



# Making sense of the ‘Human’ in human-centered AI: an Arendtian perspective

Luuk Stellinga<sup>1</sup> · Paulan Korenhof<sup>1</sup> · Vincent Blok<sup>1</sup>

Received: 3 April 2025 / Accepted: 19 November 2025  
© The Author(s) 2025

## Abstract

In recent years, ‘human-centered artificial intelligence’ (HCAI) has emerged as a prominent framing device in the societal debate on the implications of AI. By adopting this phrase, AI ethics discourses make an appeal to the notion of the ‘human,’ while failing to critically reflect on its meaning. Against this background, we pose the question as to what ‘human’ is taken to mean in the context of HCAI. We apply a critical hermeneutic approach to analyze prominent HCAI literature and identify five key characteristics that shape the meaning given to the concept of the ‘human’ in HCAI: universalism, individualism, instrumentalism, psychologism, and exceptionalism. Following this, we introduce Hannah Arendt’s philosophical anthropology to provide an understanding of human existence as essentially political life, and argue that the HCAI discourse would benefit from considering the political dimension of human existence. We conclude the paper by proposing a research program for developing a reconceptualization of HCAI informed by philosophical anthropology.

**Keywords** Human-centered artificial intelligence · Ethics of AI · Hannah Arendt · Philosophical anthropology · Zoon politikon

## 1 Introduction

New developments in artificial intelligence (AI) have led to widespread societal debate. In both academia and popular discourse, the opportunities and risks of the increasing capabilities of technologies that use AI have become hot topics. Can AI be leveraged to address societal challenges? Will it create new challenges? Or will it merely perpetuate existing injustices? These types of questions are at the core of contemporary discussions on how to responsibly develop, assess, govern, and regulate the further development of AI. In an effort to provide direction, a variety of documents articulating ethical principles and guidelines have been produced (Floridi and Cows 2022; HLEGAI 2019; UNESCO 2022). Beyond the creation of such guidelines, research

continues to respond to new AI innovations, analyzing its impact on individuals and society, and evaluating controversial cases where AI technologies demonstrate harm or appear problematic otherwise.

In the context of this societal debate, ‘*human-centered artificial intelligence*’ (HCAI) has emerged as a prominent framing device to support responsible AI development. Roughly speaking, HCAI refers to the idea that AI is currently being developed with the goal of technological progress, while human needs and interests are overlooked. In other words, the current state of AI research and development is ‘technology-centered,’ with ‘human-centered’ AI as the desired alternative. While a number of researchers have attempted to define HCAI in the last few years (Schmager et al. 2023), the phrase is commonly not clearly defined and rather used as a way to frame the debate surrounding AI. This is the case, for example, in the European Commission’s ‘AI strategy,’ which aims to respond to the challenges and opportunities of AI by “placing people at the center of the development of AI–human-centric AI” (HLEGAI 2019). Despite the widespread agreement that humans should be centered, however, it remains unclear what this means. A map of the HCAI research landscape reveals considerable ambiguity in how researchers and developers interpret the

---

✉ Luuk Stellinga  
luuk.stellinga@wur.nl  
Paulan Korenhof  
paulan.korenhof@wur.nl  
Vincent Blok  
vincent.blok@wur.nl

<sup>1</sup> Wageningen University and Research, Wageningen, Netherlands

phrase (Capel and Brereton 2023). This ambiguity has led to philosophical engagement with HCAI, whereby scholars investigate and scrutinize its concepts and assumptions (Taylor et al. 2024; Ryan 2024; Braun & Meacham 2024). While still young and therefore modest, this philosophical scholarship is promising in its careful engagement with HCAI at a conceptual level.

What has so far remained underacknowledged in the literature is that the AI ethics discourse, by adopting the phrase HCAI, makes an appeal to the concept of the ‘human.’ It is, therefore, worthwhile to reflect on what it means to be human, especially in relation to AI technologies. Current HCAI discourse, however, offers little to no reflection on the concept of the ‘human.’ This is problematic because, since foundational concept structure our thinking about specific situations, the concept of the ‘human’ provides the foundation for our thinking and understanding of HCAI. In this article, we, therefore, aim to answer the research question of how we can understand the concept of the ‘human’ in the development and deployment of AI technologies such that it reflects careful philosophical analysis. With this, the article serves two aims: the first is to identify and critically reflect on the core elements of the human image that implicitly figures in prominent HCAI literature. The second aim is to offer a complementary perspective for understanding the ‘human’ in HCAI based on the analysis. For this, we introduce Hannah Arendt’s philosophical anthropology to provide an understanding of human existence as essentially political life. An Arendtian perspective is particularly fruitful, as it allows us to both analyze the implicit understanding of the human currently present in the HCAI literature, and demonstrate the need to broaden that understanding.

The paper is structured into three sections. Section 2 provides a critical analysis of top-cited articles from the HCAI discourse to identify implicit understandings of the word ‘human.’ We present and problematize five key characteristics of the assumed human image that exist throughout the literature: universalism, individualism, instrumentalism, psychologism, and exceptionalism. Section 3 critically reflects on these five characteristics following the philosophical anthropology of Arendt. By reconstructing Arendt’s understanding of human existence as a necessarily political existence, we can illustrate the one-sidedness of the understanding of the ‘human’ in HCAI discourse, and provide opportunities for broadening this understanding. Finally, Section 4 offers a discussion and concludes by proposing a research agenda for further research in the area of HCAI.

## 2 What does ‘human’ mean in human-centered AI?

Much of the literature on ‘human-centered AI’ does not explicitly define or elaborate on the meaning of the word ‘human.’ As a result, what ‘human’ is taken to mean in HCAI is left implicit, only to be derived from the discursive context within which the word is used. A research approach that allows for investigating implicit assumptions in textual sources is critical hermeneutics, which combines the hermeneutic and critical theory traditions to reveal and look critically at assumptions underlying the language used in a given discourse (Korenhof et al. 2021; Timmermans and Blok 2021). The approach is hermeneutic in its commitment to reconstructing a deeper understanding and interpretation of the text under consideration, and critical in analyzing and challenging the claims it recognizes as being taken for granted. In this article, we apply the critical hermeneutic approach to interpret representative texts in the field of HCAI to identify how prominent articles in HCAI literature implicitly conceptualize the ‘human.’

