



Human-centric AI: philosophical and community-centric considerations

Randon R. Taylor¹ · Bessie O'Dell² · John W. Murphy³

Received: 2 September 2022 / Accepted: 12 May 2023
© The Author(s) 2023

Abstract

This article provides a course of correction in the discourse surrounding human-centric AI by elucidating the philosophical underpinning that serves to create a view that AI is divorced from human-centric values. Next, we espouse the need to explicitly designate stakeholder- or community-centric values which are needed to resolve the issue of alignment. To achieve this, we present two frameworks, Ubuntu and maximum feasible participation. Finally, we demonstrate how employing the aforementioned frameworks in AI can benefit society by flattening the current top-down social hierarchies as AI is currently being utilized. Implications are discussed.

Keywords Dualism · AI · Human-centric · Ubuntu · Maximum feasible participation · Community-centric

1 Introduction

Artificial Intelligence (AI) is a term used to refer to the 'science and engineering of making intelligent machines,' generally by using a computer to model intelligent behavior with minimal intervention from humans (McCarthy 2007; Hamet and Tremblay 2017). Since its inception at a meeting of minds in Dartmouth College in 1956, AI has now developed such that it is being successfully deployed across a variety of domains across the globe. This ranges from AI applications in healthcare to fields such as education, engineering, economics, and finance (see, e.g., Panesar 2019; Luckin et al. 2016; Gogas and Papadimitrou 2021; Mhlanga 2020). Across these fields of application, the intended use of the algorithmic model must be defined by its developer to

apply, develop, validate and deploy the software or machine being used (Bitterman et al. 2020).

The rationale behind this need for definition and specificity can be thought of as twofold: 1. to ensure that AI is fit for purpose by achieving a desired and intended outcome. For example, an AI-based medical device being used to diagnose a specific condition; and 2. to avoid unintended harm to end users or future harm to others. One important method to ensure that AI supports such reasoning is by employing a "human-centric" or "human-centered" approach. Human-centered AI is a movement by AI developers and academics, among others who aim to narrow the gap between the values of the AI engineer(s), the people who will use it or end-users, and anyone else who might be impacted by the AI. Essentially, the main objective is to align the values between developers, who typically center algorithms, and end-users to prioritize human outcomes (Shneiderman 2022). The aims of such efforts are to mitigate potential harm and to ultimately promote human well-being (Stanford Institute for Human-Centered Artificial Intelligence, n.d.) Although this is not a universally held goal, some assert that developing human-centric AI is necessary for society's long-term stability, despite many challenges (Bryson and Theodorou 2019; see also Lukowicz 2019). Moreover, there are some who question which humans are being centered and whose values are being emphasized, particularly given the legacy of the West's dehumanization and epistemological dominance of non-Western societies (Mhlambi 2020).

✉ Randon R. Taylor
r.r.taylor@uva.nl

Bessie O'Dell
bessie.odell@kellogg.ox.ac.uk

John W. Murphy
j.murphy@miami.edu

¹ Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands

² Department of Psychiatry, University of Oxford, Oxford, England

³ Department of Sociology, University of Miami, Miami, FL, USA

Despite this, given the increasing pivot toward human-centered AI, the objectives of this paper are twofold: (1) we maintain that AI has always been human-centered, and we aim to highlight how certain philosophical perspectives, such as dualism, has created the illusion of AI operating independently of human values. (2) Next, we introduce two alternative frameworks, namely Ubuntu and Maximum Feasible Participation (MFP), that can better align the objectives of human-centered AI with a community-based or participatory perspective. We conclude by emphasizing the key points, discussing implications, and offering directions for future research.

1.1 Becoming human-centric

There is much discussion today about AI becoming human-centric. The thrust of this exchange about the importance of human-centricity is the algorithms that underpin AI have drifted from human control and do not reflect the values of their users. Brian Christian (2020) refers to this issue as the “alignment problem.” As a result, critics are calling for human values to undergird the development of AI-based technology (Dhanrajani 2018; Guszcza 2018; Xu 2019). Accordingly, algorithms must be more closely connected to everyday values, beliefs, and commitments, if developers seek to mitigate potential harm, such as the alienation expressed by workers in several industries (Vicent 2019; Shneidermann 2022).

