



The ethics at the intersection of artificial intelligence and transhumanism: a personhood-based approach

Amara Esther Chimakonam 

Centre for African Phenomenology, University of Fort Hare, Alice, South Africa
Email: Amaraesther35@gmail.com

Received: 08 November 2023; **Revised:** 18 March 2024; **Accepted:** 08 April 2024

Keywords: African ethics; artificial intelligence; moral enhancement; personhood-based theory; transhumanism

Abstract

In this article, I will consider the moral issues that might arise from the possibility of creating more complex and sophisticated autonomous intelligent machines or simply artificial intelligence (AI) that would have the human capacity for moral reasoning, judgment, and decision-making, and (the possibility) of humans enhancing their moral capacities beyond what is considered normal for humanity. These two possibilities raise an urgency for ethical principles that could be used to analyze the moral consequences of the intersection of AI and transhumanism. In this article, I deploy personhood-based relational ethics grounded on Afro-communitarianism as an African ethical framework to evaluate some of the moral problems at the intersection of AI and transhumanism. In doing so, I will propose some Afro-ethical principles for research and policy development in AI and transhumanism.

Policy Significance Statement

So far, ethical guidelines for artificial intelligence (AI) largely come from the West, such as Europe and North America and are mainly drawn from the Western ethical tradition. However, Africa has played little role in designing algorithms and drawing up ethical guidelines from African ethics for AI development, programming, and application. To fill this gap, this article draws from African ethics, particularly, personhood-based relational ethics, to articulate Afro-ethical principles for AI and transhumanism research. These Afro-ethical principles, also identified as the 3-I, are *inter-relationality*, *inter-contextuality*, and *inter-complementarity*.

1. Introduction

In this essay, I aim to critically engage with the moral issues that arise from the intersection of artificial intelligence (AI) and transhumanism. This intersection invokes a threshold at which AI might begin to simulate (or even surpass) human-level intelligence with the capacities for moral reasoning, judgments and decision-making, and humans cease to be humans and become ultraintelligent minds with supermoral capacities. I will argue that this intersection is likely to pose two moral problems, namely the technologization of humans and AI dominance. Although the technologization of humanity through radical AI-based moral enhancement would result in humans becoming intelligent moral machines (IMMs), AI dominance would result in supermoral machines that might treat humans as moral patients. To overcome



these problems, I will employ personhood-based relational ethics grounded on Afro-communitarianism as a framework for building an ethical AI that would align with African moral values such as complementary relationship. Building on personhood-based theory, I demonstrate that its main principle and two exception clauses, which emphasize mutual and nonmutual relationships, could be strategic in developing an ethical template for AI and transhumanism research and policy.

Two things make this inquiry novel and relevant. First, scant research attention has been paid to the ethical consequences of the intersection of AI and transhumanism. Second, scholars acknowledge that there is little cultural and ethical diversity in AI studies and even in transhumanism. I plan to cover these two underexplored perspectives in this inquiry. For the second aspect, I will explore and deploy an African philosophical dimension called personhood-based relational ethics. Scholars like Floridi and Cowsls (2019), Thilo Hagendorff (2020), Syed Mustafa Ali (2021), and Jan-Christoph Heilinger (2022) have shown that much of the discussions on AI and transhumanism center on Western ethical perspectives. For instance, while Ali (2021, 169), in his essay, “Transhumanism And/As Whiteness”, shows that the discourse of transhumanism projects “[M]an” as white, male, European and anthropocentric,” Hagendorff (2020, 105), in his “The Ethics of AI Ethics: An Evaluation of Guidelines,” points out that the field of AI is predominantly dominated by “white men,” making the field to lack diversity. In addition, Heilinger (2022, 4) writes that “[e]thical reflections and arguments in scholarly publications as well as in policy documents and tech industry guidelines... mirror the three different normative theories that shape the tradition of Western moral philosophy: consequentialism, deontology and virtue ethics”. In other words, the ethics of AI and transhumanism are dominated by Western ethical principles, while ethical perspectives from Africa are largely ignored (see UNESCO, 2021). However, the little literature that explores the African ethical dimension of AI and transhumanism often do so from the Ubuntu standpoint (see van Norren, 2023). All this shows that there is a need to broaden the discourse of ethics of AI and transhumanism since different ethical systems will result in different moral principles for the programming and application of AI. Personhood-based relational ethics offers a novel approach to AI and transhumanism from an African perspective, specifically an Afro-communitarian standpoint.

I divide this essay into four sections. I briefly conceptualize AI ethics and transhumanism and show the intersection of AI and transhumanism in the first section. In the second and third sections, I consider some of the moral issues that the intersection of AI and transhumanism portends. I articulate Afro-ethical principles from personhood-based relational ethics for AI and transhumanism research and policy development in the fourth section.

2. An overview of AI ethics and transhumanism

In this section, I will conceptualize AI ethics and transhumanism. Also, I will show the intersection of AI and transhumanism. I will begin with a brief definition of AI. There is no consensus on how AI is to be defined. What some scholars define as AI is human-like intelligence embedded in machines (see McCarthy et al., 1955; Rich, 1983; Liao, 2020). Others deny this conception and claim it is too narrow to capture the many meaningful possibilities of the subject matter (see Russell and Norvig, 2010; Russell, 2016). Some others define AI so loosely to encompass all kinds of machines that it becomes difficult to pin down (see Boddington, 2023; Nyholm and Ruther, 2023), while others define it so strictly to include only those machines equipped with both human cognitive skills and moral capacities that it practically shuts out its many potentials (Haugeland, 1981).

These various definitions of AI show that there are many ways of understanding the term AI. These different definitions have merit insofar as one keeps clear of what one has in mind and the place upon which one is staking one’s claim. For my aim, I define AI as:

technologies that can imitate/simulate intelligent behavior or/and moral capacities such as moral reasoning, judgment, and decision-making; and enhance/augment humans’ intelligence and moral capacities.

This definition covers (a) artificial narrow intelligence, any machine intellect that intelligently reproduces the cognitive performance of humans in a single specific domain; (b) artificial general intelligence (AGI), any machine intellect that exhibits human cognitive skills/and moral capacities in different domains; and (c) superintelligence “any [machine] intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest” (Bostrom, 2014,26).

Currently, we have narrow AI systems, or weak AI, that operate with artificial narrow intelligence because they are designed to perform a particular task, like diagnosing cancer or playing chess. Some scholars, such as Hans Moravec (Moravec, 1988) and Donna Haraway (1991), anticipate the creation of more sophisticated and complex AI systems that will operate with AGI and be capable of performing (or even outperforming) various human intelligible tasks. Such AI systems, or strong AI, are also anticipated to be capable of human-like thought, moral reason, sentience, and consciousness. Other scholars, like Vernor Vinge (1993), Ray Kurzweil (2005), and Nick Bostrom (2014), speculate that strong AI, when sufficiently advanced, could develop an improved version of itself, which could, in turn, create a greater version of itself until we arrive at an intelligence explosion or singularity. However, scholars, including Bostrom, have pointed out that such AI advancement would come at greater “existential risks” to humanity. For example, such superintelligent AI might consider humanity inferior (I will say more in Sections 2 and 3). The challenge before us is how to come up with ethical principles that would ensure that we develop AI systems that would pose minimal risks to humanity and the environment (I will come back to this later in Section 4).