We applied this approach to a set of publications that we believe to be prominent and influential articles in the HCAI discourse. This set of publications consists of the ten most cited journal articles that explicitly present, discuss, or argue for HCAI. We collected these articles through a search using a combination of Scopus and Google Scholar, where we looked for publications that contain the term “Human-Centered AI”, “Human-Centered Artificial Intelligence”, “Human-Centred AI”, or “Human-Centred Artificial Intelligence” in their title, keywords, or abstract. We read the abstracts of the first 100 results and excluded articles that do not explicitly present HCAI as a framework or approach. From the remaining results, the ten most cited articles were analyzed. These texts are valuable, not merely due to the high number of citations, but also because they explicitly present and develop a conceptualization of HCAI through the articulation of specific perspectives and frameworks. These two factors together render them representative of the HCAI discourse as a whole. In our analysis, we looked for excerpts of the texts that provide insight into what is assumed about the meaning of the word ‘human.’ These excerpts were copied and categorized into a separate document, which allowed us to recognize which assumptions are held in the majority of the ten articles and constitute trends. The approach did not result in identifying a coherent meaning of the ‘human,’ but rather led to the identification of five key characteristics that shape the meaning given to the concept of the ‘human’ in HCAI.

This research is explorative in nature and does not intend to suggest that HCAI literature as a whole holds a uniform view of the human (as there is no discussion, we also do not

expect there to be agreement on the concept). Our aim is to draw out the concept and its key characteristics based on prominent articles to draw attention to the implicit assumptions shaping the concept of the ‘human’ and the implications thereof. We argue that the selected literature provides enough input to conclude that the assumptions can materialize in the discourse in the absence of critical engagement with the notion of the ‘human,’ and we provide examples to illustrate this. The ensuing discussion aims to enrich the discourse by considering the potential problematic outcomes of not engaging critically with its implicit conceptualization of the ‘human.’ This paper aims to make this conceptualization explicit, challenge it, and contribute towards articulating an explicit philosophical anthropology to ground the concept of HCAI.<sup>1</sup>

## 2.1 Universalism

The first characteristic that shapes the concept of the ‘human’ in the selected literature is that of universalism. We draw this characteristic from the level of generality at which claims about humans are made as well as from the lack of consideration for diversity between different humans. For example, the literature states that “humans generally derive pleasure from seeking mastery, improving their skills, and feeling fully engaged” (Shneiderman 2020a) or “seek explanations when there is a gap of understanding” (Liao and Varshney 2022) and does not regard possible deviations from those sentiments. The word ‘human’ seems throughout the literature posited as a universal category, implying the existence of a universally shared human experience at least insofar as it concerns the development and use of AI technologies. This suggests that there is such a thing as *the* human that can and should be placed at the center of AI development. A prominent figure in Ozmen Garibay et al. further demonstrates this, presenting an illustration of the AI development cycle orbiting a minimalist reproduction of Leonardo Davinci’s *Vetruvian Man* (Ozmen Garibay et al. 2023, 329). Davinci’s *Man*, stripped of defining characteristics, is taken to stand for the human experience, which is assumed to be universal and accessible to AI developers.

A universalist approach to the ‘human’ appears problematic when considering that human experiences are deeply heterogenic in actual society, including in relation to technology. Cultural diversity, neurodiversity, gender diversity, and many other kinds of differences that exist between humans challenge the idea of human-centered AI as being about *the* human. Based on cultural background, knowledge,

experiences, and situatedness in society, humans as users or ‘targets’ of AI may have different requirements for HCAI to connect to their needs and ensure that they as a human are indeed considered at the center of AI technologies (Ess 1999). For example, people with disabilities may suffer when AI technologies are imagined and developed with only able-bodied human beings in mind (Shew 2020). The risk of leaving an implicit universalist approach to the ‘human’ unchallenged consists of a reduction of the heterogeneity of human experiences to one homogenous experience. Such a one-size-fits-all mentality is common in design thinking and has been critiqued for reflecting a privileged perspective, perpetrating societal biases, and harming marginalized perspectives (Place 2023).

## 2.2 Individualism

The second characteristic we identified from the literature is that there seems to be an overall focus on the ‘human’ as an individual being. The approach of the ‘human’ as individual being is apparent in two ways. First, the literature largely refers to humans at the level of individuals, focusing on people in an individual capacity, as “user,” “operator,” or “client” (Riedl 2019). In the studied literature, we did not come across vocabulary that refers to humans in collective capacities, for example as ‘family’ or ‘community’. While the literature analyzed does sometimes acknowledge the importance of the wellbeing of ‘society’ or ‘humanity,’ especially in introductory paragraphs, it does not operationalize a collectivist perspective in developing HCAI approaches. The second way in which we were able to identify a focus on the individual surfacing in the literature is in discussions on the relationship between humans and AI systems, which takes place at the level of the interaction between an individual person and technological artifact. In the literature, ‘placing humans at the center’ is typically taken to most immediately require improved user-interface design such that systems are “human-controllable by keeping humans as the ultimate decision makers, usable by providing effective interaction design, and explainable by providing understandable output to humans” (Xu et al. 2023). This leads to a dominant focus on the interface of AI technologies and an understanding of its impact as immediate impact on those human beings that are in direct contact with the system, in turn reflecting an individualist perspective.<sup>2</sup>

<sup>1</sup> We take ‘philosophical anthropology’ to refer to the critical reflection on the concept of the ‘human,’ or ‘anthropos,’ broadly speaking, rather than to a particular philosophical discipline or tradition.

<sup>2</sup> It should be noted that outside of the analyzed literature there exist proposals that take a more collectivist perspective, such as the “AI4People” framework. The existence of such proposals does not invalidate the claim that an individualist assumptions looms, and in fact demonstrates the need for critically engaging with the question as to whether human existence can and ought to be conceived of as individual existence in AI ethics.