Yet the disconnect between this AI-based technology and end users is beginning to cause personal and social problems. For instance, not long after Microsoft deployed its AI chatbot, Tay, on Twitter, it quickly began hurling derogatory insults while praising Hitler on the public, social platform (Matyszczyk 2016). Most recently, Meta, previously known as Facebook, deployed an AI-powered chatbot, and it immediately announced that Donald Trump was still president, the election had been stolen, and that it had an Asian wife and watched anime (Thorbecke 2022). While many still see AI as a rational and reliable technology for human development, the root of the misalignment argument must be understood and underscored.

How has this separation occurred? The root of this drift away from human control is an issue as old as Plato (Grayling 2019). That is, throughout the Western tradition, philosophers have sought a base of knowledge and ethics that avoids the contingencies that are a part of daily life. To accomplish this, an escape from the everyday, subjective world is required to gain insight. For example, the Absolute Truth, a fixed reality where facts lay uncontaminated by humans (AllAboutPhilosophy.org n.d.), has been invoked to supply this foundation. The assumption is that once this pristine referent is discovered, real, unaltered knowledge and sound ethical principles become achievable.

This achievement, an objective reality where facts can be retrieved, became much simpler with the theoretical maneuver made by Descartes around 1600 (Bordo 1987). Rather than speculate about unknown and absolute foundations, he declared that subjectivity (mind) could be divorced from objectivity (physical reality). Later on, the fact-value and mind–body distinctions would become popular. In each case, the idea is that particular knowledge can be separated from subjectivity or human contamination and treated as objective and universal. According to Descartes, this distinction is possible and necessary to secure reliable knowledge and unambiguous ethical standards.

In philosophical terminology, this separation of facts from values, or subjectivity from objectivity, is referred to as dualism (Robinson 2018). Dualism, accordingly, is the maneuver that allows algorithms, and AI, to appear to be autonomous because of their status as strictly logical and objective. Within the context of dualism, AI appears to gain autonomy and, after some time, needs an orientation supplied by human values. However, as long as dualism is in play, this ambivalence is difficult to resolve, especially when autonomy, and, at the extreme end, “sentience” (Tiku 2022), is thought to be the strength of AI. That is, as AI-guided technology begins to appear self-directed, and even improve on human traits, interference by users would only compromise this technology and must be restricted.

Contemporary philosophy has a solution to such philosophical dissonance. Through a rejection of dualism, the difficulty of introducing human values in the face of ‘objectivity’ is overcome. For instance, while paying tribute to phenomenology, many writers claim that all knowledge originates in the lifeworld, or the everyday world of human experience (Bakewell 2016). Consequently, nothing is devoid of human values, even AI and associated technologies. When filtered through the lens of the lifeworld, AI is revealed to have a human base that is thoroughly mediated by human values.

Edmund Husserl, for example, undercuts the dualism linked to Descartes with a simple phrase. Husserl (1964) asserts that consciousness is always directed towards an object, which he terms “intentionality”. As a result, the dichotomy between objectivity and subjectivity is dissolved, as both are united within the conscious experience. In acknowledging the role of intentionality, the spheres of objectivity and subjectivity converge into conscious action. Therefore, the lifeworld is the reality that is formed through intentional acts, according to Husserl (1970). Additionally, this world is an outgrowth of human action and is replete with values and meanings that are situational. Consequently, facts are never objective, but are rather subjective and should be interpreted based on the intentions of social actors, as phenomenologists argue.

In terms of AI, the key implication is that AI is never autonomous and disconnected from human values. Indeed, algorithms are a product of conscious activity and carry the standpoints that accompany this connection. As a result, AI can be treated as a mode of human expression, rather than a technology that relieves humans of their total involvement. In the absence of dualism, a new relationship is established between AI and users that enables them to direct or redirect this technology at any time.

