hand, human experience and sense-making and, on the other hand, what AI "knows" and does, since it remains a challenge for humans to make sense of the kind of knowledge provided by AI—at least the kind of AI that is not based on human reasoning and decision-making (e.g., decision trees), but that is based on machine learning. Especially, deep learning with neural nets seems to pose a problem here, since it is not possible for humans to understand how the machine comes to a decision. More generally, we must ask what we can do and should do with this kind of statistical knowledge produced by machine learning. Shall we (co-)create narratives in which correlations and probabilities play an increasing role? What place do we give this kind of quantitative knowledge in our lives? What narratives about our personal future do we want? As AI moves into the medical sphere (e.g., diagnosis based on data from image recognition, genome, etc.) and in our daily lives (analysis of data coming from our self-monitoring, consider again quantified self), this becomes a very practical and urgent question that may soon become relevant to everyone. Both at a personal and cultural level, it may well transform our self-understanding dramatically. Are these tensions there to remain, perhaps tragically so, or can they be overcome—without denying our humanity as social, experiencing, and interpreting beings? We will be faced with these problems, whether we want it or not. Next to responsibility to deal with moral issues, we have the hermeneutic-existential[1] task of making sense of ourselves and our lives given this new form of knowledge.

For this reason, next to taking care of moral and relational responsibility, it is important to take up this normative (but not necessarily moral) task at all levels. At a cultural level, we need to scrutinize the grand narratives about AI that currently pervade AI discourse. Consider for example again the transhumanist narrative about superintelligence, but also the humanist Frankenstein-like narrative of AI taking over and the posthumanist fairy tales of friendly AI-others we live with: all these narratives stand in need of interpretation and evaluation. We therefore better give both technology developers and citizens the education that gives them the hermeneutic tools to take critical distance from these narratives, revise these narratives, or imagine new, better narratives for the technological future. Classic humanist media such as books can and should still play a role in such an education. At the personal, interpersonal, and community level, we need to discuss what kind of plots and stories we want to create with or without AI, and what role we give AI and the kind of knowledge it creates in our lives and communities.

If we fail to exercise this hermeneutic and narrative responsibility, if we fail to make sense, AI may emplot and organize us in stories created by those who make profit from the technology[2] or who may have other aims that are not in line with our own aims. Then we leave the creation of narrative to big tech and its transhumanist supporters, to governments, and other players that wish to shape our lives. We may even end up playing the non-human character in the story: raw resources for data. And if we do not even know that story or if we are suddenly confronted with it when it is already too late, it may leave us in a state of hermeneutic ignorance and existential crisis, potentially leading to despair and anxiety. We would then be in the situation that we are living in the narrative that someone else wrote for us, *without even (fully) knowing it*. Or worse: we may expect meaning from technology, but technology *alone* will not and cannot provide narrative meaning. Then we end up in a situation of nihilism, in particular the passive nihilism Gertz (2018) warns for: we are unwilling to take responsibility for our lives and try to make the machines responsible (which, as I assumed here in this paper, is not even possible). The danger of falling into this void is not only a moral problem but also a hermeneutic one, especially in modernity. While having someone else write the story of our lives and communities was quite a familiar experience in pre-modern times, when people believed that there was a divine Author or Authors who would eventually pull things together and enable meaning, hermeneutically speaking, *for* us, it is hard to live and think like this today. And even many pre-modern people thought that they had a role to play, had to participate

[1] Note that more work is needed on the existential dimension of hermeneutic responsibility and, more generally, on the existential dimension of our relation to digital technologies and AI. Some points of departure may be Lagerkvist's *Existential Media* (2022) and my *Human Being @ Risk* (2013).

[2] This point could be further developed by making connections with political economy.

in the divine meaning-making process via rituals. In (late?) modern times, we wish to be the first author of our lives, even if it may no longer be possible to be its sole author. Today, AI is our co-author, and often an uninvited one. Yet, we wish to make sense of our lives and we wish to create *a narrative that we can live with*—regardless of further moral, relational, and scientific considerations. Therefore, and also taking into account the mentioned limits of AI due lack of consciousness, subjectivity, and experience (while acknowledging its participations and mediations in the relevant senses explained), this hermeneutic work cannot and should not and cannot be outsourced: neither to technology nor to the tech barons and politicians of this world, who may be expected to play the role of a deus ex machina that will save us. Again: humans are the main meaning-makers and, given that compared to AI only they have consciousness and experience, meaning-making always has to take place through them. Humans are the experiencers and meaning-makers. With a nod to Sartre (2007), who made a similar claim in the moral sphere when talking about freedom, I conclude that *we are condemned to make sense*. And in the age of AI, we are condemned to make sense of, with, or against AI. No one else or nothing else will do that or can do that in our place, least of all AI itself. It is our own, narrative responsibility.

