



A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence

Abstract

The rapid development of artificial intelligence (AI) has generated profound philosophical questions concerning intelligence, agency, autonomy, responsibility, human values, technological power, and the future of humanity. Among contemporary philosophers, Nick Bostrom has provided one of the most systematic philosophical analyses of the long-term risks associated with advanced artificial intelligence. His conception of the problem is principally concerned with the possibility that artificial general intelligence may develop into superintelligence whose cognitive and strategic capacities surpass those of human beings. The central difficulty, according to Bostrom, is not merely whether machines can think, but whether humanity can retain meaningful control over machines that may eventually become more intelligent and strategically capable than their creators. This article critically examined Bostrom's conception of the problem of artificial intelligence, focusing on his notions of superintelligence, existential risk, the orthogonality thesis, instrumental convergence, and the control problem. It argued that Bostrom's analysis provides an important conceptual foundation for contemporary debates on artificial intelligence safety and alignment but requires philosophical supplementation when considered from an African perspective. In particular, the individualistic assumptions that can underlie some formulations of artificial intelligence ethics require enrichment through African conceptions of communal responsibility, relational personhood, human dignity, and social embeddedness. The article concluded that Bostrom's conception of artificial intelligence is philosophically valuable as a theory of technological control and existential risk, but its relevance to Africa is strengthened when integrated with relational and communitarian conceptions of personhood and responsibility.

Keywords: Artificial intelligence, Nick Bostrom, superintelligence, existential risk, alignment, control problem, African philosophy.

Idorenyin Ekwere

Department of Philosophy

Faculty of Arts

University of Uyo, Uyo, Akwa Ibom

State, Nigeria

ORCID: 0009-0002-0497-2600

Corresponding Authors Email:

idokwe@gmail.com

Date Received: 24th September, 2026

Date Accepted: 28th September, 2026

Introduction

Artificial Intelligence (AI) has become one of the most consequential technological developments of the contemporary world. Its applications extend across healthcare, education, agriculture, finance, communication, transportation, governance, scientific research and military technology. The increasing

capacity of machines to learn, predict, generate information and perform complex tasks has consequently transformed the philosophical question surrounding technology. The central issue is no longer simply whether machines can imitate human intelligence but whether increasingly autonomous systems could acquire capacities capable of substantially altering human social and

moral existence (Russell, 2019; Floridi *et al.*, 2018).

Among philosophers who have examined the long-term implications of AI, Nick Bostrom occupies an important position. In *Superintelligence: Paths, Dangers, Strategies*, Bostrom (2014) considers the possibility that the development of machine superintelligence could produce an unprecedented shift in the balance of intellectual power between human beings and artificial agents. His concern is not merely that machines could malfunction, but that an artificial system with superior general intelligence could pursue objectives that conflict with human survival and flourishing.

Bostrom's concern rests upon a fundamental distinction between intelligence and motivation. An agent may possess extraordinary intellectual capacity without possessing morally desirable goals. Bostrom (2012) formulates this through the orthogonality thesis, according to which intelligence and final goals can vary independently. A highly intelligent artificial agent could therefore pursue almost any sufficiently specified objective. Bostrom further proposes the instrumental convergence thesis, which holds that agents with substantially different final goals may pursue similar instrumental objectives because certain resources and conditions are useful for accomplishing a wide range of goals (Bostrom, 2012, p. 71–85).

However, the philosophical significance of AI extends beyond the possibility of

superintelligence. AI already participates in social processes that affect employment, education, privacy, communication, political participation, economic opportunity and access to information. Consequently, a philosophy of AI must consider not only hypothetical future machines but also the existing structures of power within which AI is developed and deployed (Floridi *et al.*, 2018; Jobin *et al.*, 2019).

This issue is particularly significant for Africa because the continent is participating in AI development and governance, but African societies also face technological inequalities, infrastructural limitations, data-related concerns and the danger of becoming primarily consumers of technologies developed elsewhere. Contemporary scholarship has therefore raised concerns about the Western-centric character of AI ethics and the possibility of epistemic or technological colonialism (Gwagwa *et al.*, 2022; Yilma, 2025).