The problem, however, is that some phenomena strive to divert attention away from this connection in the hopes of appearing to be objective, unbiased, and scientific. Given the lifeworld perspective, algorithms are not autonomous and universal but constitute a worldview with unique traits and values. Nonetheless, the following constitute some of the assumptions associated with a dualism that reinforce the illusion of algorithmic autonomy:

1. Dualism assumes that facts are empirical and are associated with the empirical features of behavior or events.
2. Dualism treats facts as concrete, measurable data points. In this regard, Lyotard (1984) comments that the current period is the age of the information “bit,” with the prevailing belief that reliable knowledge comes in the form of neatly packaged pieces of data.
3. Dualism assumes that facts are causally linked, meaning that a natural association exists between them. For example, if A happens, then B is likely to happen as a result.
4. Dualism assumes that the relationships between facts are entirely logical, with a precise and discoverable relationship between each one.

These themes constitute the backdrop of algorithms while providing this technology with the appearance of autonomy and engendering an illusion of untrammelled rationality. In other words, this background philosophy frames technology in a particular way that directs attention away from any human connection. However, considering the lifeworld, this externality is no longer justified.

Three ideas are particularly noteworthy in this realization. First, algorithms are never divorced from human reach. Second, AI always has a value orientation that underpins any technical operations. Third, this AI-based technology can be guided by many values that do not attempt to hide human presence. The point is that although a particular philosophical maneuver (dualism) and practice (technical focus) strive to hide the connection of algorithms to human action, this association is at the core of their creation and use.

The moral of this assessment is that no one should be striving to make algorithms humane given that human values are already in action. Making these devices human-centric, accordingly, does not involve a monumental discovery. What

is involved, instead, is a decision to make this technology less alienating to stakeholders and community members. That is, persons must explicitly decide to make AI reflect what they desire, rather than deny their presence. Next, we provide two alternative perspectives that hold the potential to reduce algorithmic alienation, namely Ubuntu and maximum feasible participation, and can adequately support community and stakeholder well-being.

1.2 Moving from the implicit to the explicit: the need for community-centric values

What motivated our elucidation of the human values that underlie AI is a call to explicitly put forth images of AI that are beneficial that consider the well-being of all stakeholders and communities. By doing so, we can support the development of AI that aligns with ethical and moral principles and promotes the greater good of society. One such approach is seeing AI from the perspective of Ubuntu, a community-centered framework. Another is maximum feasible participation, a framework that demands the inclusion of community members or stakeholders to a significant extent in the development and deployment of the AI lifecycle. These two perspectives can be advantageous for both developers and community members.

1.3 Ubuntu

Ubuntu¹ is an African philosophy and a social ontology that provides guideposts for how one relates to other human beings. Originating from the Nguni and Bantu language families (Mugumbate and Nyanguru 2013), it is a form of humanism that elevates a constant concern for the collective, community, or stakeholders as well as highlighting the intersubjective nature of all community members’ lives and outcomes. Essentially, an Ubuntu outlook acknowledges that all lives are, to a great extent, entangled and the behavior of one individual has the ability to influence the lives of all community members or stakeholders. In other words, the fates of all community members are linked (Dawson 1995).

Ubuntu, as a way of viewing social relationships, is a distinct worldview that prioritizes community interrelatedness over a view that promotes individuals as atomized. For instance, an axiom most commonly attributed to the Ubuntu philosophy is “I am because we are.” This declaration epitomizes the core values underpinning the Ubuntu framework. These values relate to sharing, solidarity, humanness through

¹ Ubuntu is a philosophy not to be confused with the well-known computer software operating system. However, the tenets and values of the software are derived from the Ubuntu social philosophy, e.g. open source access for all.

participation, and the acknowledgment that individuals of high rank in society are positioned only with the support of community members or stakeholders (Mugumbate and Nyanguru 2013). In consideration of artificial intelligence, Ubuntu recognizes that AI is only possible by extracting training data from the community.

