



Trust norms for generative AI data gathering in the African context

Abiola Joseph Azeez¹  and Tosin Adeate² 

¹Philosophy Department & Canadian Robotics and Artificial Intelligence Ethical Design Laboratory, University of Ottawa, Ottawa, ON, Canada

²Graduate School of Business Leadership, University of South Africa, South Africa

Corresponding author: Abiola Joseph Azeez; Emails: Aazee020@uottawa.ca; Abiolaaz045@gmail.com

Received: 29 November 2023; **Revised:** 11 March 2024; **Accepted:** 28 March 2024

Keywords: Afro-epistemic-moral theory; data gathering; generative AI; knowledge forms; trust culture

Abstract

Can trust norms within the African moral system support data gathering for Generative AI (GenAI) development in African society? Recent developments in the field of large language models, such as GenAI, including models like ChatGPT and Midjourney, have identified a common issue with these GenAI models known as “AI hallucination,” which involves the presentation of misinformation as facts along with its potential downside of facilitating public distrust in AI performance. In the African context, this paper frames unsupportive data-gathering norms as a contributory factor to issues such as AI hallucination and investigates the following claims. First, this paper explores the claim that knowledge in the African context exists in both esoteric and exoteric forms, incorporating such diverse knowledge as data could imply that a GenAI tailored for Africa may have unlimited accessibility across all contexts. Second, this paper acknowledges the formidable challenge of amassing a substantial volume of data, which encompasses esoteric information, requisite for the development of a GenAI model, positing that the establishment of a foundational framework for data collection, rooted in trust norms that is culturally resonant, has the potential to engender trust dynamics between data providers and collectors. Lastly, this paper recommends that trust norms in the African context require recalibration to align with contemporary social progress, while preserving their core values, to accommodate innovative data-gathering methodologies for a GenAI tailored to the African setting. This paper contributes to how trust culture within the African context, particularly in the domain of GenAI for African society, propels the development of Afro-AI technologies.

Policy Significance Statement

This article highlights the role of Afro-trust norms in facilitating and governing data-gathering processes for building an Afro-Generative AI. Exploring the significance of knowledge forms—esoteric and exoteric—in the African setting generates discussions around whether Afro-knowledge forms are AI data-worthy while providing a narrative of data-gathering challenges in Africa for Generative AI (GenAI) projects. This article undertakes a novel issue by focusing on the challenges of a GenAI data collector in the African setting and how possessing the knowledge of the Afro-trust norms, with much recalibration, can improve the relationship between the data collector and the community members toward an Afro-tailored Generative AI technology.



Introduction

In this paper, we examine the relationship between trust norms in the African moral systems and data gathering for Generative AI (GenAI) within African society. The rise of large language models like ChatGPT and Midjourney has brought attention to the issue of AI hallucination, which can undermine trust in AI. We aim to explore how trust norms impact data collection in the African context and their potential to shape the development of indigenous GenAI technologies in Africa, specifically, in data gathering. To set up the central focus of this paper, we frame unsupportive data-gathering norms as a contributory factor to the problem of AI hallucination, focusing on various narrative challenges that beset data gathering for AI technology projects, such as GenAI. The capability of GenAI is limitless. GenAI technologies can learn from existing artefacts and generate new, realistic creations that capture the characteristics of the training data without merely duplicating it. This technology can produce diverse forms of original content, including images, videos, music, speech, text, software code, and product designs. Most GenAI models are the Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), which we shall discuss in this section only. The VAE consists of an encoder network that maps the input data to a latent space and a decoder network that generates output data from points in the latent space, while the GANs consist of a generator network that generates synthetic samples and a discriminator network that tries to distinguish between real and synthetic samples (Dhoni, 2023). The main challenge of this paper, in the African context, is data gathering for GenAI projects, exploring in different sections the nature of African knowledge, narrative challenges of data gathering in Africa, and adaptable trust norms to facilitate an easy data gathering process. This paper makes a substantial contribution by examining how the integration of trust culture within the African context, particularly in the domain of GenAI for African society, can inform the development of AI development and further propel the development of indigenous AI technologies in Africa. Our contribution is significant given this contemporary era, where the significance of data has risen to unparalleled heights, vividly underscored by the inception of OpenAI's ChatGPT, an exemplar of Natural Language Models (NLMs) birthed from the cradle of data-driven AI (Sambasivan et al., 2021; Oubibi et al., 2022; Dhoni, 2023). Even though it is outside of this study's scope, it is worth noting that although the trust norms proposed in this study mainly resonate with African cultures, they do the same in other regions, particularly in Asian cultures.

What does Generative AI mean in the African setting?