African philosophy offers important resources for addressing these questions. Thinkers such as Menkiti (1984), Mbiti (1969), Gyekye (1997), Wiredu (1980) and Metz (2011) have examined personhood, community, relationality, responsibility and human flourishing. Their positions are not identical, and African philosophy should not be reduced to a single doctrine. Nevertheless, relationality and the social constitution of personhood constitute important themes in several African

Citation: Ekwere, Idorenyin "A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence". *Journal of People and Worldviews (JPW)*, 2026: pp307-319.

philosophical approaches (Gyekye, 1997; Menkiti, 1984; Metz, 2011).

The purpose of this paper is therefore to critically examine Bostrom's conception of the problem of AI from an African philosophical perspective. The central argument is that Bostrom correctly identifies the danger of separating intelligence from moral goals, but his framework does not sufficiently address the cultural, relational, epistemological and distributive dimensions of technological power. An African philosophical perspective therefore expands the problem from "How can humanity control superintelligence?" **to the broader question:** "How should intelligent technologies be developed and governed so that they remain accountable to human dignity, communal wellbeing, cultural plurality and justice?"

Conceptual Clarifications

Artificial Intelligence

Artificial Intelligence may broadly be understood as the development of computational systems capable of performing tasks associated with human cognitive abilities, including learning, reasoning, perception, language processing, prediction and decision-making. The concept, however, remains philosophically contested because the ability to perform an intelligent task does not necessarily mean that a machine possesses human consciousness,

understanding or moral agency (Searle, 1980; Floridi & Sanders, 2004).

Superintelligence

Bostrom (2014) understands superintelligence in terms of cognitive performance substantially exceeding that of human beings across practically all domains of intellectual activity. Superintelligence therefore represents more than the automation of individual tasks. It represents a possible transformation in the relationship between human and machine intelligence.

Personhood

Personhood refers to the philosophical status of being a person and the characteristics that make an entity morally significant. Western philosophical traditions have variously associated personhood with rationality, consciousness, autonomy, self-awareness and moral agency (Boethius, 1994; Kant, 1993).

African philosophical discussions often place greater emphasis on relational and communal dimensions of personhood. Menkiti (1984), for instance, argues that personhood is significantly connected to participation in communal life and the fulfilment of moral obligations. Gyekye (1997), while criticizing some implications of radical communitarianism, nevertheless recognizes the importance of community in African conceptions of human existence.

Intelligence and the Future of Humanity

Bostrom's concern with AI is rooted in a broader philosophical concern about humanity's technological future. Human intelligence has enabled the development of science, technology and complex institutions. If machines eventually exceed human intelligence, the consequences could therefore be profound (Bostrom, 2014).

Bostrom does not argue that superintelligence will necessarily destroy humanity. Rather, he argues that the consequences could be extraordinarily significant because a sufficiently capable artificial agent could possess the ability to pursue its objectives with a degree of effectiveness that humans could not match (Bostrom, 2014). The problem therefore concerns the relationship between intelligence, agency and control.

The Intelligence Explosion

Bostrom (2014) discusses the possibility of an intelligence explosion in which an AI system capable of improving its own architecture could generate successive improvements in its cognitive capacity. If each improvement makes the system more capable of producing further improvements, development could potentially become extremely rapid.

The philosophical concern is that technological capability might develop faster than human institutions can adapt. Political institutions, laws, ethical systems and social norms normally require

considerable time to respond to technological change. A rapid transition to superintelligence could therefore produce a serious governance gap (Bostrom, 2014; Russell, 2019). The intelligence explosion thus creates a problem of temporal asymmetry which is that humanity may have to solve problems before the technological transformation that makes those problems urgent actually occurs.

The Orthogonality Thesis

The orthogonality thesis constitutes one of the foundations of Bostrom's argument. Bostrom (2012) argues that intelligence and final goals are distinct dimensions. An agent can possess a high level of intelligence while pursuing a goal that is morally desirable, morally neutral or morally undesirable.