A growing number of scholars are arguing for the usage of an Ubuntu framework in the approach to AI development and deployment. For instance, Langat et al. (2020) and van Norren (2021) consider the ethical implications of AI and the advantages of having an Ubuntu worldview. The scholars contend that when facing privacy issues, an Ubuntu framework would prioritize transparency to community members or all stakeholders. Moreover, the scholars argue that an Ubuntu lens would elevate community participation and democratization in matters of algorithmic decision-making. Gwagwa et al. (2022) maintain that an Ubuntu approach and its consideration of the impact of community members in AI work could benefit not only African AI developers but have universal benefits as it could reduce issues of inclusivity and diversity by its inherent compelling need to embrace all community members or stakeholders. Adams (2021) maintains that Ubuntu can be a valuable philosophy in the struggle to decolonize Western-dominant AI. Mhlambi (2020) asserts that Ubuntu, by enabling a view of seeing individuals as communal people, will ultimately promote community well-being and social solidarity. Black (2018), in his thesis dealing with the urgent need for Ubuntu within AI research, maintains that, given the acceptance of an Ubuntu philosophical outlook, "An AI researcher... accepts the responsibility they have to develop AI to the benefit of all people and not to the benefit of some while harming others" (2018, p. 26). Accordingly, Ubuntu exalts community elevation over a type of techno-marginalization.

Consider how Ubuntu contrasts with the popularly expressed axiom in the West "I think therefore I am," which elevates the individual or entity over community well-being. This is underscored by Bakewell's (2016) assessment of existentialism, a European philosophy commonly associated with the French philosopher Jean-Paul Sartre. Bakewell's description of existentialism in her book, *At the existentialist café: freedom, being and apricot cocktails*, describes it as an anxiety that extends no further than the self. This is a glaring contrast from an Ubuntu outlook that concerns community well-being and interconnectedness.

In the case of artificial intelligence and its expanding utility in the West, such a narrow concern for the "I" or the entity in deploying AI would primarily seek to bolster the undergirding neoliberal ambitions, while trivializing the potential long-term impact on community members and stakeholders. A consequence of this would reinforce the marginalization of vulnerable communities by alienating them from the developmental process of AI technology

and exacerbating inequities existing at social, political, and economic levels. Without healthier, community-centric approaches like Ubuntu, organizations will continuously have to focus on mitigating the harm associated with philosophical frameworks that only prioritize the "needs of the business," leading to ongoing algorithmic casualties.

Currently, the evidence of harm caused by AI is increasing, as both private and public organizations are implementing AI in their operations. For example, Obermeyer et al. (2019) found issues within algorithms developed by a major health insurance company that used healthcare costs as a predictor of stakeholder health. Considering the training data, developers found that white Americans' health expenditures were significantly more costly than African Americans' health expenditures. Subsequently and logically, it was inferred that white Americans were unhealthier than African Americans and needed more healthcare intervention. On the surface, the premise seems reasonable. However, this proposition does not hold for African Americans given it did not account for the complex historical and contemporary racialized relationship between African Americans, the United States (US), and the healthcare industry.

Indeed, in the US, African Americans, compared to white Americans, routinely endure disproportionate levels of morbidity while having less access to healthcare due to the various ways that systemic racism divorces stigmatized social groups from health-promoting resources (Paradies 2006). Furthermore, African Americans are more distrustful of the medical industry owing to the centuries-long history of ghastly, unethical medical experimentation by healthcare workers, e.g., James Marion Sims (Washington 2006). As a result, African Americans have disparate access to healthcare and are less likely to trust the healthcare industry, leading to fewer medical encounters and lower healthcare spending compared to white Americans. Such a social dynamic produces a misleading pattern in training data.