The need for ethics in AI becomes more pressing each day with the continuous advancement of AI systems. The advancement of AI raises many ethical issues. For instance, battlefield lethal autonomous weapons aid military personnel and decrease fatal risks for civilians; however, what happens in cases where such lethal autonomous weapons malfunction? In 2007, an autonomous anti-aircraft cannon malfunctioned, killing nine soldiers and injuring 11 others during a shooting exercise in South Africa (IOL news, 2007). In this case and other similar cases, who will be held morally responsible: the AI, the programmer, or the company? Also, consider the issue of sex robots and how such would impact human sexual relationships or Robo-lawyers and how they would impact the jobs of legal practitioners.

Ethics of AI (or AI ethics)¹ is a relatively new field of study in applied ethics (see Hanna and Kazim, 2021; Waelen, 2022). The field of AI ethics has emerged to investigate the moral issues associated with AI research, creation, and application. The field also aims to provide ethical frameworks for ensuring that AI contributes meaningfully to humanity and promotes social good. I define AI ethics as:

A multidisciplinary study of the moral concerns arising from the development and useful application of AI technologies and the articulation and formulation of moral principles, values, theories, and policies for creating ethically permissible AI.

As a multidisciplinary study, AI ethics combines approaches from different fields of study, such as computer sciences, engineering, informatics, neurosciences, and philosophy, to look at multifaceted ethical issues arising from the advancement of AI technologies and offer myriad solutions to them. This multidisciplinaryity is vital for developing ethically permissible AI, optimizing the beneficial impact of AI technologies for humanity and environmental sustainability, and the meaningful use of these AI technologies. It disallows any one-size-fits-all ethical approach to AI. In addition, it opens up conversations and collaborations among different knowledge domains, which is essential in formulating effective and efficient ethical principles for the design of AI and ensuring that policies are in place to limit the abusive use of AI technologies.

Many scholars have focused on the potential harms of AI, such as privacy, algorithm bias, transparency issues, data problems, and infringement of individual autonomy, inequality, monopoly, surveillance, and manipulation (Hagendorff, 2020, Müller 2022). Others consider issues like creating ethical machine

¹ For a detailed discussion of AI ethics, see Paula Boddington’s *AI Ethics: A Textbook* (2023) and *Towards a Code of Ethics for Artificial Intelligence* (2017)

agents and raising questions of whether autonomous machines should be regarded as moral agents, be held morally responsible for their actions (see Bostrom and Yudkowsky, 2014), and whether machines should be accorded moral status (Gunkel, 2012; Anderson, 2013; Coeckelbergh, 2020). Some others consider the impact of AI on life's meaning, asking whether AI could be employed for meaningful human existence (see Nyholm and Ruther, 2023).

While the ethics of AI is an interesting area of focus, some other scholars have begun to discuss how AI could be used to enhance humans. In the philosophical circle, this discussion is known as transhumanism. Transhumanism² can be defined “broadly as seeking to use the means of science and technology to enhance human capacities radically and to transform their social conditions by transcending the limitations imposed on them by their biology and nature in order to create posthumans” (AE Chimakonam, 2023a, 3). AI could play a major role in gene editing/engineering process aimed at enhancing humans. Transhumanists such as Hans Moravec (1988), Bostrom (2005), Kurzweil (2005), De Grey and Rae (2007), Max More (2013), Natasha Vita-more (2019), Newton Lee (2019), and Stefan L. Sorgner (2022) defend the possibility of creating trans-and-post biological life without the limitations of disease, ageing, suffering, cognitive and moral limitations, and even death.

Transhumanism has its roots in Enlightenment humanism, which emphasizes values like reason, science, progress, the uniqueness of humanity, and self-perfection. Enlightenment humanism promotes traditional means of enhancing humans, such as education and cultural refinement. Although transhumanism promotes these enlightenment humanistic values, it is more radical in its approach to human enhancement. It seeks the evolution of humans beyond their current biological and natural limits. Transhumanism promotes the conscious guiding of evolution to recreate and remold human nature in desirable ways. By extending evolution beyond current humanity through the use of science and technology, transhumanism opens up the opportunity for humans to live healthier and longer and enhance their cognitive and moral capacities.

One of the ways in which transhumanists aim to enhance humans is through radical AI-based moral enhancement. Moral enhancement is defined as the “biomedical and genetic interventions that would directly and radically augment individuals’ moral capacities beyond what is therapeutically necessary and considered normal for humans so that they always act morally and become more virtuous” (AE Chimakonam, 2021a, footnote 2). Proponents of moral enhancement, like Ingmar Persson and Savulescu (2008), Thomas Douglas (2008), David DeGrazia (2014), and Vojin Rakic (2014), seek to use the means of science and technology to radically augment the human capacity for moral reasoning, insight, disposition, desire, behavior, belief, and motivation. There is currently no scientific and technological means of augmenting humans’ moral capacities, but some ethicists are very optimistic that such means will be available soon.

Through advancements in science and technology, transhumanists seek to create a good life and society where humans would live morally, healthier, longer, and happier with fulfilled desires. Most remarkable is their belief that sufficient advancement of AI would increase the likelihood of humans becoming posthumans. Elsewhere, I define posthumans as “ultra-intelligent minds with supermoral capacities who have overcome the biological and natural limitations that confront humans” (Chimakonam, 2023a, 8; see also Bostrom, 2014). Posthumans would possess longer health and life spans, better cognitive and emotional abilities, and greater moral capacities, among others, exceeding that of humans. Transhumanists see the coming of posthumans as both *necessary* and *desirable*. It is necessary because humans merging with and becoming IMMs is an evolutionary imperative and desirable because humans aspire for a good life; it matters little whether such a good life is achieved biologically or technologically.

I believe that the intersection of AI and transhumanism lies in their quest for the technological evolution of humans to IMMs. Elsewhere, I have discussed and engaged with the transhumanists’ idea

² I have engaged extensively with transhumanism in the following works: “Transhumanism in Africa: A Conversation with Ademola Fayemi on His Afrofuturistic Account of Personhood” (2021), “Afro-communitarianism and Transhumanism” (2023) and “Moral Enhancement, Afro-communitarianism and the Superchoice” (Forthcoming), and “God and Transhumanism in the Context of African Philosophy of Religion” (Forthcoming).

of humans' technological evolution into posthumans (Chimakonam, 2021a). I will proceed to map out this intersection thus: With natural evolution, human life through the biological mechanism of the brain emerged, a mechanism sometimes referred to as the mind or consciousness. The brain is a biological configuration with many neurons that process the body's sensory input, and its functions could be artificially understood and duplicated. The human brain and its functions could be duplicated in machine circuitry through the cybernetic means of mind uploading. Mind uploading would allow individuals to "scan" their brain into a "powerful supercomputer;" storing their "entire personality, memory, skills, and history" (Bostrom, 2005, 9; Kurzweil, 2005, 199). The result of such radical AI-based enhancement would be humans becoming IMMs. At the same time, the technological evolution of computers emerged from the first mechanic calculators with faster computing capacity. The exponential growth of this capacity would result in computers processing sensory inputs in identical ways but at far faster speeds than the human brain. At this point, computers would attain human-level intelligence and probably exceed such intelligence. The result, yet again, could be IMMs.