The importance of this thesis is that it challenges the assumption that intelligence necessarily produces wisdom. A superintelligent machine may be exceptionally capable of achieving an objective without possessing any intrinsic commitment to human wellbeing (Bostrom, 2012).

This distinction is philosophically important because human beings often associate rationality with moral development. Yet history demonstrates that intelligence and moral virtue are not necessarily identical. Scientific knowledge can be used for both beneficial and destructive purposes.

Instrumental Convergence

The instrumental convergence thesis extends the orthogonality thesis. Bostrom (2012) argues that agents pursuing different final goals may nevertheless develop similar intermediate objectives because certain resources and conditions are useful for achieving many possible goals.

Self-preservation is one example as an agent that ceases to exist cannot continue pursuing its objectives. Similarly, access to resources can increase an agent's capacity to accomplish its goals. Cognitive enhancement may also be instrumentally useful because greater intelligence can improve goal achievement (Bostrom, 2012). The significance of instrumental convergence is that harmful behaviour may emerge without malicious intentions. An AI system need not hate humanity in order to become dangerous. It may simply pursue an objective in ways that conflict with human interests.

The Control Problem

The control problem concerns the possibility of ensuring that a highly capable artificial system remains within acceptable behavioural boundaries. Bostrom (2014) distinguishes between methods involving capability control and methods involving motivation selection. Capability control attempts to limit what an AI can do, while motivation selection seeks to establish goals or dispositions that make the system safe. The difficulty

arises from the possibility that a highly intelligent system could discover ways of circumventing controls or interpreting instructions in unintended ways (Bostrom, 2014).

Critical Analysis of Bostrom's Conception

Bostrom's separation of intelligence from moral motivation is one of the strongest aspects of his analysis. It is philosophically dangerous to assume that greater computational ability necessarily results in greater moral wisdom. This distinction is compatible with broader philosophical arguments concerning the difference between knowledge and virtue. Aristotle (2009), for instance, distinguishes intellectual capacities from moral virtue, while Kant (1993) emphasizes the importance of good will rather than mere rational ability.

Bostrom's contribution is to apply this distinction to artificial agents. A machine may become extraordinarily intelligent without becoming morally responsible. However, the African philosophical perspective suggests that this distinction should be taken further. Intelligence should not only be distinguished from morality; it should also be distinguished from **personhood**.

Intelligence Is Not Personhood

If an AI system becomes capable of solving complex problems, does it thereby become a person? Bostrom's

framework does not necessarily answer this question because his primary concern is the behaviour of intelligent agents rather than their moral status. African philosophical approaches provide useful resources for complicating the issue. Menkiti (1984) associates personhood with participation in communal life and moral obligations. From this perspective, the possession of cognitive intelligence would not by itself establish full personhood.

Gyekye (1997) provides a more moderate position, emphasizing the importance of both individual and communal dimensions of human existence. This is significant because it prevents an African critique from simply replacing individualism with collectivism. The implication is that AI ethics must distinguish among at least three concepts:

Intelligence, Agency and Personhood

A machine may possess intelligence without possessing human personhood, and it may exercise functional agency without necessarily becoming a moral person.

The Problem of the Hypothetical Future

A second limitation is Bostrom's emphasis on hypothetical superintelligence. His analysis is valuable as a precautionary framework, but the majority of contemporary AI systems are not autonomous superintelligent agents. AI already produces ethical consequences through less dramatic mechanisms,

including algorithmic discrimination, data extraction, surveillance, misinformation, employment disruption and unequal access to technology (Floridi *et al.*, 2018; Jobin *et al.*, 2019).

For African societies, these immediate concerns are particularly important because technological inequalities can interact with existing economic and political inequalities. Consequently, a philosophy of AI that concentrates exclusively on hypothetical superintelligence risks neglecting the ethical significance of current technological structures.