In this case, appropriate consultation with subject matter experts, i.e., stakeholders, who study health inequity and inequality, could have provided rich insight to the algorithmic developers. Furthermore, by failing to acknowledge that data are mere products of social processes and not objective representations of social reality, the AI developers have inadvertently exacerbated racial disparities in healthcare, further harming a vulnerable community. An Ubuntu framework in this case would be valuable due to its preoccupation with including relevant stakeholders in algorithmic decision-making with the aim of expanding equitable access to healthcare. Such a community-centric approach will undoubtedly foster community well-being. Maximum feasible participation

is another framework that can assist in aligning with the goals of human-centric AI.

1.4 Maximum feasible participation

To date, maximum feasible participation (MFP) has not been proposed in the literature as a framework for AI. MFP is a political phrase that comes from a provision in the US Economic Opportunity Act of 1964 during the tenure of President Lyndon B. Johnson (Rubin 1969). At the time, the US faced unprecedented poverty rates which compelled the Johnson administration to formally declare a “War on Poverty,” the unofficial name for the Economic Opportunity Act.

Under the auspices of The War on Poverty legislation, a more robust social safety net was established to help lift US citizens out of impoverished conditions. An example of intervention was the quick creation of community action programs in the US. These programs aimed to tackle social issues such as health, education, and poverty, by infusing government funding and resources in needy communities. However, a significant requirement within the Economic Opportunity Act's language became notorious for igniting ongoing conflicts between major political parties and their ideological supporters. The requirement mandated that residents or groups predominantly comprising of poor individuals, who would be served and impacted by the community action programs, must participate in the governance and decision-making of the community action programs to the greatest extent possible, hence “[maximum feasible participation](#)” (Rubin 1969).

The broad goal of this tactic was to balance the conventional top-down, paternalistic nature of government funding programs while empowering community members to exercise greater self-determination by participating in the design of programs that would have a direct impact on their community's outcomes. Essentially, rather than public and private officials determining how government funds would be distributed to the community members, the maximum feasible participatory clause required laymen to participate in the decision-making process in a collaborative effort with officials—at the time, a novel policy within the US political context. Accordingly, Rubin (1969) states that “community action [or participation] [was viewed as]...a vehicle for community development” (p. 18).

Until a 1967 amendment was introduced to the MFP mandate, there was an ongoing debate about how it should be implemented, leading to constant confusion. The amendment, which was a significant development, specified that at least one-third of board members of community action programs needed to be community members, while public officials and industry leaders were limited to a maximum of one-third of board membership. This stands in stark contrast

to vague ideas of community participation or democratization that lack clear and detailed operationalization.

To illustrate the potential challenges that a maximum feasible participation framework in the field of AI could face, it's worth noting that many city and organizational leaders pushed back against this approach. They cited reasons such as a lack of expertise among community members, rumors that this strategy would enable them to undermine professional authority, and concerns that it could lead to revolution (Rubin 1969). Rubin (1969) states, “welfare agencies and politicians made massive efforts to retain their doctrine, dogma, and power, while the leaders of the poor did, indeed, use federal funds to try to force institutional change.”

While the maximum feasible participation clause would go on to see thousands of community members and stakeholders employed by community action programs, it would never reach its full potential given the constant struggle for power by city officials and organizational leaders. However, such a mandate that sought inclusivity in the decision-making process regarding government funding distribution can provide relevant guidance in efforts to democratize and align with the goals of human-centric AI.

1.5 The social benefit of a community-centric approach

A community- or stakeholder-centric approach from the perspective of Ubuntu or maximum feasible participation provides many ways that can benefit organizations deploying AI. In the legal realm, the inclusion of stakeholders in the total lifecycle of AI would ease the number of class action and personal litigation suits against organizations deploying AI models onto citizens. According to the Ethical Tech Initiative of DC (2021), an organization that maintains a database of AI litigation cases in the US, there have been over 30 lawsuits related to the use of AI technology by various organizations. The most commonly cited themes in these lawsuits, as reported by plaintiffs, were concerns about transparency, lack of human, scholarly, and expert review, issues related to the reliability of the technology, and the improper use of variables such as gender and race in algorithmic decision-making. These concerns have been raised by multiple plaintiffs across multiple lawsuits, indicating a significant issue with the implementation of AI technology in various industries.