A risk of taking an individualistic perspective in HCAI is that the social embeddedness of human existence is neglected. The selected HCAI literature, drawing largely from the field of human–computer interaction (HCI), offers no handholds or views on how to deal with the ways in which human actions and perceptions, as well as their needs and desires, are dynamically shaped by their social context. As a result, HCAI treats the relationship between individuals and society in a way that echoes methodological individualism. This suggests an underlying understanding of human existence as one in which society refers simply to the multiplicity of many individuals, and can consequently be acknowledged by accounting for the sum of individual needs. While pervasive in modern societies, this individualist perspective is often criticized in social scientific discourses for mischaracterizing the relationship between individuals and societies and ignoring the ways in which individual identities are socially shaped (Cortois 2017). Individualism can be challenged by contrasting it with perspectives that acknowledge the social embeddedness of human existence, as, for example, in philosophical research on collective intentionality (Searle 2002; Zahavi 2021). Taking seriously such a collectivist perspective problematizes the individualist assumption of HCAI, and calls for critical reflection.

### 2.3 Instrumentalism

Throughout the selected literature, the phrase ‘human-centered AI’ is used almost interchangeably with ‘user-centered AI.’ In this, we read an implicit conceptualization of ‘human’ as ‘user,’ further demarcating the understanding of human existence in HCAI literature. With this, the relationship between humans and AI technologies is primarily framed as a relationship between a user and the product or tool that they are using. Shneiderman accentuates this focus when he writes: “HCAI’s design thinking approach puts the human users at the center, emphasizing user experience design, measuring human performance, and celebrating the new powers that people have” (Shneiderman 2020b). The focus on use is further illustrated by the prominent goal of usefulness: “to expand the usefulness of AI and ensure that future applications are human-centered” (Ozmen Garibay et al. 2023). In a similar way, the literature frequently moves from a concern for humans to a concern for consumers, as exemplified by the assertion that AI technologies should be developed with the intent of serving consumer needs. Xu states that “AI solutions must have a clear purpose. By providing useful AI, such solutions can match user needs and thus gain acceptance and generate economic benefits” (Xu 2019). Equating humans to users and consumers propagates an understanding of the human–AI relationship as tool-use and consumption. As such, the human is framed in an instrumentalist perspective on human–technology relations.

What is not captured in this instrumentalist picture are all the other ways in which human beings can relate to AI technologies. The idea that human beings ‘use’ or ‘consume’ AI implies a particular type of engagement with the technology, akin to how we may use a tool or a scientific instrument, or consume digital media content. This does not map onto the variety of other ways in which human beings can relate to AI systems, which are often more reciprocal or occur through the integration of AI systems in larger networks. Because of this, some have suggested retiring the term ‘user’ altogether in favor of words like ‘humans’ or ‘people’ (Majewski 2024). On top of this, the development and functioning of AI systems involves labor that is often neglected in popular discourse on AI (Crawford 2021; Irani and Silberman 2013). The workers who do this labor are not only essential for the existence of AI systems by performing crucial tasks from mining rare earth materials to labeling data, but are also greatly affected in their wellbeing. Importantly, these workers are not ‘users’ of the technology and will, therefore, be overlooked if HCAI’s focus is limited to use and consumption. In general, an instrumentalist approach of the human in HCAI risks a problematic simplification of the complexity of human–technology relationships.

### 2.4 Psychologism

A fourth characteristic we derive from the analyzed literature is a psychologistic approach to humans. In the literature, we encountered a recurring appeal to scientific models of cognitive and behavioral psychology. After acknowledging the importance of taking the ‘human’ into consideration, HCAI literature often refers to cognitive and behavioral psychology to provide scientific input on how human beings process information and behave in interaction with AI technologies. Liao and Varshney, for example, suggest that many of the current problems with AI deployment arise from a disconnect between the assumptions that underly technical approaches and people’s cognitive processes (Liao and Varshney 2022). In response, they appeal to dual-process theories to provide a scientific understanding of how people process information and make decisions. This theoretical perspective then forms the basis for human-centered design interventions, such as ‘nudging’ people into an analytical mode of thinking. Similarly, Riedl suggests that if AI technologies were designed to “model sociocultural beliefs and norms”, they could “make more educated guesses on how to anticipate and respond to human needs” (Riedl 2019, p. 34). Argumentative steps like this imply that ‘centering humans’ in the development and deployment of AI technologies is possible on the basis of an understanding of human beings as characterized by their cognition and behavior according to scientific theories. We refer to this perspective on human existence as ‘psychologism’.

The psychologicistic perspective of HCAI runs the risk of overlooking aspects of human existence that are not, or cannot, be captured by models of cognitive and behavioral psychology. This is not to say that such models, or the research that informs them, are intrinsically flawed. Rather, HCAI's willingness to draw upon these models for understanding human existence (instead of engaging in stakeholder engagement, for example) reflects a particular understanding of human beings as scientifically 'modellable.' While useful, restricting HCAI to this perspective on human existence also has drawbacks. First, psychological theories are likely to reflect the values embedded in the context within which they are developed. Consequently, models of human cognition and behavior can be biased towards certain parts of the human population. Marginalized communities especially may not 'fit' the models that are developed in vastly different circumstances (Henrich et al. 2010). Another reason why relying solely on a psychologicistic approach to conceptualizing the 'human' can be problematic follows from the more radical claim that certain aspects of human existence are intrinsically irreducible to scientific theories. This stance, according to which humans are inherently ambiguous and indeterminable, can be grounded in Arendt's philosophical anthropology and we will return to it as such in the following section. For either of these two reasons, the psychologicistic framing, manifesting itself in the attempt to predict human–AI interaction according to psychological models, is philosophically precarious.

## 2.5 Exceptionalism

Finally, a last characteristic we identified as being present in the background of the complete body of literature analyzed is the claim of human moral exceptionalism. The literature reflects the general idea that the ethical aspects of AI technologies can be adequately addressed through consideration of human needs and values alone. This implies that moral concerns are intimately connected to human experiences, and that 'human', therefore, refers to a morally exceptional category. Following this, it makes sense to address the ethical aspects related to AI technologies by shifting from 'technology-centered AI' to 'human-centered AI.' We derived this characteristic from the literature's general lack of consideration for nonhuman animals and the natural environment. In all of the documents analyzed, none dedicate attention to the impact of AI technologies on the wellbeing of nonhuman animals. Where the natural environment is considered, notably through the claim that AI ought to be sustainable, it is framed as instrumentally valuable to the greater goal of improving human wellbeing (Ozmen Garibay et al. 2023).