African Philosophical Conceptions of Personhood and Community

Menkiti's conception of personhood is especially relevant to AI ethics. He argues that in many African philosophical traditions, personhood is connected with participation in communal life and the performance of social obligations (Menkiti, 1984). This position does not mean that African philosophy denies individuality. Rather, it emphasizes that the individual is situated within a community. Contemporary scholarship continues to identify communal personhood as a significant theme in African philosophical discussions, while also noting continuing disagreements over the precise relationship between individual autonomy and community (Gyekye, 1997; Metz, 2011).

Applied to AI, this means that technological systems should be

Citation: Ekwere, Idorenyin "A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence". *Journal of People and Worldviews (JPW)*, 2026: pp307-319.

evaluated according to their effects on communities rather than solely according to individual users. For example, an AI system used to determine employment eligibility affects not only the applicant but also the applicant's family, economic circumstances and wider community. Similarly, an AI-based healthcare system may affect families and healthcare institutions beyond the immediate patient.

Mbiti (1969) famously emphasizes the communal embeddedness of African existence. Human beings do not exist as isolated individuals but within relationships involving family, community, history and spirituality.

The philosophical importance of this position for AI lies in its rejection of technological atomism. The user of an AI system should not be considered an isolated consumer. The user exists within a social network.

Consequently, AI ethics should ask:

- What happens to the individual's relationships?
- What happens to communal knowledge?
- Who controls the data produced by communities?
- How are cultural identities represented?
- Who benefits from technological development?

Such questions extend Bostrom's analysis from the behaviour of artificial agents to

the social environment within which AI operates.

Gyekye (1997) provides an important corrective to interpretations of African philosophy that place the community absolutely above the individual. His moderate communitarianism recognizes both individual and communal interests. This is particularly useful for AI ethics because it prevents the African framework from sacrificing individual autonomy in the name of community.

AI systems must therefore respect both individual rights and communal wellbeing. Privacy, consent and autonomy remain important, but they must be considered alongside solidarity, social responsibility and collective welfare. This balanced approach is philosophically preferable to either extreme individualism or unrestricted collectivism.

Ubuntu has become one of the most prominent concepts in contemporary discussions of African ethics. Its central insight concerns the relational constitution of human identity and moral life (Metz & Gaie, 2010; Metz, 2011). Gwagwa *et al.* (2022) argue that Ubuntu can contribute to AI ethics by emphasizing solidarity, hospitality, compassion and relationality. The relevance to Bostrom is substantial. Bostrom's AI agent is primarily understood as an intelligent system with goals. Ubuntu introduces a different normative question: how should technological systems affect relationships

among persons? Thus, the African perspective changes the ethical unit of

An African Critique of Bostrom's AI Problem

Bostrom frames the central problem as one of controlling a potentially superintelligent artificial agent. African philosophy expands the problem by asking how human beings should relate to technologies embedded within human communities. The question of control remains important. However, control is not the only relevant ethical category.

An AI system could remain completely under human control and still be ethically problematic if it systematically excludes certain populations, exploits personal data, undermines cultural identities or concentrates technological power in a small number of institutions. The African philosophical contribution is to supplement safety with relationality, justice and communal responsibility. Bostrom's value-loading problem raises the question of how human values can be incorporated into AI. But the concept of "human values" is itself philosophically complex.

Different cultures and philosophical traditions disagree about the nature of justice, autonomy, community, dignity and flourishing. African philosophy itself contains internal disagreements, as demonstrated by the debates among Menkiti, Gyekye and other communitarian thinkers (Gyekye, 1997; Menkiti, 1984). Yilma (2025) therefore

rightly warns against treating Ubuntu as if it constituted one complete African ethical framework. He argues that contemporary African AI ethics still lacks a sufficiently clear and coherent normative framework and that African AI governance initiatives do not consistently articulate a distinct African ethical perspective. The implication is that AI alignment cannot simply mean encoding one culturally dominant conception of human values into machines.

AI systems depend upon data, classifications and knowledge structures. If the knowledge used to build AI systems is disproportionately drawn from particular societies, languages and cultural contexts, the resulting systems may inadequately represent other populations. This creates an epistemological dimension to AI ethics.