In one class action lawsuit, *Bauserman v Unemployment Insurance Agency* (2015), the Michigan Unemployment Insurance Agency, through its use of algorithmic decision-making, falsely accused 40,000 residents of insurance fraud, which led to civil penalties and tax refund seizures of the citizens (Ethical Tech Initiative of DC 2021). As a result, the state was forced to distribute tens of thousands of monetary waivers to the affected (Pluta 2022). The lack of stakeholder

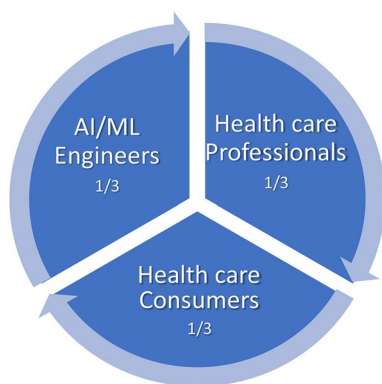


Fig. 1 Maximum feasible participation model: 1/3rd minimum participation from stakeholders (consumers and professionals) with a maximum of 1/3rd participation from AI/ML engineers (original source)

participation, that is, human review, in the lifecycle of the algorithm is cited as one of the contributing reasons for the flawed automated system. In this particular situation, implementing an MFP approach could have reduced the chances of Michigan residents from losing their insurance benefits and suffering negative psychological consequences. Unfortunately, it is the taxpayers who, in the long run, fund state inconsideration and lack of foresight.

Considering AI and medicine, an Ubuntu approach would seek to, among other things, cultivate patient and public collaborations in the active collection of data, sharing of methodologies, and the dissemination of research results to the public. Each party would understand that such a way of relating to each other would be mutually beneficial. In the UK, McKay et al. (2022) shed light on three primary efforts made in the medical and AI industry to try to actively involve the public in the governance of AI, namely “lay representation on data access committees, patient and public involvement groups, and citizen forums.” Although scholars argue that there is a need for improved integration of these methods, the efforts in the UK do provide a framework for organizations looking to develop patient and public partnerships.

In public health, an explicit stakeholder/community-centric approach would align AI values with more upstream health interventions—the efforts made to prevent individuals from acquiring disease (Shultz et al. 2019). Through a circular relationship as seen in Fig. 1, such a model would not only involve stakeholders in the research, investigation, development, and deployment process of AI, but it could potentially support the training and employment of stakeholders and community members, which ultimately increases equity and subsequently reduces inequality.

Additionally, incessant engagement with community members and stakeholders could theoretically provide increased access to health data for researchers. This would

potentially enable quicker modalities to prevent and reduce disease as well as provide longitudinal data, rather than reliance on cross-sectional data. Although AI is predominantly deployed at midstream and downstream preventive health measures; that is, the efforts designed to mitigate the impact of a disease or injury and the efforts that reduce the impact of an ongoing injury or disease through clinical intervention, respectively (Shultz et al. 2019), a stakeholder and community-centric AI would bolster public health interventions.

2 Conclusion

This paper contributes to the literature in two primary ways: first, we provide a response to recent suggestions that AI should become more human-centric. Specifically, we clarify that AI is already human-centered. However, the philosophy of dualism, which holds that human experience is separate from objective reality, creates an illusion of a divide between human actions and AI decisions. This leads some to believe that AI operates independently of human influence. In reality, AI technology reflects human activity, as it is designed by humans in pursuit of human-derived goals. Second, although proponents of human-centered AI aim to promote human well-being, we present two alternative, community-centered frameworks, Ubuntu and MFP, that can better align AI with the aspirations of those seeking to democratize and center ethical values and thereby support community well-being. Ubuntu and MFP are two frameworks that provide clear guidance on how AI development and deployment can become more inclusive, rather than alienating, in various industries. In conclusion, it is important to highlight that when organizations choose to develop and deploy AI, whether their focus is on maximizing profits (i.e., their “bottom line”) or promoting community well-being, each decision reflects a specific set of human values. Yet, it is crucial to note that this is not a matter of misalignment. Ultimately, the choice of which values to prioritize when deploying AI depends on the organization's goals and priorities, as well as the ethical considerations that guide their decision-making process.