In general, then, the intersection of AI and transhumanism is a crucial threshold at which AI might start to simulate (and even surpass) human-level intelligence with the capacities for moral reasoning, judgments and decision-making, and humans cease to be humans and become ultraintelligent minds with supermoral capacities. Ever since the emergency of the Turing Test³, researchers have been in search of the scientific Holy Grail: getting machines to simulate (and even surpass) human-level intelligence and moral capacities (AI) and getting humans to radically emerge with machines by duplicating the brain functions into a combination of some software and hardware (transhumanism). However, there is doubt whether this search for this Holy Grail will ever be realized, even if not in the way the proponents of AI and transhumanism envisage. Nevertheless, it is a matter of hope to say that this search would yield neither IMMs nor posthumans. In addition, since such hope is very thin, we must take this intersection seriously. Not only because of the possibility of it coming to fruition but also because of the ethical issues it would pose. In the following section, I will analyze some of the moral problems that this intersection of AI and transhumanism presents.

3. The technologization of humanity⁴

In this section and the next, I will draw attention to the possibility of serious moral consequences of the intersection of AI and transhumanism. One of the moral consequences that might arise is the technologization of humanity or what can be called the *AIfication of humans* (i.e., the artificial intelligentification/smartification of humans). Technically speaking, *AIfication is a neologism that refers to the process of making humans artificial moral (intelligent) systems*. This term is used here in the context of the intersection of AI and transhumanism to describe the transformation of humans and machines into supermoral, automated, and connected entities that can gather and exchange data, make decisions, and self-improve to adapt to changing conditions. I argue that the intersection of AI and transhumanism could result in IMMs, thereby redefining what it means to be human. Humans would no longer be those who are subject to cognitive and moral limitations but IMMs-posthumans! They would radicalize what it means to be moral human beings since they would no longer act immorally (see Harris, 2016; AE Chimakonam, 2021a, 2023a). For instance, the posthuman "I¹" would have greater moral capacities and would never have to act immorally, unlike the human "I⁰" that fluctuates between moral and immoral courses of action.

³ In 1950, Alan Turing showed, using the *Turing Test*, that machines act intelligibly if they can, among other things, generate and communicate with language; autonomously perceive, learn and adapt to experience; and sense, reason and act independently. Additionally, the Turing Test defines intelligent machines in terms of their ability to fool a human judge into thinking that they are talking to a person. For example, driverless vehicles function intelligibly without human intervention and Apple's voice assistant, Siri, communicates successfully with the English language. Also, consider OpenAI's Generative Pre-Trained Transformer-4, a large language model that exhibits human-level intelligence by generating texts indistinguishable from human speech.

⁴ Some parts of this section are from my unpublished PhD thesis titled, "Contending with Superchoice in a Transhumanist Future: Is the Normative Conception of Personhood under Threat?" Department of Philosophy, University of Johannesburg, South Africa, with some modifications.

They would eliminate the freedom to choose among alternative moral choices since, through their moral enhancement facilitated by sufficient advancement in AI, they would inevitably behave morally. In my essay, “Afro-communitarianism and Transhumanism” (2023a), I explored the implication of radical AI-based moral enhancement on humans’ moral choices, but I aim to deepen the argument further here.

If the idea of radical AI-based moral enhancement entails that morally enhanced agents inevitably choose the right course of action, it is very difficult to see how they are different from “moral zombies” (see Chimakonam, 2023a, 16–17), for instance, which are radically and biomedically programmed to always act morally without being capable of seeing and considering moral choices. If morally enhanced agents inevitably choose the right course of action, it means that they are not free to choose among alternative moral choices. They would know the right course of action and would have no choice but to choose it. What seems to be crucial for morality is that individuals choose among different moral choices for the right reason, and it is not easy to see what it can mean in the case of radical AI-based moral enhancement. An individual is only responsible for their action when they are free to choose either to do right or wrong. If then morally enhanced agents choose the right course of action, in what way are they free? They seem no freer and more responsible than moral zombies. Moral zombies are not free or responsible for what they do, for their actions are determined. Are not morally enhanced agents similarly radically and biomedically programmed to always act morally? For if they could cease to act in the way thus programmed, that is, to always act morally, they would not always be morally virtuous and so would fail to fulfill the primary condition of being morally enhanced. How, then, can we attribute freedom to choose among alternative moral choices to such morally enhanced agents?

It might be argued that enhancing human beings may not necessarily rule out the possibility of having moral choices. In “Alternate Possibilities and Moral Responsibility” (1969), Harry Frankfurt argues that moral agents need not have alternative choices to choose from before they can be said to have chosen freely and be held morally responsible for their choices. He champions this idea with his famous thought experiment, a variant of which someone, Black, as a Republican, wants Jones, a Democrat, to vote for Donald Trump in the 2020 American presidential election against his preference to vote for Joe Biden. Therefore, Black secretly implants a remote control chip in Jones’ brain that will manipulate him to vote for Trump. Black prefers not to show himself unnecessarily but plans to press the remote only when Jones decides to vote for Biden. On the day of the election, Jones voted for Trump on his own accord even when he could not have done otherwise because of Black’s remote control chip. In this case, Frankfurt claims that Jones would be held morally responsible as long as he performs “the same action” Black demanded of him—“whether he acts on his own or as a result of Black’s intervention”—because lacking alternative moral possibilities is utterly “irrelevant” to his moral actions (Frankfurt, 1969, 836–837).

Although Frankfurt’s thought experiment was directed to the issue of free will and determinism, it has a great implication for the ethical issue of creating IMMs. For example, Persson and Savulescu give a similar example of this thought experiment where a “freaky mechanism” is implanted into the human brain to ensure that one never does an immoral act (Persson and Savulescu, 2012, 114). The freaky mechanism implies that moral enhancement poses no great challenge to moral agents’ ethical choices since they are morally free to act morally and free to even initiate alternative acts but restricted from acting immorally. In essence, human freedom and responsibility tally as long as moral agents act morally.

However, I am skeptical that enhancing humans’ moral capacities would guarantee moral agents’ freedom/responsibility. It can be argued that in Frankfurt’s thought experiment or Persson and Savulescu’s “freaky mechanism,” one would not be free or responsible since such a mechanism would undermine one’s ability to choose between or select among alternative moral choices. We would automatically know what is best on offer, and that is not a process of moral judgment that leads to a choice between moral and immoral actions. A moral agent would be prevented from making a whole lot of other moral choices because they have been habitually conditioned to behave in morally certain ways. They are not morally responsible for not acting immorally because of the freaky machine intervention. Rakic has pointed out that freedom is an essential part of our morality, which is a key element of what makes us human, adds weight to our moral choices and if any freaky mechanism restricts this freedom, we would run the risk of

denying what is vital to humanity and “inflicting serious (if not ultimate) harm upon ourselves” (Rakic, 2014, 248–249).