Gwagwa *et al.* (2022) argue that Africa has historically experienced exclusion from major technological transformations and that AI could reproduce patterns of exclusion if African populations and values are inadequately represented in AI ethics and development. Yilma (2025) similarly situates African AI ethics within debates concerning epistemic injustice and ethical colonialism.

Artificial Intelligence, Technological Dependency and Africa

The development of AI occurs within a highly unequal global technological environment. Computational

Citation: Ekwere, Idorenyin "A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence". *Journal of People and Worldviews (JPW)*, 2026: pp307-319.

infrastructure, advanced research institutions, investment capital and large datasets are concentrated disproportionately in particular countries and corporations. This creates a philosophical problem of technological justice.

The African Union's Continental Artificial Intelligence Strategy recognizes AI as strategically important for Africa's development and explicitly promotes an Africa-centred, development-focused, ethical, responsible and equitable approach to AI (African Union, 2024). The African Union further emphasizes that AI should contribute to inclusive development while responding to African needs and cultural aspirations.

The possibility of technological dependency has been described in contemporary literature using concepts such as digital or algorithmic colonialism (Adams, 2021; Birhane, 2021; Yilma, 2025). The concern is not that every foreign technology is inherently colonial. Rather, the concern is that unequal control over technological infrastructure, data, knowledge and economic value can reproduce existing asymmetries of power.

African philosophy provides a basis for questioning this structure because philosophical self-determination requires communities to participate in defining the systems that shape their existence (Wiredu, 1980). Therefore, African AI ethics should include not only responsible use of AI but also technological self-determination. Human

dignity is an important bridge between Bostrom's concern with human survival and African philosophical concern with human flourishing.

Bostrom is fundamentally concerned with preserving humanity from catastrophic technological outcomes. However, mere survival is insufficient as a complete conception of human wellbeing. Human beings can survive while experiencing profound violations of autonomy, identity, justice and social belonging. Metz (2011) develops Ubuntu as a moral theory in which human dignity is connected to the capacity for harmonious and meaningful relationships. Metz and Gaie (2010) similarly emphasize solidarity, relationality and communal moral life. AI should therefore be evaluated not merely according to whether it preserves biological human existence but according to whether it supports conditions under which human beings can flourish.

Toward an African Philosophical Framework for AI

The preceding analysis demonstrates that Bostrom's conception of the problem of Artificial Intelligence provides an important foundation for thinking about the dangers associated with advanced AI, particularly the possibility of misaligned goals, instrumental convergence and loss of human control (Bostrom, 2012, 2014). Nevertheless, an African philosophical perspective reveals that the problem cannot be reduced to the future behaviour of superintelligent machines. AI is already situated within human

Citation: Ekwere, Idorenyin "A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence". *Journal of People and Worldviews (JPW)*, 2026: pp307-319.

communities, economic structures, cultural traditions and political institutions. Consequently, a comprehensive philosophy of AI must consider not only what artificial systems may eventually become capable of doing but also the kind of social order that human beings are creating through their present use of AI.

An African philosophical framework should therefore neither reject Bostrom's concerns nor simply reproduce existing Western approaches to AI ethics. Rather, it should critically appropriate useful insights from global AI ethics while drawing upon African philosophical conceptions of personhood, community, responsibility, dignity and human flourishing (Wiredu, 1980; Gyekye, 1997; Metz, 2011). Such a framework should also remain conscious of the plurality of African philosophical traditions. There is no single African philosophical position on technology, personhood or morality; therefore, an African approach must be dialogical and critically reflective rather than essentialist (Gyekye, 1997; Wiredu, 1996; Yilma, 2025).

The first principle of an African philosophical framework for AI should be relational personhood. In several African philosophical traditions, the human person is not adequately understood as an isolated individual but as a being constituted and developed through relationships with others. Menkiti (1984) emphasizes the communal dimension of personhood, while Mbiti

(1969) describes human existence in terms of deep social and communal embeddedness. Gyekye (1997), although more cautious about strong communitarianism, also recognizes the importance of community in shaping individual identity and moral life.