6 Conclusion

This paper has distinguished various notions of responsibility: causal responsibility, moral responsibility, relational responsibility, and what I have called "narrative responsibility" as a form of "hermeneutic responsibility." I have noted that moral responsibility implies relational responsibility and shown how, in some accounts at least, this radically changes what we understand by moral responsibility. I also asked some *political* questions, usually neglected by standard accounts of moral responsibility. Yet my main aim in this paper was to develop the notion of "narrative responsibility" and to show how AI challenges these different notions of responsibility and underlying assumptions about humans and (knowledge of) the world. For example, the notion of causal responsibility becomes problematic when AI produces a different kind of knowledge. The new notion developed in this paper, however, was *narrative responsibility* as a specific form of *hermeneutic responsibility*. Using Ricoeur (and making a connection to Wittgenstein) and building on ongoing work in hermeneutics of technology, I have argued why we need the notions of narrative responsibility and hermeneutic responsibility next to other notions of responsibility, I have highlighted the role of imagination in exercising these responsibilities, and I have discussed the role of AI vis-à-vis these kinds of responsibility. This has led me to argue that while (1) AI participates in meaning-making and narration, (2) humans as conscious beings and beings-in-the-world are the necessary and main meaning-makers and narrators through which the hermeneutic process of meaning-making—for example by means of narrative—always has to pass and attains its completion. Therefore, I concluded that humans have narrative responsibility and, more generally, a hermeneutic responsibility: we are responsible to create a narrative we can live with, and to tell a story that makes sense of, with, or against AI. This revealed the normative dimension of hermeneutic responsibility.

Which narrative should we create? Answering this question goes beyond the scope of this paper, but throughout the paper, I already raised the question and have indicated a number of conflicting "grand" narratives: humanist, posthumanist, and transhumanist ones. Each of them do not only relate to a particular view of humans and the world, but also incur normative visions about the future of humans and the future of technology. However, nothing said here limits our narrative–hermeneutic space to these narratives. On the contrary, if we have the narrative responsibility I conceptualized, we must critically discuss the narratives and explore new narratives.

My general conclusion is therefore that to exercise our responsibility for AI and towards others, it is not sufficient to exercise our causal, moral, and relational responsibility. It is also important to connect moral and relational responsibility to the hermeneutic role we have as humans. Taking seriously this hermeneutic responsibility—and understanding the ways it is woven together with other kinds of responsibility and with normativity—is essential to our further efforts to engage with AI not only in morally and politically responsible ways, but also in meaningful ways, ways that make sense to us as humans. In the form of narrative responsibility, the concept of hermeneutic responsibility invites us to make our narratives about humans and technology explicit, interpret them, argue about them, and mobilize them in normatively relevant contexts—for instance, democratic discussions about technology. And if we reject the current narratives, for example those written by big tech, we have to create new ones and better ones. No one will do this for us. It is up to us to create the new stories and, for example, define the role of AI in those stories and in the writing of those stories. Thus, while anthropocentrism might be morally and politically problematic, hermeneutically speaking it is unavoidable: we are the main meaning-makers and story tellers. Meaning has to pass through us. However, if the more posthumanist directions in the hermeneutics of technology mobilized in this article are right, this will always have to be done in "co-authorship" with AI and other technologies of our time.

Funding Open access funding provided by University of Vienna. No funding was received for this work.

Data and code availability Not applicable.

Declarations

Conflict of interest No conflict of interest.

References

Aristotle (1984) Nicomachean ethics. In: Barnes J (ed) The complete works of Aristotle, vol 2. Princeton University Press, Princeton, pp 1729–1867

Bartoletti I (2020) An artificial revolution: on power, politics and AI. The Indigo Press, London

Berlin I (1997) Two concepts of liberty. In: Berlin I (ed) The proper study of mankind. Chatto & Windus, London, pp 191–242

Bijker WE, Hughes TP, Pinch T (eds) (1987) The social construction of technological systems: new directions in the sociology and history of technology. The MIT Press, Cambridge, MA

Braidotti R (2017) Posthuman critical theory. Journal of Posthuman Studies 1(1):9–25

Coeckelbergh M (2013) Human being @ risk: enhancement, technology, and the evaluation of vulnerability transformations. Springer, Dordrecht, New York

Coeckelbergh M (2018) Technology games: using Wittgenstein for understanding and evaluating technology. Sci Eng Ethics 24(5):1503–1519

Coeckelbergh M (2020) Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci Eng Ethics 26:2051–2068

Crawford K (2021) Atlas of AI: power, politics, and the planetary costs of artificial intelligence. Yale University Press, New Haven, London

Dennett D (1984) Elbow room: the varieties of free will worth wanting. MIT Press, Cambridge, MA