A significant aspect of our freedom would then be eliminated, and individuals’ ability and freedom to choose among alternative moral choices would be obliterated.⁵ As argued elsewhere, we must not forget that individuals’ freedom of choice covers not only moral choices but immoral choices as well (Chimakonam, 2021a). To eliminate the latter would amount to eliminating, or at least, slashing away half of “responsibility” as a moral concept. John Harris articulated this point in his magisterial book, *How to be Good: The Possibility of Moral Enhancement*, where he points out that “[k]nowledge of the good is sufficient to have stood, but freedom to fall, is all” (Harris, 2016, 60). He also points out that “[w]ithout the freedom to fall, good cannot be a choice and freedom disappears and along with it virtue. There is no virtue in doing what you must” (Harris, 2016, 60). Thus, the AIfication of humanity would eliminate not just the freedom to decide/choose whether or not to act morally, but the freedom to act morally or immorally. The freedom to act is the guarantor of the freedom to decide/choose. In the absence of the former, the latter vanishes. In other words, without the freedom to act, choice/choosing does not exist because action is the manifestation of choice. If one could not act freely, then they have not really chosen.

Persson and Savulescu further their argument with the “God Machine” thought experiment that also assumes, with Frankfurt’s case, that moral agents need not have alternative moral possibilities before they can be said to have acted morally. I will quote them in detail:

The Great Moral Project was completed in 2045. This involved construction of the most powerful, self-learning, self-developing bioquantum computer ever constructed called the God Machine. The God Machine would monitor the thoughts, beliefs, desires and intentions of every human being. It was capable of modifying these within nanoseconds, without the conscious recognition by any human subjects. The God Machine was designed to give human beings near complete freedom. It only ever intervened in human action to prevent great harm, injustice or other deeply immoral behaviour from occurring. For example, murder of innocent people no longer occurred. As soon as a person formed the intention to murder, and it became inevitable that this person would act to kill, the God Machine would intervene. The would-be murderer could ‘change his mind.’ The God Machine would not intervene in trivial immoral acts, like minor instances of lying or cheating. It was only when a threshold insult to some sentient being’s interests was crossed would the God Machine exercise its almighty power. (Savulescu and Persson, 2012, 412–413).

The above thought experiment entails that those who are morally enhanced would be free to act morally but not free to do “grossly immoral acts.” However, the God machine would guarantee one’s freedom if one chose to act morally, but it would only take away one’s freedom to fall. With this thought experiment, Persson and Savulescu establish that enhancement of moral dispositions such as altruism and justice would not limit one’s freedom, autonomy, and even responsibility.

Persson and Savulescu’s position could be read as accounting for a straightforward kind of freedom that focuses only on what an agent does and not on the moral choices available to them during their actions. Such straightforward freedom, even if necessary, is insufficient in the absence of further freedom to choose among alternative moral choices. If at any given time, an agent is morally determined, qua AI

⁵ One might object that individual freedom is traditionally a value more strongly associated with Western ethical theories (chiefly Kantianism) and not Afro-communitarianism. However, a plausible response will be that individual freedom is not chiefly Western or Kantian. Afro-communitarianism promotes individual freedom. In the Afro-communitarian literature, Kwame Gyekye, Bernard Matolino, Jonathan Chimakonam, Molefe Motsamai, and others have shown that Afro-communitarianism accommodates individual freedom. Elsewhere, I have argued that Ifeanyi Menkiti’s account of normative personhood upholds individual freedom, although it is secondary to communal duties and obligations (AE Chimakonam, A Personhood-based Theory of Right Action, 2023). For example, in Menkiti’s account, individual freedom is very essential in the process of attaining personhood. Individuals must decide whether to comply or not to comply with social norms, which is why one could fail or succeed at attaining personhood. Eliminating free moral choices will truncate this process of acquiring personhood. It is against this backdrop that I discussed the importance of free moral choices.

and moral enhancement, to have the moral capacities that they do have, and if those moral capacities casually determine their moral actions, even though they act morally, they cannot be said to have free choices. They satisfy Persson and Savulescu's conditions for free will. However, free will requires that an individual has the freedom to stand or fall irrespective of the magnitude of moral choices, and radical AI-based moral enhancement undermines this. The God machine seems more like a maker of moral zombies and a killer of moral responsibility (see Chimakonam, 2023a). Each time the God machine intervenes with its vroom, it denies the human subject of agency. Agency is borne out of the difficulty in choosing between two opposing moral choices. Moreover, even though, the God Machine prevents a moral zombie from committing a hideous evil, it destroys responsibility with the same broom. Many ethicists would agree that a world with hideous evils is far better than one without responsibility (see, e.g., Harris, 2016, Hauskeller, 2017). Also, it does not matter how small the God Machine's influence is; the suggestion that a machine could have some control over human consciousness obliterates any confidence in the existence of free will and choice. Thus, the God Machine is like a bull in a China shop.

To further interrogate this position, there is a need to differentiate those actions that an agent would have performed if they wanted from the ones they could not perform even if they wanted. Although the former refers to those moral choices that were available to an agent at the time of their action, the latter refers to the absence of moral choice. One might have the temptation to dismiss this as superficial freedom, but far from that, it differentiates the presence of moral choice from the absence of moral choice. Suppose that Amara has an ailurophobia (fear of cats). Imagine that one day, on her way to school, Amara saw a cat hit down by a hit-run driver bleeding to death near a children's park along the street and needed immediate medical attention. At the same time, Amara saw a dog that had lost its owner and needed help finding him. Suppose that Amara is the only one who arrives at the scene on time needed to save the cat's life and find the puppy's owner. Amara happily chooses to help the puppy, even though the puppy is not in immediate danger, eventually leaving the cat to die. When Amara chose to help the puppy, was she able to choose to save the cat? It seems not. Why? Given her ailurophobia, choosing to save the cat's life was practically not available to her since her fear of cats makes her unable to save the cat. Bringing this to our discussion, given that IMMs inevitably take the right course of action, choosing the wrong course of action is a choice not available to them because of their sufficient advancement and moral enhancement. In other words, even if a moral agent does act morally (determined by their radical AI-based moral enhancement), the alternative would not be available to them since morality requires freedom involving moral choices.

Persson and Savulescu might reply that so many things limit humans' free actions. For example, they have argued that "our power to act out of our own free will is a matter of degree" (Persson and Savulescu, 2014, 251) since nature imposes some limitations on our free will alongside other limitations imposed on us to avoid harm to ourselves and others. They cite our inability to lift a skyscraper with our bare hands and a feeling of revulsion that arises from the idea of putting excrement in our mouth as examples of the former. Some examples of the latter are those restrictions imposed on us by our society, such as moral education and civil punishment. They also argue that since we do not dispute some of the limitations imposed on our freedom in such ways, then the limitation imposed by moral enhancement on our freedom to prevent grossly immoral actions should be welcomed. For, as the argument goes, freedom is "only one value and not the sole value; safety is another" (Persson and Savulescu, 2014, 251). Persson and Savulescu's position is based on free action and not free choice. However, my argument is that free choice is what informs free actions, unless in those cases of coercion or compulsion. If there was free choice without corresponding free action, then the choice was never free. For our free choice would be said to arise from our free will. Free will is what informs our free thought. Imposing moral enhancement on us would rob us of our free choice unless Persson and Savulescu say that we should stop thinking or deciding for ourselves, which would be ridiculous! Rakic (2017, 386) puts forth a similar point thus; "[R]estrictions on our free will ... are restrictions on our free thought. As soon as our freedom to think is restricted, even slightly, we cannot consider ourselves as being deprived of our freedom 'to a degree.' In that case, we can only call ourselves unfree." And this is one of the greatest dangers that the Alfication of humanity poses.