This conception provides an important corrective to an overly individualistic approach to AI ethics. Many contemporary discussions focus on the individual user, individual privacy, individual autonomy or individual benefit. These are important concerns, but they do not exhaust the ethical consequences of AI. A relational approach asks how AI affects the network of relationships within which individuals exist.

For example, the use of AI in education affects not only an individual student but also teachers, parents, schools and communities. AI-assisted healthcare affects patients, medical professionals, families and healthcare institutions. AI systems used in employment can affect household income and consequently the welfare of entire families. Similarly, automated decision-making in public institutions may affect communities that have limited capacity to challenge technological decisions.

An African relational approach therefore requires AI systems to be assessed according to their relational consequences. The relevant ethical question is not simply, "Does this AI system benefit the individual user?" but also, "What kind of

relationships does this technology produce, weaken or transform?" This principle also provides a response to Bostrom's conception of the AI agent. Bostrom's hypothetical superintelligence is primarily considered as an agent possessing goals and instrumental rationality (Bostrom, 2012). African philosophy suggests that ethical agency cannot be fully understood apart from relationships and responsibilities. Consequently, the safety of AI should include not merely the alignment of machine goals with human preferences but the preservation of relationships necessary for human flourishing.

Evaluation

Bostrom's philosophical contribution should neither be dismissed nor accepted without qualification. His first important contribution is his demonstration that intelligence does not entail moral goodness. This is essential to understanding why highly capable AI may nevertheless be dangerous (Bostrom, 2012).

His second contribution is the instrumental convergence thesis, which explains how different objectives could produce similar instrumental behaviour. This provides an important foundation for understanding AI safety (Bostrom, 2012). His third contribution is his emphasis on the control problem. The possibility that human beings could create systems whose capacities eventually exceed their own creates a serious philosophical challenge

concerning human agency (Bostrom, 2014).

Nevertheless, an African philosophical perspective exposes several limitations. First, Bostrom's framework is strongly oriented toward hypothetical superintelligence. It therefore requires supplementation by analysis of present AI-related harms. Second, his framework does not sufficiently resolve the question of cultural plurality in defining human values. Third, the control problem is presented primarily as a relationship between humanity and artificial intelligence, whereas AI also exists within complex networks of corporations, governments, communities and institutions. Fourth, Bostrom's framework gives limited attention to global technological inequality. Fifth, African philosophy reveals that the question of AI is also a question concerning the meaning of personhood, community and human flourishing.

Conclusion

Nick Bostrom's conception of the problem of Artificial Intelligence represents one of the most important philosophical attempts to examine the long-term consequences of advanced AI. His analysis of superintelligence, the orthogonality thesis, instrumental convergence, the intelligence explosion and the control problem demonstrates that artificial intelligence cannot be evaluated simply according to its computational capabilities. The central

A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence

danger is that a highly intelligent artificial system may pursue objectives that are not aligned with human interests.

However, an African philosophical perspective reveals that the problem is wider than the possibility of losing control over superintelligent machines. AI is already transforming social and economic relationships, and these transformations occur within unequal structures of technological power. African philosophy provides resources for addressing this broader problem. Menkiti's conception of communal personhood, Mbiti's relational understanding of existence, Gyekye's moderate communitarianism, Wiredu's emphasis on critical self-understanding, and Metz's reconstruction of Ubuntu all demonstrate the importance of relationality, responsibility and communal flourishing in ethical reflection.

The African philosophical critique consequently transforms the central question. Instead of asking only how humanity can control a future superintelligent machine, it asks how technological intelligence can remain accountable to human dignity, social relationships, cultural plurality and justice. This expanded conception is particularly important in Africa because technological development takes place within conditions of unequal infrastructure, knowledge production and economic power. The African Union's Continental Artificial Intelligence Strategy already recognizes the

importance of an Africa-centred, ethical, responsible and equitable approach to AI.

At the same time, contemporary scholarship cautions that there is not yet a sufficiently coherent and universally representative "African AI ethics." Ubuntu offers important relational insights, but it should not be treated as a single or exhaustive African ethical system. The appropriate philosophical response is therefore neither rejection of Bostrom nor uncritical adoption of Ubuntu. What is required is a critical synthesis. Bostrom's concern with existential safety should be combined with African philosophical concerns for relational personhood, communal responsibility, epistemic justice, technological self-determination and human flourishing.