There is a controversy around the notion of GenAI: This concerns whether GenAI is about generating new content, a not-seen-before content, or about humans building a new technology that does something extraordinary in the history of technology, a not-seen-before technology (Srbic, 2018; West et al., 2023). This controversy further shrouds what GenAI means in the African setting because the consequences of using AI in a purely African context are under-researched although there has been ample work on what AI should look like for the African audience. For more than half a decade, the debate within the Afro-cultural consideration of AI has focused on how designing AI for the African audience may not embrace Afro-existential norms such as communality, togetherness, and brotherhood. These studies include the misalignment between second-wave AI and Afro-existential (Azeez and Adeate, 2020), and marginalisation of non-Western knowledge systems in the study of AI ethics (Segun 2021). In more recent times, however, this framing has expanded, encompassing issues of algorithm colonisation of Africa (Birhane, 2020), the importance of local data and knowledge (Abebe et al., 2021), empowering local talent (Ade-Ibijola and Okonkwo, 2023), and technological colonialism (Ugar, 2023).

Pan Dhoni (2023) describes GenAI as a technology that can learn from existing artefacts and generate new, realistic creations that capture the characteristics of the training data without merely duplicating it (Dhoni, 2023, 7). In general, GenAI is a type of technology that can produce diverse forms of original content, including images, videos, music, speech, text, software code, and product designs (ibid). Predominantly, within the domain of GenAI, two prominent models stand out: Variational Autoencoders (VAEs) and GANs. These models are pivotal in the advancement of AI due to their unique architectures and functionalities.

The Variational Autoencoder (VAE) is structured around a dual-network framework. One component, known as the encoder network, serves to map input data into a latent space. This latent space acts as a compressed representation of the input data, capturing its essential features and patterns. In contrast, the decoder network reconstructs output data from points within this latent space, effectively generating new data samples that closely resemble the original input.

On the other hand, GANs operate on a fundamentally different principle. GANs comprise two interconnected neural networks: a generator and a discriminator. The generator network's primary function is to synthesise artificial data samples by transforming random noise into meaningful representations. Simultaneously, the discriminator network is trained to differentiate between real data and the synthetic samples produced by the generator. Through an adversarial process, where the generator aims to fool the discriminator while the discriminator improves its discernment skills, GANs iteratively enhance the realism of their generated outputs.

By exploring the intricate workings of Variational Autoencoders (VAEs) and GANs, we gain valuable insights into the diverse methodologies employed within the landscape of General Artificial Intelligence. These models not only facilitate the generation of realistic data representations but also contribute significantly to various AI applications and advancements.

In the African setting, GenAI can be subsumed under the general notion of AI for ontological categorisation. In other words, a form of AI—whether recommender systems, GenAI, or similar systems—emerges from the Western individualist ontology and it does not negate the idea that AI predominantly has a non-collective ontological basis. Therefore, reacting to GenAI in an African setting would mean that GenAI is non-relational, non-communal, and non-collective and, as such, would not reflect some of the core characteristics of African thought. This would further imply that GenAI lacks existential, moral, and epistemic grounding in African society. However, an African-tailored GenAI would mean to the African audience, a technology that can interact at the esoteric and exoteric levels of knowledge. In other words, it must possess data that are both suitable to adapt to the public and restricted to a close and small society. Speaking of data restricted within a close society, Fayemi and Azeez (2021) have described it as epistemic occlusion, emphasising that it is, as seen in Barry Hallen's work, a form of data that is not available for public consumption. Recognising the potential implication of epistemic occlusion, this paper underscores the frustration of building a robust GenAI technology tailored to the African setting, citing the inaccessibility to the esoteric forms of knowledge. Nevertheless, this study explores how the norms of building trust in the African setting can help a GenAI data collector interact with the custodians of both esoteric and exoteric knowledge, towards building a robust African-tailored GenAI technology.

Is knowledge in the African context Generative AI data-worthy?

Epistemology is generally regarded as the theory of knowledge (Pritchard, 2016; Bewaji 2007). It is about what we know, the nature of knowledge, and the process leading to it. Before we claim to know, the medium of validating that belief or the process that affirms that certainty is important to the enterprise of epistemology—no wonder knowledge and justification is a focus in epistemology (Dancy, 2005). However, it is not stated when knowledge along with its process of acquisition becomes data. Speaking about data, we refer to raw facts, observations, measurements, values that have little meaning, or the basic building block of information and it typically in various formats such as numbers, text, images, speech, and so on.

For African epistemology, the idea of community is central to its focus, that is, knowledge and justification. This is so, because of the nature of knowledge in the African philosophical traditions. Similarly, the idea of communalism is an essential reference in understanding African epistemology. As with community, the nature of knowledge in African thought entails a system that recognises communal relations in how beliefs are identified, shared, and certified to be true. The communal relation is grounded on the idea of interconnectedness, which is a dominant feature of many African thought systems. Traditional African societies emphasised the relationship and communion among its members. What is

worth challenging is whether data can be arrived at through such a feature as interconnectedness. For what is known, data are described as pieces of information, and in most cases, units of information. The community system is essential to the African social system, and it is believed that there can be a developed human being without a community. An illustration of the ‘Robinson Crusoe’ model of human development will help clarify. The case of Robinson Crusoe on a deserted island that lacks human intervention in a process of growth portrays a scenario that shows that mental capacity and mental development, moral growth, and the acquisition of skills and etiquettes of social relations are acquired and nurtured not through the individual rational forte alone but social interaction and engagement with the community. However, it needs to be pointed out that learning communal ethics and values does not rub the individual off her existential capacities such as self-awareness, choices, personal responsibility, and so on. Therefore, the idea of togetherness and learning from one another is not against individual autonomy, but to polish our sense of community.