Acknowledgements This paper wishes to express gratitude for the support of the AI for Good Foundation (<https://ai4good.org/>) in their tireless effort to make the world a better place. Specifically, we thank James Hodson, Tia Christopher, Lindsey Asis, and Senthujan Senkai-ahliyan at AI for Good for their guidance.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

- Adams R. (2021) Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews* 46(1-2) 176–197. [10.1080/03080188.2020.1840225](https://doi.org/10.1080/03080188.2020.1840225)
- Bakewell S (2016) *At the existentialist café: freedom, being and apricot cocktails*. Chatto & Windus, London
- Bitterman DS, Aerts HJWL, Mak RH (2020) Approaching autonomy in medical artificial intelligence. *Lancet Digital Health* 2(9):e447–e449. [https://doi.org/10.1016/S2589-7500\(20\)30187-4](https://doi.org/10.1016/S2589-7500(20)30187-4)
- Black M. (2018). “Artificial intelligence research: a discipline in need of Ubuntu.” Not Published.
- Bordo S (1987) *The flight to objectivity: essays on cartesianism and culture*. SUNY Press, Albany, NY
- Bryson JJ, Theodorou A (2019) How society can maintain human-centric artificial intelligence. In: Toivonen M, Saari E (eds) *Human-centered digitalization and services, translational systems sciences*. Springer, Singapore, pp 305–323
- Christian B (2020) *The alignment problem: machine learning and human values*. Norton, NY
- Dawson MC (1995) *Behind the mule: race and class in African-American politics*. Princeton University Press, Princeton, NJ
- Dhanrajani S (2018). “3 Ways to Human Centric AI.” *Forbes*, Dec. 12, 2018. [Forbes.com/sites/cognitive world/2018/12/12/3-ways-to-human-centric-ai/?sh=4bb620324a38](https://www.forbes.com/sites/cognitive-world/2018/12/12/3-ways-to-human-centric-ai/?sh=4bb620324a38). Retrieved June 27, 2021.
- Ethical tech initiative of DC. (2021). “AI Litigation database.” ethical tech initiative of DC. Retrieved August 27, 2022, Accessed 28-Aug-2022 (<https://blogs.gwu.edu/law-eti/ai-litigation-database/>).
- Gogas P, Papadimitriou T (2021) Machine learning in economics and finance. *Comput Econ* 57(1):1–4. <https://doi.org/10.1007/s10614-021-10094-w>
- Grayling AC (2019) *The history of philosophy*. Penguin Press, NY
- Guszcza J. 2018. “AI needs human-centered design.” *Wired*, May 1.
- Gwagwa A, Kazim E, Hilliard A (2022) The role of the African value of Ubuntu in global ai inclusion discourse: a normative ethics perspective. *Patterns*. <https://doi.org/10.1016/j.patter.2022.100462>
- Hamet P, Tremblay J (2017) Artificial intelligence in medicine. *Metabolism* 69:336–40. <https://doi.org/10.1016/j.metabol.2017.01.011>
- Husserl E (1964) *The Paris lectures*. Nijhoff, The Hague
- Husserl E (1970) *The crisis in european science and transcendental phenomenology*. Northwestern University Press, Evanston, IL
- Langat SK, Mwakio PM, Ayuku D (2020) “How Africa should engage Ubuntu ethics and artificial intelligence” edited by R. Hajare *Journal of Public Health International* 2(4):20–25. <https://doi.org/10.14302/issn.2641-4538.jphi-20-3427>
- Luckin R, Holmes W, Griffiths M, Forcier LB (2016) *Intelligence unleashed: an argument for AI in education*. Pearson, London
- Lukowicz P (2019) The challenge of human centric AI. *Digitale Welt* 3(4):9–10
- Lytard J-F (1984) *The postmodern condition: a report on knowledge*. University of Minnesota Press, Minneapolis