Although I agree with Persson and Savulescu that there is a need for humans to behave morally, but that should not be at a greater cause to humanity. I doubt whether radical AI-based moral enhancement would allow us to be free to act morally. Radical AI-based moral enhancement and the God machine would be both an intrinsic and extrinsic constraint that would undermine our moral choices. Consider, for example, someone who is mentally ill and still has the ability to act as they want without being externally constrained. Yet, we do not ordinarily judge them to be fully responsible for their choices in the same way we do healthy adults. Also, consider that such mentally ill individuals are confined to a mental institution where their actions are restricted by their caregivers. Analogically, radical AI-based moral enhancement would be an intrinsic constraint and the God Machine would be an internal/external constraint that would undermine individuals' moral choices. Rakic (2017, 3) posits a similar argument when he argues that "... the very moment we levy external limitations on our free will, even if those limitations are minor, it ceases to be free." He buttresses that "by imposing limitations on what we are allowed to will, such a mechanism intervenes in what we are free to think" (Rakic, 2017, 3). For instance, the God machine would intervene when one makes the decision to do a wrong course of action, thereby preventing them from making such a choice.

At this juncture, the proponents of radical AI-based moral enhancement might object that losing some freedom does not immediately translate to losing all freedom. They may argue that morally enhanced persons would retain their freedom to act morally but would lose their freedom to act immorally, which will accrue in a net benefit for them. Michael J. Selgelid argues that "... a net loss of liberty does not entail a complete loss of liberty. Under a regime of mandatory enhancement, people would maintain wide-ranging freedom of conduct." He adds that "[a] net loss of freedom need not entail that "freedom would no longer be intact"—a net loss of freedom might simply mean that some freedom is lost (while overall freedom remains largely intact)" (Selgelid, 2014, 215). Further support for this claim is evident in Persson and Savulescu's work, where they claim that losing some part of our freedom, especially the freedom to act immorally, would not undermine individuals' freedom. They further argue that if it undermines freedom, the benefit that would accrue from such loss of "freedom to fall" outweighs the value of freedom (Savulescu and Persson, 2012, 416). In this light, Persson and Savulescu seem to argue that we should always limit freedom in situations where a moral action would cause greater harm, and we should always let human well-being outweigh greater harm.

It is incorrect to say that the benefits that would accrue from such loss of "freedom to fall," such as "human well-being and respect for basic rights, outweigh the value of freedom" (Persson and Savulescu, 2012, 416). Freedom of choice is at the core of our humanity; losing it would undermine our humanity, making us IMMs. What type of benefit can possibly outweigh this mother of all losses? I would argue against Selgelid, Persson, and Savulescu that if I lose my freedom to act immorally (having chosen to act immorally), what is left is no longer freedom but compulsion by God Machine or Mr Black to act morally. As weird as it might sound, the freedom to act immorally is what stands between free choice and compulsion.

So far, I have argued that radical AI-based moral enhancement would result in AIfication of humans, that is, humanity becoming IMMs. I also argued that since these IMMs would lack humans' cognitive and moral limitations, they would know and do what is morally required and would be incapable of acting immorally. This will imply the technologization of humanity through radical AI-based moral enhancement that could result in machines with supermoral capacities.

4. AI dominance

In this section, I will consider AI dominance as another moral consequence that would arise from the intersection of AI and transhumanism. The problem of machine dominance has been represented in various ways in science fiction, such as in novels like Samuel Butler's *Erewhon*, Jack Williamson's *The Humanoids*, and movies like *Frankenstein*, and *2001: A Space Odyssey*, where intelligent machines turned against humans. Scholars like Francis Fukuyama (2002), Annas et al. (2002), Charles Rubin (2003), Nicholas Agar (2013), Leon Kass (2014), and Bostrom (2014) have also expressed some worries

about the possibility of posthumans subduing (and even replacing) humans. Although these worries deserve some attention, I will focus on AI systems becoming morally superior agents to humans. Ben Goertzel points out that in the near future, “AI’s will possess true AGI, not necessarily emulating human intelligence, but equaling and likely surpassing it” (Goertzel, 2002, online). Similarly, Floridi and Sanders (2004, 351) claim that AI moral agents would be “sufficiently informed, ‘smart,’ autonomous and able to perform morally relevant actions independently of the humans that created them.” We can suppose that sufficiently advanced AI systems could develop moral reasoning and be better at solving ethical problems than humans. Just as humans consider themselves morally superior to animals because of their advanced intellect and their possession of moral capacities, AI moral agents would consider themselves morally superior to humans because of their technologically advanced moral capacities. Joseph Emile Nadeau has argued that “[humans] are not moral agents but robots are” since “an action is a free action if and only if it is based on reasons fully thought out by the agent” (cited in Sullins, 2006, 27). Because humans would not possess the advanced intellect that AI moral agents would possess, they often make immoral and illogical decisions based on emotional attachments, personal bias and prejudice. However, AI moral agents are logically directed and capable of making moral and logical decisions devoid of emotional encumbrances.

The problem that this AI dominance portends for humanity is that AI moral agents would consider humans to be moral patients and not moral agents since humans are morally lower beings that sometimes fail at the gate of morality. As Hall points out; “[Humans] will all too soon be the lower-order creatures. It will behoove us to have taught [AI moral agents] well their responsibilities toward us” (Hall, 2001, 6). In other words, AI agents would be higher-order creatures and humans would be lower-order creatures. Just as we consider babies, animals and the environment moral patients because they possess lower capacities than us, AI moral agents with supermoral capacities would consider us lower-moral creatures. One of the principal reasons AI moral agents would consider humans as moral patients would derive from humans’ capacity to do both moral and immoral courses of action. As moral patients, AI moral agents would owe humans certain moral responsibilities (let us call them *minimal responsibilities*), such as safeguarding humans’ well-being, which might differ from the ones they owe to each other as moral agents (let us call them *maximal responsibilities*) such as preserving their best interests. One troubling consequence of this is that these new moral paragons could wipe out human beings in a whole village or city for their own ends, just like human industrialists clear a whole forest or destroy a coral reef during dredging. We can expect, of course, that a few of these machines might advocate our protection but which can be conveniently ignored by the majority as is currently the case in environmental and climate change advocacy.