The question is whether this exceptionalist approach is ethically viable given AI's impact on the world beyond

humans. We consider that at least some other entities besides human beings are morally relevant and impacted by AI technologies. In this case, the idea of human-centered AI falls short of capturing the morally relevant impact of AI, and is instead grounded in an implicit anthropocentrism. This stands in contrast with growing awareness of the impact that AI technologies have on nonhuman animals and the natural environment. Most notably, the contribution of the development of AI technologies to greenhouse gas emissions and high energy consumption demonstrate its considerable and alarming environmental costs (Strubell et al. 2019; van Wynsberghe 2021). In addition, AI technologies are found to impact both domesticated animals and wildlife in significant and problematic ways, for example by allowing for unethical treatment through sensor technologies or by perpetuating speciesist biases (Hagendorff et al. 2023; Singer and Tse 2023). The tension between the anthropocentric perspective present in the analyzed literature and the explicitly non-anthropocentric research revealing AI's harm towards nonhumans calls for a need to critically reflect on whether humans should or should not be understood as morally unique. As long as this is insufficiently addressed, the risk of anthropocentrism remains in HCAI.

The five characteristics that shape the notion of the 'human' in prominent HCAI literature discussed in this section do not refer to explicitly held beliefs in the HCAI discourse. Instead, the five characteristics reflect implicit conceptualizations of human existence, revealed through interpreting the way in which human beings are presented in the literature analyzed. This does not mean that these characteristics are unilaterally present throughout the entirety of the AI ethics discourse, as some of the research we have already appealed to demonstrates. The different characteristics can be taken as part of one human image, as they may rely upon or strengthen one another, but can also be taken separately and critiqued separately. The point that we want to make is that each of these characteristics have drawbacks and are problematic if left implicit in HCAI discourse, as they may materialize in HCAI unchallenged. It is, therefore, pivotal to actively consider, and if needed, challenge the characteristics that shape the meaning of the 'human' in HCAI. We argue that HCAI should be founded in a philosophical anthropology, so as to provide an informed understanding of what 'human' should mean in HCAI.

## 3 An Arendtian perspective on human-centered AI

In this section, we provide a complementary philosophical perspective on what 'human' might mean in the context of HCAI. Having concluded that prominent HCAI literature currently conceptualizes what it means to be human along



the lines of universalism, individualism, instrumentalism, psychologism, and exceptionalism, and having posed critical questions concerning these characteristics, this section provides an alternative perspective on human existence based on the philosophical anthropology of Hannah Arendt, presented in her book *The Human Condition* (Arendt 2013).<sup>3</sup> While Arendt's understanding of human existence is by no means the only understanding available in philosophy, we believe it is particularly valuable for the purposes of this paper, as it affords a complementary perspective to the human image underpinning prominent HCAI literature, while also demonstrating a need to go beyond this image. More specifically, Arendt acknowledges both the economic and political dimensions of human life. In reconstructing Arendt's understanding of human existence in the modes of labor, work, and action for critical analysis and rethinking of developments in technology and innovation, this paper resonates with recent scholarship that performs a similar gesture (von Schomberg and Blok 2023; Waelen 2025).

In this section, we reconstruct Arendt's philosophical anthropology to critically reflect on the framing of the 'human' in prominent HCAI literature, and illustrate how it has forgotten political life, i.e., the human as *zoon politikon*.

### 3.1 *Vita activa*: labor work and action

In *The Human Condition*, Arendt makes an attempt to provide a conception of the fundamental and basic capabilities that all human beings share. Arendt argues that philosophers in the Western tradition have long focused their investigations on the particular human activities of thought, reflection, and understanding. By focusing on these activities, the life of contemplation, or *vita contemplativa*, has come to be seen in philosophy as superior to the active life, or *vita activa*, through which humans engage in/with the world. According to Arendt, the prioritization of the *vita contemplativa* has led philosophers to disregard (political) action as secondary. Arendt rejects this hierarchy, and instead holds that both contemplative and active life are important for being human, but that a reconsideration of the active life is necessary to recover the political dimension of human existence. As such, Arendt aims her attention at the *vita activa* (leaving the consideration of the *vita contemplativa* to her final work *The Life of the Mind* (Arendt 1981)), and provides an investigation of the basic human ways of acting in the world.

Arendt identifies three basic activities: labor, work, and action. Each activity corresponds to a mode of human relating to the world, as *animal laborans*, *homo faber*, and *zoon politikon*, respectively. In this section, we will explore these three basic activities to conceptualize Arendt's concept of human living and acting in the world, which can inform our critical reflection on the image of the 'human' in HCAI in the subsequent section.

Labor is the activity that corresponds to the biological existence of human beings. Human beings labor when they respond to the necessities of living on earth. This includes what humans do to maintain survival, such as eating, sleeping, and nurturing, but also refers to the activities that humans engage in to make a living, to earn an income so that they can buy food and secure shelter to survive. Labor sustains the life process of the individual organism, their society, and ultimately the species. Characteristic of all labor is its repetitiveness, which reflects the cyclical nature of the metabolism of the life process. Accordingly, the activity of labor is always temporal and fleeting, with nothing remaining beyond the metabolic process (Arendt 2013, p. 88). Anything produced during the laboring process is immediately consumed within the metabolic cycle of labor itself. In their capacity as laboring and consuming, human beings appear as *animal laborans*, 'laboring animal,' and relate to their environment only as fertile ground.

Human beings maintain a different relation to their environment when they engage with it in the mode of work, which refers to the making and fabrication of durable things. The main difference between labor and work lies in the fact that while labor does not produce anything beyond what is immediately consumed, work always results in an end-product. A process of labor such as cleaning a room has no durable result beyond making a living for the cleaner, but a process of work like building a table also results in a material object (a table) that will remain even when the table builder is no longer there. Like labor, work involves getting one's hands dirty, but unlike labor, it always produces something that remains and is not directly consumed in the metabolic process. Consequently, while labor is cyclical, work is characterized by a linear temporality, with the making and production of material objects as its end. Those material objects are objects of use, which are not consumed immediately, but in principle persist beyond the period of their use.<sup>4</sup> The totality of the products of work come together to form

<sup>3</sup> In reading Arendt's work as presenting a philosophical anthropology, rather than a political theory, we follow prominent interpreters such as Peter Gratton, Margaret Canovan, and Dana Villa, who argue that Arendt's major contribution in *The Human Condition* consists of a critical analysis of a misunderstanding of the human condition in her time (Gratton 2021; Canovan 1992; Villa 1995).