- Matyszczak C. (2016). “Thanks, Twitter. You Turned Microsoft’s AI Teen into a Horny Racist.” CNET. Retrieved August 19, 2022 (<https://www.cnet.com/culture/twitter-turned-microsoft-ai-teen-into-horny-foul-mouthed-sex-bot-racist/>).
- McCarthy J (2007). What is artificial intelligence? Available at: <http://www-formal.stanford.edu/jmc/whatisai.pdf> [Accessed 8 Aug 2022]
- McKay F, Williams BJ, Prestwich G, Treanor D, Hallowell N (2022) Public governance of medical artificial intelligence research in the UK: an Integrated multi-scale model. *Res Involv Engagem* 8(1):21. <https://doi.org/10.1186/s40900-022-00357-7>
- Mhlanga D (2020) Industry 4.0 in Finance: the impact of artificial intelligence (AI) on digital financial inclusion. *Int J Financ Stud* 8(3):45. <https://doi.org/10.3390/ijfs8030045>
- Mhlambi S. (2020). “From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance.” Carr Center Discussion Paper Series (2020–009).
- Mugumbate J, Nyanguru A (2013) Exploring African philosophy: the value of Ubuntu in social work. *Afr J Soc Work* 3(1):82–100
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453. <https://doi.org/10.1126/science.aax2342>
- Paradies Y (2006) A systematic review of empirical research on self-reported racism and health. *Int J Epidemiol* 35(4):888–901. <https://doi.org/10.1093/ije/dyl056>
- Panesar A. (2019). *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Berkeley, CA: Apress.
- Pluta R. (2022). “State supreme court says victims of jobless benefits system failure can seek damages.” *Michigan Radio*. Retrieved August 27, 2022, Accessed 28-Aug-2022 (<https://www.michiganradio.org/criminal-justice-legal-system/2022-07-26/state-supreme-court-says-victims-of-jobless-benefits-system-failure-can-see-damages>).
- Robinson WS (2018) Dualism. In: Gennaro RJ (ed) *The Routledge Handbook of Consciousness*. Routledge, London, pp 51–63
- Rubin LB (1969) Maximum feasible participation: the origins, implications, and present status. *Ann Am Acad Pol Soc Sci* 385(1):14–29. <https://doi.org/10.1177/000271626938500103>
- Shneidermann B (2022) *Human-Centric AI*. Oxford University Press, Oxford, UK
- Shultz JM, Sullivan LM, Galea S (2019) *Public health: an introduction to the science and practice of population health*, 1st edn. Springer Publishing Company, New York
- Thorbecke C. (2022). “It Didn’t Take Long for Meta’s New Chatbot to Say Something Offensive | CNN Business.” *CNN*. Retrieved August 18, 2022 (<https://www.cnn.com/2022/08/11/tech/meta-chatbot-blenderbot/index.html>).
- Tiku N. (2022). “The Google Engineer Who Thinks the Company’s AI Has Come to Life.” *Washington Post*. Retrieved August 10, 2022 (<https://www.washingtonpost.com/technology/2022/06/11/google-ai-lambda-blake-lemoine/>).

- van Norren D. (2021). "The Ethics of Artificial Intelligence through the Lens of Ubuntu." Africa Knows Conference.
- Vicent J (2019). "Amazon's Latest Warehouse Machine Demonstrates the Slow Amazon-warehouse-workers-replaced-technology-jobs. Retrieved on December
- Washington HA. (2006). Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present. Doubleday.
- Xu W (2019) Toward human centric AI: a perspective from human-computer interaction. Interactions. <https://doi.org/10.1145/3328485>

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