Dostoyevsky F (1972) Notes from underground (trans. Coulson J). Penguin Books, London

Dreyfus H (1972) What computers can't do: the limits of artificial intelligence. Harper & Row, New York

Duff RA (2005) Who is responsible, for what, to whom? Ohio State J Crim Law 2:441–461

Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge University Press, Cambridge

Frankfurt H (1969) Alternate possibilities and moral responsibility. J Philos 66(23):829–839

Gertz N (2018) Nihilism and technology. Rowman & Littlefield, London

Gilligan CC (1982) In a different voice: psychological theory and women's development. Harvard University Press, Cambridge, MA

Hakli R (2019) Moral responsibility of robots and hybrid agents. Monist 102(2):259–275

Harari YN (2015) Homo Deus. Vintage, London

Heney C, Banks W, Zimbardo P (1973) Interpersonal dynamics in a simulated prison. Int J Criminol Penol 1:69–97

Ihde D (1990) Technology and the lifeworld: from garden to earth. Indiana University Press, Bloomington

Ihde D (1998) Expanding hermeneutics: visualism in science. Northwestern University Press, Evanston, IL

Johnson DG (2014) Technology with no human responsibility? J Bus Ethics 127:707–715

Kudina O (2021) "Alexa, who am I?": voice assistants and hermeneutic lemniscate as the technologically mediated sense-making. Hum Stud. https://doi.org/10.1007/s10746-021-09572-9

Lagerkvist A (2022) Existential media: a media theory of the limit situation. Oxford University Press, New York

Matthias A (2004) The responsibility gap: ascribing responsibility for the actions of learning automata. Ethics Inf Technol 6:175–183

Milgram S (1963) Behavioral study of obedience. J Abnorm Soc Psychol 67:371–378

Nortvedt P, Hem MH, Skirbekk H (2011) The ethics of care: role obligations and moderate partiality in health care. Nurs Ethics 18(2):192–200

Pereboom D (2001) Living without free will. Cambridge University Press, Cambridge

Puig de la Bellacasa M (2017) Matters of care: speculative ethics in more than human worlds. University of Minnesota Press, Minneapolis, London

Reijers W, Coeckelbergh M (2020) Narrative and technology ethics. Palgrave, New York

Ricoeur P (1983) Time and narrative—volume 1 (McLaughlin K, Pellauer D, trans.). The University of Chicago, Chicago

Romele A (2020) Digital hermeneutics: philosophical investigations in new media and technologies. Routledge, New York, Abingdon

Rosenberger R, Verbeek P-P (eds) (2015) Postphenomenological investigations: essays on human-technology relations. Lexington Books, London

Rudy-Hiller, F. 2018. The epistemic condition for moral responsibility. Stanford Encyclopedia of Philosophy. Retrieved on April 13, 2021 from https://plato.stanford.edu/entries/moral-responsibility-epistemic/

Santoni de Sio F, Mecacci G (2021) Four responsibility gaps with artificial intelligence: why they matter and how to address them. Philos Technol. https://doi.org/10.1007/s13347-021-00450-x

Santoro M, Marino D, Tamburrini G (2008) Learning robots interacting with humans: from epistemic risk to responsibility. AI Soc 22(3):301–314

Sartorio C (2007) Causation and responsibility. Philos Compass 2(5):749–765

Sartre J-P (2007) Existentialism is a humanism (Macomber C, trans.). Yale University Press, New Haven, CT and London, England

Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP (2002) Instant neural control of a movement signal. Nature 416(6877):141–142. https://doi.org/10.1038/416141a

Sparrow R (2021) Why machines cannot be moral. AI Soc. https://doi.org/10.1007/s00146-020-01132-6

Talbert, M. 2019. Moral responsibility. Stanford Encyclopedia of Philosophy. Retrieved on April 12, 2021 from https://plato.stanford.edu/entries/moral-responsibility/

Thaler RH, Sunstein CR (2008) Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven, CT

van de Poel I, Royakkers L, Zwart SD (2015) Moral responsibility and the problem of many hands. Routledge, New York

van Inwagen P (1983) An essay on free will. Oxford University Press, New York

von Schomberg R (ed) (2011) Towards responsible research and innovation in the information and communication technologies and security technologies fields. European Commission, Brussels. Retrieved on April 13, 2021 from http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf

Véliz C (2020) Privacy is power: why and how you should take back control of your data. Penguin/Bantam Press, London

Winner L (1980) Do artifacts have politics? Daedalus 109(1):121–136

Wittgenstein, L. (2009). Philosophical investigations (revised 4th edn, Anscombe GEM, Hacker PMS, Schulte J, trans.). Wiley, Malden, MA

Yeung K (2018) A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Retrieved on April 12, 2021 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027