<sup>4</sup> Arendt acknowledges that all objects of use, like objects of consumption, are not impervious to wear and tear, and ultimately decay. Nonetheless, the nature of worldly things is that they endure at least a while beyond the initial period of use. Arendt uses the example of worn-down shoes to make this point. While shoes ultimately wear down until they are unusable, the point of shoe-making is to create something able to withstand natural forces (Arendt 2013, p. 138).

the artificial world of things in which we live and act. In their capacity as working and using material things, human beings do not relate to their environment as *animal laborans*, but instead as *homo faber*, ‘man the maker,’ actively creating an artificial world of things, shaping their surroundings according to their desires. As *homo faber*, human beings appear as “the lord and master of the whole earth” (Arendt 2013, p. 139).

The activities of labor and work belong to the private realm of economic life, which Arendt contrasts with the public realm of political life. In the private realm, human beings live according to their individual needs and desires. As such, the private realm refers to places where humans literally exist in private, such as within the confines of their family homes (the *oikos*), but also refers to economic life (*oikonomia*). In economic life, like exchanging goods and services in the market place, or engaging in the work of building technical artifacts, people are in company of other people, but this being-together is primarily determined by the objective to satisfy their own needs. In the public realm, on the other hand, human beings are in the company of others not primarily to serve their own interests, but to deal with matters that are necessarily shared with others and cannot be reduced to individual needs and desires. Labor and work, alongside their counterparts of consumption and use, are ultimately individual activities, and thus belong to the private realm of economic life. They describe human existence only insofar as humans seek to satisfy individual needs in a material world of things that are continuously produced, used, and consumed. A philosophical anthropology that views human existence only as a composite of the activities of labor and work would, therefore, result in a view of human existence as purely economic existence, characterized solely by engaging in economic activities, while missing the reality of political life.

In other words, the activities of labor and work are necessary dimensions of human existence, but they do not cover the full extent of the *vita activa*. If it is through labor that human beings are capable of ensuring their survival, and through work that they can build a material world, then it is through the activity of *action* that humans are capable of giving meaning to that world. Action refers to the human capacity of making sense of one’s environment through speech and deed. The result of action is not the material world of things, like in the case of work, but instead an immaterial world of human affairs produced in stories, opinions, and judgments of human actors, who appear in their capacity as acting as citizens in a participatory democratic society, or *zoon politikon*. According to Arendt, acting is inherently a *shared* activity of humans living and acting in a shared world, and always takes place in the presence of others. The meaning of our living and acting in the world, Arendt argues, can only be established by “talking about”

it, which involves the perspective of others as a principle (Arendt 2013, p. 183). In the mode of action, then, human beings always appear among others, conditioned by plurality, which Arendt defines as “the fact that men, not Man, live on earth and inhabit the world” (Arendt 2013, p. 7). Therefore, in action, human beings operate not in the private realm of economic life, but in the public realm of political life as equal but distinct citizens with heterogeneous perceptions, visions, and values. Taking action to be a fundamental human activity leads to a broader understanding of human existence beyond *homo economicus* as an essentially political existence.

According to Arendt, the plurality of action is accompanied by unpredictability. Action provides humans the means for making sense of their worldly environment, and in turn allows them to disclose themselves to others. This is possible because each human being is unique and brings a new perspective to the world, which is what makes it possible to ask a human being *who* they are, while of other entities it only ever makes sense to ask *what* they are.<sup>5</sup> Through action, then humans are capable of revealing their personal distinctiveness and bringing something radically new into the world. Because each human being acts from their own unique perspective, the outcome of any given action is inherently unpredictable. According to Arendt, the capability of action involves unpredictability and improbability, as it is informed by the uniqueness of one’s perspective (Arendt 2013, p. 178). As *zoon politikon*, then a human being appears in the world not only in the context of the presence of others, but also as spontaneous and unpredictable. The activities of *animal laborans* and *homo faber* are determined by biological necessity and the material world, respectively, but action reveals the human as free, as it is the only activity that humans initiate themselves in the shared space of political existence.

### 3.2 *Vita activa* and human-centered AI

Arendt’s philosophical anthropology, expressed as *vita activa*, allows us to critically reflect on and challenge the current conceptualization of the ‘human’ in prominent HCAI literature, and specifically with regard to the characteristics of universalism, individualism, instrumentalism, and psychologism. The characteristic of exceptionalism will be briefly considered in the discussion.

In the previous section, we discussed the tendency in HCAI literature to adopt a universalistic perspective by taking the term ‘human’ to refer to a universal category.

<sup>5</sup> “While in *The Human Condition* Arendt presents action as a uniquely human capability, in *The Life of the Mind* she appears more open to idea that nonhuman animals are also capable of action. For a detailed discussion see (Rossello, 2022).

The risk of this universalism concerns a reduction of the heterogeneity of human experiences to one homogenous experience, in turn overlooking diversity between humans. From an Arendtian perspective, we can now scrutinize the universalist approach philosophically. While it is true that all human beings share the same existential conditions and consequently engage in the basic activities of labor, work, and action, Arendt rejects the idea of a single human experience (Gratton 2021). In fact, Arendt writes that “we are all the same, that is, human, in such a way that nobody is ever the same as anyone else who ever lived, lives, or will live” (Arendt 2013, p. 8). In other words, the fact that human beings are never identical to one another is a crucial aspect of human existence. Furthermore, human beings are capable of expressing their distinctiveness exclusively in the mode of action. A universalist conceptualization of the ‘human’, thus, disregards diversity and individual distinctiveness, and in effect is unable to acknowledge human existence in the mode of acting, as *zoon politikon*.

This brings us to the individualist conceptualization of the ‘human,’ which takes for granted the possibility of accounting for human existence at the level of individual persons. The risk that we identified in discussing the individualist approach is a neglect of the social and cultural embeddedness of human existence. From an Arendtian perspective, we can challenge this assumed individualism by referring to the public realm of political life, wherein which humans engage with public affairs irreducible to individual concerns. The public realm consists of shared meanings enacted in the interaction of people, which ultimately forms an intangible “web of relationships” (Arendt 2013, p. 181). It is only in the context of this web of relationships, which already exists before any individual human being exists, that humans can act, give meaning to the environment around them, and disclose who they are to others. In neglecting this embeddedness and implicitly treating human existence purely as individual existence, the prominent HCAI literature appears to substitute the private realm for the public realm, and again forgets the political dimension of human existence.