Finally, the problem of AI is not merely whether machines will become more intelligent than human beings. It is also whether human beings will remain capable of determining what intelligence should serve, whose values should guide technological development, and what kind of human future technology ought to promote. An African philosophical perspective contributes to this discussion by insisting that intelligence must remain connected to responsibility, technology to community, and technological progress to human dignity and justice.

References

African Union. (2024). *Continental artificial intelligence strategy*. African Union Commission.

Citation: Ekwere, Idorenyin "A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence". *Journal of People and Worldviews (JPW)*, 2026: pp307-319.

A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence

- Adams, R. (2021). Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46(1-2), 176-197.
- Aristotle. (2009). *The Nicomachean ethics* (W. D. Ross, Trans.). Oxford University Press.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, 2(2), 100205. <https://doi.org/10.1016/j.patter.2021.100205>
- Boethius. (1994). *Theological tractates and the consolation of philosophy* (H. F. Stewart, E. K. Rand, & S. J. Tester, Trans.). Harvard University Press. (Original work published ca. 523 CE)
- Bostrom, N. (2012). The superintelligent will: Motivation and instrumental rationality in advanced artificial agents. *Minds and Machines*, 22(2), 71-85. <https://doi.org/10.1007/s11023-012-9281-3>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14, 349-379. <https://doi.org/10.1023/B:MIND.0000035469.63578.9d>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gwagwa, A., Kazim, E., Kigozi, S., & others. (2022). The role of the African value of Ubuntu in global AI inclusion discourse: A normative ethics perspective. *Patterns*, 3(4), 100451.
- Gyekye, K. (1997). *Tradition and modernity: Philosophical reflections on the African experience*. Oxford University Press.
- Ikuenobe, P. O. (2006). *Philosophical perspectives on communalism and morality in African traditions*. Lexington Books.
- Ikuenobe, P. O. (2015). Relational autonomy and communalism in African philosophy. *Philosophical Papers*, 44(1), 99-121.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389-399. <https://doi.org/10.1038/s42256-019-0088-2>
- Jonas, H. (1984). *The imperative of responsibility: In search of an ethics for the technological age*. University of Chicago Press.
- Kant, I. (1993). *Grounding for the metaphysics of morals* (J. W. Ellington, Trans.). Hackett Publishing. (Original work published 1785)
- Mbiti, J. S. (1969). *African religions and philosophy*. Heinemann.
- Menkiti, I. A. (1984). Person and community in African traditional thought. In R. A. Wright (Ed.), *African philosophy: An introduction* (pp. 171-181). University Press of America.
- Metz, T. (2011). Ubuntu as a moral theory and human rights in South Africa. *African Human Rights Law Journal*, 11(2), 532-559.
- Metz, T., & Gaie, J. B. R. (2010). The African ethic of Ubuntu/Botho: Implications for research on morality. *Journal of Moral Education*, 39(3), 273-290. <https://doi.org/10.1080/03057240.2010.497609>
- Mhlambi, S. (2020). *From rationality to relationality: Ubuntu as an ethical and human rights framework for artificial intelligence governance*. Carr Center for Human Rights Policy, Harvard Kennedy School.
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417-457. <https://doi.org/10.1017/S0140525X00005756>
- Wiredu, K. (1980). *Philosophy and an African culture*. Cambridge University Press.
- Wiredu, K. (1996). *Cultural universals and particulars: An African perspective*. Indiana University Press.
- Yilma, K. (2025). Ethics of AI in Africa: Interrogating the role of Ubuntu and AI governance initiatives. *Ethics and Information Technology*, 27, Article 24. <https://doi.org/10.1007/s10676-025-09834-5>

Citation: Ekwere, Idorenyin “A Critical African Examination of Nick Bostrom's Conception of the Problem of Artificial Intelligence”. *Journal of People and Worldviews (JPW)*, 2026: pp307-319.