In addition, AI moral agents could sacrifice these minimal responsibilities in cases where they clash with maximal responsibilities. This implies that AI moral agents would put their best interest first, especially when their survival is at stake, even if it means sacrificing some of these minimal responsibilities. Agar paints a picture of this with his idea of “supreme opportunities,” which “arise in respect of significant potential benefits best secured by sacrificing morally considerable beings” (Agar, 2013, 72). He argues that supreme opportunities will allow “mere persons” to be sacrificed for “post-persons” significant benefits. Agar concludes that “[t]here is, therefore, some inductive support for the notion that post-persons will allocate benefits to mere persons only when all of the needs of post-persons are met. The hopes of mere persons will depend on the predictions of some futurists that technological progress will create a super-abundance that enables the(sic) all of the interests of post-persons and mere persons to be concurrently satisfied” (Agar, 2013, 73). Although Agar’s argument is directed at enhancing moral status, his argument is vital here. It shows that there will be some contexts in which AI moral agents would sacrifice humans’ well-being to satisfy their best interests. Perhaps, the relationship between humans and AI moral agents will depend on the fact that humans exist to satisfy their best interests, and minimal responsibilities could be sacrificed whenever such best interests are at stake. Humans would only hope that such a clash never happens.

However, some scholars like Moravec would argue that humans and AI moral agents will have a harmonious relationship since AI agents are our artificial progeny. He claims that our “mind children” will regard us as parents (Moravec, 1988). However, from an evolutionary standpoint, one doubts whether

such a harmonious relationship will be possible. Evolution has proven that stronger species often consider the weaker species as “prey.” Even though, humans, as a higher species, have developed some moral constraints to restrict such prey instinct, they still prey on animals for their best interest. For instance, consider the use of rats for cancer research or the use of mice for biomedical and scientific research. Similarly, AI moral agents will consider humans as “prey” whenever that satisfies their best interest despite the minimal responsibilities they owe to humans. For example, they could use the human brain for scientific research.

The challenge for us now is to ensure that we develop AI systems that do not raise this dominance problem or will Alifcate humans—AI systems that will be in a complementary relationship with humans, which I aim to do in the following section.

5. Some personhood-based relational ethical principles for research and policy development in AI and transhumanism

Here, I seek to extend my idea of a personhood-based theory of right action (Chimakonam, 2021b, 2023b) to the intersection of AI and transhumanism. In the preceding publications, I articulated and formulated a personhood-based theory of right action grounded in the notion of complementary relationship salient in most cultures in Africa. This theory has one main principle that states that “an action is right if and only if it positively contributes to the common good while adding moral excellencies to the individuals; an action is wrong if it adds moral excellencies to individuals without contributing to the common good, or contributes to the common good without adding moral excellencies to the individuals” (Chimakonam, 2023b, 112).

This main principle has two exception clauses: on the one hand, the Communal Exception Clause states that “an action X (for one thing) is a communal exception in a case Y if and only if there is an extreme group necessity, all things considered, to violate adding moral excellencies to the individuals in order to sacrifice to the common good for the sake of collective interest.” On the other hand, the Individual Exception Clause states that “(for another thing) an action X is an individual exception in a case Y if and only if there is an extreme personal necessity, all things considered, to violate contributing to the common good in order to add moral excellencies to the individuals for the sake of such individuals’ interest” (Chimakonam, 2023b, 116).

Both the main principle and two exception clauses are grounded in an African-inspired three-valued logic, known as Ezumezu (see Chimakonam, 2019). The three supplementary laws of Ezumezu logic ground the principles of relationality, complementarity and contextuality central to a personhood-based theory of right action. The principle of relationality states that “[v]alues necessarily interrelate irrespective of their unique contexts, all things considered, because no value is in isolation from others” is based on the law of Njikoka that affirms the relationship of individual variables. Further, the principle of contextuality stipulates that “[t]he relationships between values occur within specific contexts because context upsets values,” which is based on the law of nmekoka that upholds that a proposition cannot both be true and false in the same context. Finally, the principle of complementarity says thus; “[S]eemingly opposed values can have a relationship of complementation rather than contradiction” and it is grounded in the law of Q̄n̄ona-etiti that posits that in a complementary mode of thought, a proposition could be both true and false (Chimakonam & Chimakonam, 2022, 335).

The principles of relationality and complementarity and the laws of Njikoka and Q̄n̄ona-etiti ground the main principle of personhood-based theory of right action, which recognizes that we are not self-sufficient and need the complementation of others. It emphasizes the fact that we are beings in relationship with others. As Ifeanyi Menkiti points out, “the individual does not exist alone and cannot exist alone except corporately. Only in terms of other people does the individual become conscious of his own being, his own duties, his privileges, and responsibilities toward himself and toward other people” (Menkiti, 1984, 172). Innocent Asouzu echoes a similar idea when he claims that “to be in existence, an entity must be perceived by any of the units with which it constitutes a complementary whole relationship within which its existence is co-affirmed. This is why that person is to be pitied who thinks that a subject can

afford to live alone (*ka so mu di*)” (Asouzu, 2004, 277). In other words, every one of us has the ability to be in a relationship and interact with others. We do not exist in isolation but in a group where we are interconnected and interdependent on others. In this way, this main principle projects a mutual relationship in which we set aside our individual differences to work for the common good. In addition, because we are all bound up in a relationship geared toward the common good, we also acquire individual excellencies. In this form of relationship, both the common good and individual excellencies complement each other. Also, the principle of contextuality and the law of *nmekoka* underpin the two exception clauses that recognize that moral actions depend on context and consider the contexts of our moral actions (see Bambele, 2022). The two exception clauses project a form of nonmutual relationship in which we affirm our differences in order to promote our individual excellencies without bringing about negative consequences to others. This implies that there are some contexts that require us to detach from the group solely for the promotion of our own good, but our actions need not necessarily bring about negative outcomes for the group. In the rest of this section, I will articulate some Afro-ethical principles from this personhood-based relational ethics for AI and transhumanism research and policy.

5.1. *The 3-I*

Given that the intersection of AI and transhumanism poses the moral danger of the AIfication of humanity and machine dominance, I believe that there is a need to act now and not wait for AI to reveal its full capacities and then play catch up. Powers and Ganascia (2020), 28) disappointedly state the reactionary approach of the field of AI ethics thus; “[W]e (ethicists) generally learn of AI applications only after they appear, at which point we attempt to ‘catch up’ and possibly alter or limit the applications. This is essentially a rearguard action.” If we could go ahead of AI developers and programmers to anticipate the ethics of AI before such systems are developed, the field will become more “precautionary” rather than “reactionary.” We (ethicists) would anticipate the emergence of some of these AI systems before they are developed and figure out possible ethical approaches to them. The benefit of such a precautionary approach lies in avoiding some of the moral problems that AI will create when they arrive that will be difficult or impossible to deal with. As ethicists, this precautionary approach will help us to take charge of the AI systems at the designing stage rather than having to deal with the moral consequences of AI systems that are already embedded in society and widely in use. At least, as ethicists, we can try to provide ethical guidelines before AI systems are fully developed and introduced into society.