The instrumentalist framing in the HCAI discourse is reflected by the shift from the word ‘human’ to the words ‘user’ and ‘consumer.’ This suggests an approach to humans as relating to technologies primarily as tools and consumer-products, simplifying the nuanced reality of human–technology relations. While Arendt was not a philosopher of technology per se, her philosophical anthropology may, nevertheless, help us reflect on what aspects of human relating to technology can enhance current HCAI discourse.<sup>6</sup> The

activities of consumption and use are closely related to those of labor and work; labor always implies the consumption of its fruits, and the products of work are always proper objects for use. Accepting Arendt’s understanding of labor and work as belonging to the private realm of economic life, it follows that the HCAI discourse frames the relationship between humans and AI technologies primarily as one belonging to economics. What is left out here is the human mode of relating to technology in action, and consequently the role of AI in the public realm of political life. In more concrete terms, prominent HCAI literature frames the human–AI relation in such a way that it does not account for the different ways in which AI technologies are shaping the space for interaction between humans to make (political) sense of their shared world, rather focusing on the ways in which AI technologies are used and consumed economically.

The characteristic of psychologism, by which it is accepted that human existence can be acknowledged through the scientific theories of cognitive and behavioral psychology, finds its challenge in Arendt’s claim to the fundamental unpredictability of human action. As discussed in the previous section, HCAI literature frequently finds opportunities for taking humans into consideration in the development of AI technologies in theories of cognitive and behavioral psychology. Through incorporating such theories, it would be possible to anticipate how the user of an AI system will come to process information provided by an AI system, and how they will likely interact with its interface. Underlying this idea, however, lies the desire to predict how humans act. From an Arendtian perspective, this desire is problematic when considering the fundamental unpredictability of human action. In acting, human beings are inherently unpredictable, as they express their distinctiveness from others. As a result, incorporating theories of cognitive and behavioral psychology cannot, in principle, account for human spontaneity. At best, this means that prominent HCAI literature considers human existence only in the modes of labor and work. At worst, it means that the studied HCAI literature reduces human existence in the mode of action to the modes of work and labor, as such collapsing political life into economic existence.

The Arendtian reflections on the assumptions of universalism, individualism, instrumentalism, and psychologism reveal a common thread: the prioritization of economic human existence over political human existence. We should note that this section is not intended as a critique to be levied at the analyzed HCAI literature in a straightforward way. Just as the five characteristics identified in the previous section

<sup>6</sup> There exists a small but growing set of scholars who do read Arendt’s work as part of the discipline of philosophy of technology, and thus view her explicitly as a philosopher of technology (see Bas 2022; Longo 2024). While this paper is not intended to contribute to that discussion, we recognize the importance of uncovering Arendt’s

Footnote 6 (continued)

analysis of technology and human–technology relations for contemporary philosophy.



**Table 1** Research agenda for a concept of HCAI informed by philosophical anthropology

Assumption	Conceptual question	Practical question
Universalism	In what ways can we conceptualize the human condition to recognize both the shared experiences among all human beings and the unique aspects of individual lives?	What approaches can ensure that diverse human experiences, in particular those of marginalized groups, are taken into consideration in the development and deployment of AI technologies?
Individualism	What is the relationship between individual and collective human existence?	How can collective agency and community be integrated in the development and deployment of AI technologies?
Instrumentalism	In what ways can the human–technology relationship be understood beyond only use and consumption?	How to account for the impact of AI technologies on those who are not direct users or consumers, such as click-workers or members of affected communities?
Psychologism	How does human unpredictability and ambiguity challenge psychological models of human behavior and cognition?	How can human unpredictability and ambiguity be integrated in the development and deployment of AI technologies?
Exceptionalism	What is the relationship between human existence and other life forms seen from a non-anthropocentric viewpoint?	How to identify and assess anthropocentric biases in the development and deployment of AI technologies?

were derived from implicit and overarching conceptualizations of the ‘human,’ and do not refer to any particular claim, framework, or author, so too is the Arendtian critique presented in this section not directed at anyone in particular. Rather, it is a critique of the human image that functions in the background of the discourse as a whole. This article is an attempt to make that human image explicit, and to reflect on and complement it with the philosophical anthropology of Arendt. The point of this section is also not to shed HCAI in a purely negative light. While it is beyond the scope of this article to present a progressive Arendtian account of HCAI, we do believe that certain lessons can be learned from Arendt’s insights to inform an improved notion of HCAI. In particular, we believe that Arendt can help to broaden the scope of HCAI approaches beyond an economic framing and acknowledge the political dimension of human–AI relations, and can in turn broaden the scope of the ethics of AI technologies.

## 4 Discussion and conclusion

This paper responds to the observation that prominent articles in HCAI literature do not reflect on the meaning of the word ‘human.’ We have, therefore, focused on the question concerning what is meant by the word ‘human’ in the HCAI debate. Before beginning the discussion of our analysis, it is valuable to briefly return to the importance of this question. Insofar as the current societal debate on AI has adopted the phrase ‘human-centered AI’ as an overarching framing device to increase its trustworthiness and responsibility, it makes an appeal to the concept of the ‘human.’ At the same time, however, it does not actively reflect on this conceptual dependency. From a philosophical perspective, this is problematic, as basic concepts provide the background against which we come to understand and think about specific situations. If the debate on the implications of AI occurs

against the background of an unexamined conception of the ‘human,’ that conception will inevitably affect the debate, whether it is made explicit or not. The current paper demonstrates this by revealing that despite its lack of engagement with the meaning of the word ‘human,’ prominent HCAI literature, nonetheless, presupposes certain conceptualizations of the concept of the ‘human’ that can be problematized and critiqued. Making these conceptualizations explicit allows for their critical examination and creates a space for developing a philosophical anthropology to ground HCAI.

As presented in the analysis of section 2, the concept of HCAI is accompanied by five characteristics regarding the meaning of the word ‘human.’ Current conceptualizations treat human experiences as universal, approach human existence at the level of individual existence, frame human–technology relations primarily in terms of use and consumption, characterize humans by their cognition and behavior, and take humans to be morally exceptional beings. As such, prominent HCAI literature appears to be guided by a limited understanding of the human condition. We do not provide evidence to conclude that the assumptions are widely held in HCAI discourse as a whole, as the set of articles analyzed is too small to infer such a broad claim. Instead, we hope to inspire future research that analyzes the discourse more rigorously, to describe the presence of these and other characteristics in more detail. Each of these characterizations can have far-reaching implications for AI design and invites a critical perspective from philosophical anthropology. In section 3, we, therefore, introduced an Arendtian perspective according to which human existence is understood as essentially political life. This perspective is particularly valuable, not because Arendt’s philosophical anthropology is unequivocally true, but because it provides a broader perspective on the human condition that enables us to raise critical questions regarding the ways in which the ‘human’ is implicitly conceptualized in HCAI literature. The main takeaway of this section is that HCAI discourse would

benefit from taking into account the political dimension of human life in complement to the current emphasis on its economic dimension.

Connecting the identified limited conception of the ‘human’ to the broader societal debate on AI, we find that it appears to correspond to the “blind-spots” of AI ethics research (Hagendorff 2022). The field of AI ethics typically focuses on topics such as explainability, fairness, and privacy insofar as those can be framed as technical and apolitical problems, while concerns that evoke political questions are neglected. We believe that this limit is, at least in part, related to the one-sided conception of the human image that figures in the background of the current debate on AI. To give an example, current AI ethics struggles to address the way in which AI has given rise to a growing and ethically suspect click-work industry through which typically low-wage laborers from the global south provide essential annotation work for the development and deployment of AI technologies. A variety of reasons can explain why this issue is often overlooked in AI ethics, but it is worthwhile to consider that there is a connection to the literature’s implicit philosophical anthropology. Broadening the philosophical anthropology underlying the HCAI discourse, and perhaps the societal debate on the implications of AI, then, may also afford a broadening of the ethics of AI.

This paper opens up a critical perspective informed by the philosophy of Hannah Arendt, but there is more research to be done on how such a critical perspective can inform the practical contexts within which AI technologies are developed and deployed. Furthermore, the Arendtian perspective is here only introduced to reflect on the conceptualization of the ‘human’ in prominent HCAI literature, and should not be taken as the final judgment on what it means to be human. There has been much critical engagement with Arendt’s philosophical anthropology that reveals significant limitations of her work. In the context of this paper, two limitations provide fruitful ground for further nuance in reflecting on HCAI. For one, Arendt appears to posit a rigid distinction between the realms of the private and the public, which may oversimplify complex interactions and experiences of women and marginalized groups (Benhabib 2018). The rigidity of the private/public distinction may also limit our understanding of the impact of AI technologies, as they may blur the boundaries between private and public life. Second, Arendt has been critiqued for enforcing an anthropocentric thinking that puts human animals over nonhuman animals and the natural world (Plumwood 2002). Being critical of this anthropocentrism is particularly valuable for further discussion on the exceptionalist assumption in HCAI. We also recognize that many of the problems raised in this paper are discussed at length in ongoing debates on post-humanism, feminism, anti-racism, and the role of humans in the Anthropocene. For the purposes of responding to the

research question of this paper, which concerns the concept of the ‘human’ in particular, we have elected to introduce and discuss the philosophical anthropology of Arendt which offers a productive theoretical perspective on what it means to be human. We believe that this perspective opens up the possibility for further engagement with the problematic aspects of HCAI from more fine-grained perspectives relating to the topics discussed in the debates mentioned above, and encourage scholarship along these lines.

We conclude that the prominent HCAI literature reflects an implicit and questionable understanding of the meaning of the word ‘human,’ leading to a one-sided approach to human existence. In response, we claim that we need a broader understanding of the human condition to avoid blind spots in AI ethics and formulate a philosophically grounded notion of HCAI. To achieve this, we propose an agenda for future research on HCAI. The research agenda follows the structure of the five assumptions presented in this paper. For each assumption, a conceptual question is proposed to invite critical reflection on the assumptions about the ‘human’ underlying HCAI, and a practical question to reveal the connection between the philosophical assumptions and concrete AI development practices. Hereunder are the questions that we propose for further research on the philosophical anthropology of HCAI. (See Table 1)

This paper engages philosophically with one of the most prominent framing devices in the current societal debate on the impact of AI: *human-centered AI*. The emergence of and widespread commitment to HCAI as a framing device reveals the importance of the question as to what it means to be human, especially in a time where humans and society are deeply affected by AI technologies. We believe that the structural involvement of philosophical anthropology in the discourse of HCAI can guide the development of a pluralistic concept of responsible and trustworthy HCAI in the future.

**Author contributions** All authors contributed to conception and design of the research. Literature analysis and initial development of the theoretical perspective was done by LS. PK and VB supervised the drafting of the manuscript. All authors read, reviewed, and approved the manuscript.