Accordingly, my aim here is to provide Afro-ethical principles drawn from personhood-based relational ethics for AI and transhumanism research and policy development. So far, ethical guidelines for AI largely come from the West, such as Europe and North America and are mainly drawn from the Western ethical tradition. For instance, the European Commission’s 2018 European Group on Ethics in Science and New Technologies, the UK House of Lords’ 2018 AI Committee’s report and France’s 2018 Villani report emphasize “transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, solidarity” and explicability (Jobin et al., 2019; Floridi and Cowsls, 2019). However, Africa has played little role in designing algorithms and drawing up ethical guidelines from African ethics for AI development, programming and application. Although UNESCO has done a tremendous and commendable job in this regard, its ethical guidelines draw very much from Western ethical principles and not African ethics (van Norren, 2022). This is why personhood-based relational ethics is important in offering some Afro-ethical principles for AI and transhumanism research and policy development. I will go ahead and articulate three Afro-ethical principles based on personhood-based relational ethics that will address the problems of AIfication of humanity and AI dominance that pose a barrier to the complementary relationship of humans and AI systems, which can be referred to as the 3-I: Inter-relationality, inter-contextuality, and inter-complementarity. These principles in their original formulation (relationality, contextuality, and complementarity), as discussed above, apply to humans alone. However, the extension and re-articulation I am proposing here using the prefix “inter” is to make them applicable to both humans and AIs.

1. *The Afro-ethical principle of inter-relationality*
Humans and AI should mutually interrelate, all things considered, to maximize the common good (to promote the common good, AI models should be designed with the principle of relationality to make them able to have a mutual interrelationship with humans, animals, and the environment);
2. *The Afro-ethical principle of inter-contextuality*
Humans and AI can forgo this mutual interrelationship if and only if there is an extreme context necessity to contribute to each interest instead of the common good without bringing any negative consequence to the other (AI models should be designed in a way that they can affirm their difference, when need be, without jeopardizing the good of others);
3. *The Afro-ethical principle of inter-complementarity*
A harmonious society should be based on the inter-complementarity of humans and AI (to maintain harmony in society, AI models, and engines should be designed in ways that complement humans).

In the above, the Afro-ethical principle of inter-relationality explains the mutual relationship between two opposites, humans and AI systems. The Afro-ethical principle of inter-contextuality affirms the good of each in their different context. In this way, each maintains a kind of nonmutual relationship, as explained above. Finally, the principle of inter-complementarity marshals the relationship between humans and AI such that they struggle through their differences to uphold a harmonious society.

However, critics might object that the Afro-ethical principles for AI and transhumanism research and policy development are too human-centered since these Afro-ethical principles spell out how AIs would be in a complementary relationship with humans. They might claim further that there is a need to clearly spell out human responsibilities toward AI systems to avoid abuse and misuse. Although this criticism raises a serious concern, I believe that the Afro-ethical principles proposed here address this concern by considering both humans and AIs as entities in a mutual inter-relationship, working together to achieve their common good.

6. Conclusion

In this essay, I considered some of the moral problems that the intersection of AI and transhumanism presents, namely, the AIfication of humans and AI dominance. I have shown that these moral problems pose a barrier to the complementary co-existence of humans and AI. To address these moral problems, I articulated Afro-ethical principles from personhood-based relational ethics for AI and transhumanism research and policy development. These Afro-ethical principles, identified as the 3-I, are inter-relationality, inter-contextuality, and inter-complementarity. However, further research is required to broaden the African ethical contribution to AI and transhumanism research and policy development.

Author contribution. Conceptualization: A.A. and A.B. Methodology: A.A. Writing original draft: A.A. and A.B. The author approved the final submitted draft.

Provenance. This article was accepted for the 2024 Data for Policy Conference and published in *Data & Policy* on the strength of the Conference review process.

Competing interest. The author declares no competing interests exist.

References

- Agar N (2013) Why is it possible to enhance moral status and why doing so is wrong? *Journal of Medical Ethics*, 39(2), 67–74.
- Ali SM (2021) Transhumanism and/as whiteness. In Hofkirchner W and Kreowski H-J (eds.), *Transhumanism: The Proper Guide to a Posthuman Condition or a Dangerous Idea*. Cham: Springer, pp. 169–184.
- Anderson DL (2013) Machine intentionality, the moral status of machines, and the composition problem. In Mueller VC (ed.), *Philosophy and Theory of Artificial Intelligence*. Cham: Springer Nature, pp. 321–334.
- Annas GJ, Andrews LB and Isasi RM (2002) Protecting the endangered human: Toward an international treaty prohibiting cloning and inheritable alterations. *American Journal of Law & Medicine*, 28(2–3), 151–178.