**Funding** Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Nwa.1332.20.002, Nwa.1332.20.002

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License,

which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Arendt H (1981) *The life of the mind: the groundbreaking investigation on how we think*. HMH
- Arendt H (2013) *The human condition: second edition*. University of Chicago Press
- Bas M (2022) Technological mediation of politics: an arendtian critique of political philosophy of technology. April DOI 10.3990/1:9789036553667
- Benhabib S (2018) *The claims of culture: equality and diversity in the global era*. Princeton University Press, Princeton
- Braun M, Meacham D (2024) A plea for (in)human-centred AI. *Phil Technol* 37(3):97. <https://doi.org/10.1007/s13347-024-00785-1>
- Canovan M (1992) *Hannah Arendt: a reinterpretation of her political thought*. Cambridge University Press
- Capel, Tara, and Margot Brereton (2023) What is human-centered about human-centered AI? A map of the research landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–23. CHI '23. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3544548.3580959>.
- Cortois L (2017) The myth of individualism: from individualisation to a cultural sociology of individualism. *Eur J Cult Polit Sociol* 4(4):407–429. <https://doi.org/10.1080/23254823.2017.1334568>
- Crawford K (2021) *The atlas of AI: power, politics, and the planetary costs of artificial intelligence*. Yale University Press
- Ess C (1999) Cultural attitudes towards technology and communication: new directions of research in computer-mediated communication. *AI & Soc* 13(4):329–340. <https://doi.org/10.1007/BF01205981>
- Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C et al (2018) AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach* 28(4):689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Floridi L, Cowls J (2022) *A unified framework of five principles for AI in society*. Machine Learning and the City. John Wiley & Sons Ltd, USA, pp 535–545
- Gratton P (2021) *The human condition*. In: Gratton P, Sari Y (eds) *The Bloomsbury Companion to Arendt*. Bloomsbury Academic, London New York, pp 213–227
- Hagendorff T (2022) Blind spots in AI ethics. *AI and Ethics* 2(4):851–867. <https://doi.org/10.1007/s43681-021-00122-8>
- Hagendorff T, Bossert LN, Tse YF, Singer P (2023) Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals. *AI and Ethics* 3(3):717–734. <https://doi.org/10.1007/s43681-022-00199-9>
- Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behav Brain Sci* 33(2–3):61–83. <https://doi.org/10.1017/S0140525X0999152X>
- HLEGAI (2019) “Ethics Guidelines for Trustworthy AI.” European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- Irani, Lilly C., and M. Six Silberman (2013) *Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 611–20. CHI '13. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470742>.
- Korenhof P, Blok V, Kloppenburg S (2021) Steering representations—towards a critical understanding of digital twins. *Philos Technol* 34(4):1751–1773. <https://doi.org/10.1007/s13347-021-00484-1>
- Liao QV, Varshney KR (2022) Human-centered explainable AI (XAI): from algorithms to user experiences. *arXiv*. <https://doi.org/10.48550/arXiv.2110.10790>
- Longo A (2024) Reversing the primacy of political action: thinking politics and technology with Arendt. *Arendt Stud* 8:89–113. <https://doi.org/10.5840/arendtstudies20248759>
- Majewski, Taylor (2024) It's Time to Retire the Term 'User.' MIT Technology Review. August 19, 2024. <https://www.technologyreview.com/2024/04/19/1090872/ai-users-people-terms/>.
- Ozmen Garibay O, Winslow B, Andolina S, Antona M, Bodenschatz A, Coursaris C, Falco G et al (2023) Six human-centered Artificial Intelligence grand challenges. *Int J Hum-Comput Interact* 39(3):391–437. <https://doi.org/10.1080/10447318.2022.2153320>
- Place A (2023) *Feminist designer: on the personal and the political in design*. MIT Press
- Plumwood V (2002) *Feminism and the mastery of nature*. Routledge, London. <https://doi.org/10.4324/9780203006757>
- Riedl MO (2019) Human-centered artificial intelligence and machine learning. *arXiv*. <https://doi.org/10.48550/arXiv.1901.11184>
- Rossello D (2022) The animal condition in the human condition: Rethinking Arendt's political action beyond the human species. *Contemporary Political Theory*, 21(2), 219–239. <https://doi.org/10.1057/s41296-021-00495-9>
- Ryan M (2024) We're only human after all: a critique of human-centred AI. *AI Soc*. <https://doi.org/10.1007/s00146-024-01976-2>
- Schmager, Stefan, Ilias Pappas, and Polyxeni Vassilakopoulou (2023) Defining human-centered AI: a comprehensive review of HCAI literature. *MCIS 2023 Proceedings*, January. <https://aisel.aisnet.org/mcis2023/13>.
- Searle, John R (2002) *Collective intentions and actions*. Consciousness and language. Cambridge University Press, Cambridge, pp 90–105
- Shew A (2020) Ableism, technoableism, and future AI. *IEEE Technol Soc Mag* 39(1):40–85. <https://doi.org/10.1109/MTS.2020.2967492>
- Shneiderman B (2020a) Human-centered artificial intelligence: reliable, safe & trustworthy. *Int J Hum Comput Interact* 36(6):495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Shneiderman B (2020b) Human-centered artificial intelligence: three fresh ideas. *AIS Trans Hum Comput Interact* 12(3):109–124. <https://doi.org/10.17705/1thci.00131>
- Singer P, Tse YF (2023) AI ethics: the case for including animals. *AI Ethics* 3(2):539–551. <https://doi.org/10.1007/s43681-022-00187-z>
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum (2019) *Energy and Policy Considerations for Deep Learning in NLP*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, edited by Anna Korhonen, David Traum, and Lluís Màrquez, 3645–50. Florence, Italy: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1355>.
- Taylor RR, O'Dell B, Murphy JW (2024) Human-centric AI: philosophical and community-centric considerations. *AI Soc* 39(5):2417–2424. <https://doi.org/10.1007/s00146-023-01694-1>
- Timmermans J, Blok V (2021) A critical hermeneutic reflection on the paradigm-level assumptions underlying responsible

- innovation. *Synthese* 198(19):4635–4666. <https://doi.org/10.1007/s11229-018-1839-z>
- UNESCO (2022) “Recommendation on the Ethics of Artificial Intelligence.” United Nations Educational, Scientific and Cultural Organization (UNESCO). <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- van Wynsberghe A (2021) Sustainable AI: AI for sustainability and the sustainability of AI. *AI Ethics* 1(3):213–218. <https://doi.org/10.1007/s43681-021-00043-6>
- Villa D (1995) *Arendt and Heidegger: the fate of the political*. Princeton University Press
- von Schomberg L, Blok V (2023) It takes two to tango: toward a political concept of responsible innovation. *J Responsible Innov* 10(1):2264616. <https://doi.org/10.1080/23299460.2023.2264616>
- Waelen RA (2025) Rethinking automation and the future of work with Hannah Arendt. *J Bus Ethics*. <https://doi.org/10.1007/s10551-025-05991-1>
- Xu W (2019) Toward human-centered AI: a perspective from human-computer interaction. *Interactions* 26(4):42–46. <https://doi.org/10.1145/3328485>
- Xu W, Dainoff MJ, Ge L, Gao Z (2023) Transitioning to human interaction with AI systems: new challenges and opportunities for HCI professionals to enable human-centered AI. *Int J Hum Comput Interact* 39(3):494–518. <https://doi.org/10.1080/10447318.2022.2041900>
- Zahavi D (2021) We in me or me in we? Collective intentionality and selfhood. *J Soc Ontol* 7(1):1–20. <https://doi.org/10.1515/jso-2020-0076>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.