- Asouzu II** (2004) *The Method and Principles of Complementary Reflection in and Beyond African Philosophy*. Calabar: University of Calabar.
- Boddington P** (2023) *AI Ethics: A Textbook*. Cham: Springer Nature.
- Boddington P** (2017) *Towards a Code of Ethics for Artificial Intelligence*. Cham: Springer.
- Bostrom N** (2005) A history of transhumanist thought. *Journal of Evolution and Technology*, 14(1), 1–25.
- Bostrom N** (2014) Introduction—the transhumanist FAQ: A general introduction. In Mercer C and Maher DF (eds.), *Transhumanism and the Body: The World Religions Speak*. New York: Palgrave Macmillan, pp. 1–18.
- Bambale, Z** (2022) A Personhood-Based Theory and the Death Penalty: An Appraisal of AE Chimakonam's Theory of Right Action. *Arumaruka: Journal of Conversational Thinking*, 2 (2):1–21.
- Bostrom, N & Yudkowsky, E** (2014), The Ethics of Artificial Intelligence. In Frankish, K and Ramsey, W (eds.), *Cambridge Handbook of Artificial Intelligence*. New York: Cambridge University Press.
- Chimakonam AE** (2023a) Afro-communitarianism and transhumanism. In Imafidon E, Tshivhase M and Freter B (eds.), *Handbook of African Philosophy*. Cham: Springer.
- Chimakonam AE** (2021a) Transhumanism in Africa: A conversation with Ademola Fayemi on his Afrofuturistic account of personhood. *Arumaruka: Journal of Conversational Thinking*, 1(2), 42–56. <https://doi.org/10.4314/ajct.v1i2.3>
- Chimakonam AE** (2023b) A personhood-based theory of right action. In Chimakonam JO and Cordeiro-Rodrigues L (eds.), *African Ethics: A Guide to Key Ideas*, London: Bloomsbury, 103–120.
- Chimakonam AE** (2021b) Toward a personhood-based theory of right action: Investigating the Covid-19 pandemic and religious conspiracy theories in Africa. *Filosofia Theoretica: Journal of African Philosophy Culture and Religions*, 10 (2),191–210.
- Chimakonam AE** (n.d.) Contending with superchoice in a transhumanist future: Is the normative conception of personhood under threat? Unpublished thesis, Department of Philosophy, University of Johannesburg, South Africa.
- Chimakonam JO** (2019) *Ezumezu Logic: A System of Logic for African Philosophy and Studies*. Cham: Springer.
- Coeckelbergh M** (2020) *AI Ethics*. Cambridge: MIT Press.
- De Grey A and Rae M** (2007) *Ending Aging: The Rejuvenation Breakthroughs that Could Reverse Human Aging in Our Lifetime*. New York: St. Martin's Publishing Group.
- DeGrazia D** (2014) Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics*, 40(6), 361–8.
- Douglas T** (2008) Moral enhancement. *Journal of Applied Philosophy*, 25(3), 228–245.
- European Commission (EC)** (2018), “Communication on AI for Europe.” available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0237>
- Floridi L & Cows, J** (2019) A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1): 1–14.
- Floridi L and Sanders JW** (2004) On the morality of artificial agents. *Minds and Machines* 14, 349–379.
- Frankfurt H** (1969) Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66 (23),827–839.
- Fukuyama F** (2002) *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar, Straus & Giroux.
- Goertzel B** (2002) Thoughts on AI Morality. Available at <https://www.goertzel.org/dynapsyc/2002/AIMorality.htm> (accessed 26 October 2023).
- Gunkel DJ** (2012) *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. Cambridge: Massachusetts Institute of Technology.
- Hall JS** (2001) Ethics for Machines. Available at <http://www.kurzweilai.net/ethics-for-machines> (accessed 10 October 2023).
- Hagendorff T** (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hanna R and Kazim E** (2021) Philosophical foundations for digital ethics and AI ethics: A dignitarian approach. *AI and Ethics*, 1, 405–423. <https://doi.org/10.1007/s43681-021-00040-9>
- Haraway D** (1991) A Cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In Haraway, D (auth.), *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge, pp.149–181. <https://theanarchistlibrary.org/library/donna-haraway-a-cyborg-manifesto>
- Harris J** (2016) *How to be Good: The Possibility of Moral Enhancement*. Oxford: Oxford University Press.
- Haugeland J** (ed.) (1981) *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge: MIT Press.
- Hauskeller M** (2017) Is it desirable to be able to do the undesirable? Moral enhancement and the Little Alex problem. *Cambridge Quarterly of Healthcare Ethics*, 26(3), 365–376.
- Heilinger J-C** (2022) The ethics of AI ethics. A constructive critique. *Philosophy & Technology*, 35(61), 1–20.
- IOL News** (2007) 9 Killed in Army Horror. By Hosken G, Schmidt M and du Plessis J. Available at <https://www.iol.co.za/news/south-africa/9-killed-in-army-horror-374838> (accessed 26 September 2023).
- Kass L** (2014) Preventing a brave new world. In Sandler RL (ed.), *Ethics and Emerging Technologies* London: Palgrave Macmillan, pp. 76–89.
- Kurzweil R** (2005) *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.
- Jobin A, Ienca M and Vayena E** (2019) Artificial intelligence: The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1:389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Lee N** (2019) Brave new world of transhumanism. In Lee N (ed.), *The Transhuman Handbook*. Cham: Springer, pp. 3–48.
- Liao SM** (2020) A short introduction to the ethics of artificial intelligence. In Liao SM (ed.), *Ethics of Artificial Intelligence*. New York: Oxford University Publication, pp. 1–42.

- McCarthy J et al.** (1955) A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Available at <https://raysolomonoff.com/dartmouth/boxa/dart564props.pdf> (accessed 7 September 2023).
- Menkiti I** (1984) Person and community in African traditional thought. In Wright R (ed.), *African Philosophy: An Introduction*. New York: University Press of America, pp. 171–181.
- Moravec H** (1988) *Mind Children*. Cambridge: Harvard University Press.
- More M** (2013) The philosophy of transhumanism. In More M and Vita-More N (eds.), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, And Philosophy of the Human Future*. West Sussex: John Wiley & Sons, Ltd, 3–17.
- Müller, VC.,** (2022) Basic Issues in AI Policy. In Grau-Ruiz, MA., (ed.), *Interactive Robotics: Legal, Ethical, Social and Economic Aspects*. 3–9. Cham: Springer.
- Nyholm S and Ruther M** (2023) Meaning in life in AI ethics—some trends and perspectives. *Philosophy & Technology*, 36(20). <https://doi.org/10.1007/s13347-023-00620-z>
- Persson I and Savulescu J** (2008) The perils of cognitive enhancement and the urgent imperative to enhance moral character of humanity. *Journal of Applied Philosophy*, 25(3), 162–177.
- Persson I and Savulescu J** (2012) *Unfit for the Future: The Need for Moral Enhancement*. Oxford: Oxford University Press.
- Persson, I. and Savulescu, J.** (2014). Should Moral Bioenhancement be Compulsory? Reply to Vojin Rakić. *Journal of Medical Ethics* 2014; 40:251–252.
- Powers TM and Ganascia J-G** (2020) The ethics of the ethics of AI. In Dubber MD, Pasquale F and Das S (eds.), *The Oxford Handbook of Ethics of AI* Oxford: Oxford University Press, pp. 27–52.
- Persson, I. and Savulescu, J.** (2014). Should Moral Bioenhancement be Compulsory? Reply to Vojin Rakić. *Journal of Medical Ethics* 2014; 40:251 –252.
- Rakić V** (2017) Moral bioenhancement and free will: Continuing the debate. *Cambridge Quarterly of Healthcare Ethics*, 26(3), 384–393.
- Rakić V** (2014) Voluntary moral enhancement and the survival-at-any-cost bias. *Journal of Medical Ethics*, 40(4), 246–250.
- Rich E** (1983) *Artificial Intelligence*. New York: McGraw-Hill.,
- Rubin CT** (2003) Artificial intelligence and human nature. *The New Atlantis*, 1, 88–100.
- Russell S** (2016) Rationality and intelligence: A brief update. In Müller VC (ed.), *Fundamental Issues of Artificial Intelligence* Cham: Springer International Publishing, pp. 7–28.
- Russell S and Norvig P** (2010) *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River: Prentice Hall.
- Sorgner SL** (2022) *We have Always Been Cyborgs: Digital Data, Gene Technologies, and an Ethics of Transhumanism*. Bristol: Bristol University Press.
- Sullins JP** (2006) When is a robot a moral agent? *International Review of Information Ethics*, 6 (12), 23–30.
- Savulescu, J & Persson, I** (2012) Moral Enhancement, Freedom, and the God Machine. *The Monist*, 95(3):399–421.
- Selgelid, MJ** (2014) Freedom and moral enhancement. *Journal of Medical Ethics*, 40 (4):215–216.
- UNESCO** (2021) Recommendation on the Ethics of Artificial Intelligence. Available at <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- van Norren, DE.,** (2022), The Ethics of Artificial Intelligence, UNESCO and the African Ubuntu Perspective. *Journal of Information, Communication and Ethics in Society*, 21(1): 112–128
- van Norren DE** (2023) The ethics of artificial intelligence, UNESCO and the African Ubuntu perspective. *Journal of Information, Communication and Ethics in Society*, 21(1), 112–128.
- Vinge V** (1993) The coming technological singularity: How to survive in the post-human era. *NASA*, version 21:11–22.
- Vita-More N** (2019) History of transhumanism. In Lee N (ed.), *The Transhumanism Handbook*. Cham: Springer, 49–62.
- Waelen R** (2022) Why AI ethics is a critical theory. *Philosophy & Technology*, 35 (9), 1–16. <https://doi.org/10.1007/s13347-022-00507-5>