Recognising the protection of individual autonomy may yet suggest that communal knowledge can be converted into units of data. In our opinion, the foregoing can further suggest an awareness of the need to respect and protect individual autonomy, highlighting the significance of personal freedoms and choices. At the same time, it hints at the possibility that, even when prioritising individual autonomy, communal knowledge might be considered for conversion into data units. This perspective reflects a nuanced approach that seeks to balance the safeguarding of individual rights with the potential benefits of organising communal knowledge in a measurable and structured way, perhaps through data representation. This statement implies an openness to exploring ways in which collective wisdom can be systematically analysed or categorised while maintaining the importance of individual autonomy. Nevertheless, this is not explicitly stated, however, it is worth considering as possible. The emphasis on community affirms that the epistemological process or adventure must involve what the people can confirm to be true and affirmed to be useful for the well-being of the community of humans among others. The ideas stand out in the above. Togetherness for the acquisition and affirmation of knowledge, and collective relevance of what is claimed to be knowledge, in such a way that the ontological conditions of the claim to knowledge such that people’s existential condition is not threatened, and moral strength of the community not weakened. Knowledge in African thought, therefore, holds both theoretical and practical values; it is not knowledge for knowledge’s sake, it must possess moral value and bear the burden of promoting humanity. We will address the moral/practical aspect of knowledge first, to show what make data or knowledge valuable in the African context before continuing attention on the issue of knowledge validation through the idea of togetherness.

To address the moral relevance of the knowledge system in the African space, there is a need to discuss the emphasis on values—moral values, and human values in African thought. There are social expectations. The idea of social expectation connotes the standards and values set by the community or society, that its members are supposed to adhere to or maintain. In African society, certain norms and morals are written and conventions in the community. These norms guide individual private conduct, and public interaction, with fellow members and society. An example of this is the way the community frowns at the actions and thoughts of taking one’s life by considering life as sacred. While taking one’s life is an issue of private affairs that does not affect the community directly, the sacredness attached to individual life is a known phenomenon that the community values. The notion of sacredness suggests that there might be a conflict between data as information units and efforts gathering data in the African setting. This is owing to the assumption that not all data in the African setting are to be made available for public consumption. In reality, some data would be regarded as exclusive to a small circle of people (Fayemi and Azeez, 2021). Arising from the foregoing, knowledge must have a humanistic value. It must promote what is good within the cultural context and not promote actions considered antithetical to life and human flourishing. Human advancement in science and technology must not only promote the preservation of life but also must possess an epistemic virtue that makes individual life comfortable, and productive, while not undermining the capacity for self-development through labour. In African philosophy as Chimakonam and Ogbonaya note, we go an extra mile beyond understanding the nature of knowledge and justification to, first backwards, investigate the components of epistemology and the ontology of knowledge; and then

forward, to investigate the interrelatedness of knowledge and the moral dimension of knowledge (Chimakonam and Ogbonnaya, 2021, 203).

Having considered the point that knowledge must produce a moral appearance that keeps the individual and community members in African thought, we will now return to the discussion on the justification of knowledge and the place of community in it. Knowledge in African thought emphasises the place of joint action and testimonies of other epistemic agents to be validated. It is grounded on the philosophy of interconnectedness aptly captured in John Mbiti's dictum: "I am because we are and since we are, therefore, I am" (Mbiti, 1970: 141). It is this philosophy of interconnectedness that influences how Africans think about numerous questions ranging from the political, moral, and social, to name a few. It demonstrates the significance of mutual relationships as an essential element of African ethical theories. The pattern of reasoning on issues centres around thinking from the concern of the collective to the individual. This is why traditional African epistemologies are largely communal and their mode of justification relies on the coherent relation in the testimonies of the collectives (Adeate and Sewchurran, 2023, 11). The idea of interconnectedness in African epistemology for validation does suggest that an individual does not have the capacity for knowledge but for opinions or beliefs. For instance, does it mean that a person cannot hold the belief that he dropped a plate on the dining table, and be justified for having that belief because it is true that he left a plate on the set after eating? The significance of the African mode of knowledge to epistemological development is that toeing the line of a third-party agent validation gives more credibility to the knowledge claim. It is essential for the acceptance of beliefs we intend to recommend for others and not for ourselves alone.

Knowledge in African thought does not only give less priority to self-verified or self-acclaimed facts but jointly verifiable claims or facts, it recognises the importance and contribution of different aspects of an idea, approach or source of knowledge and largely viable contributors to the quest for consumable knowledge. This affirms why in African epistemology, epistemes are a derivation of both sense experience and reason, which gives a sense of seeing rationalism and empiricism, as not distinct sources of knowledge but as complements. Both find their fulfilment at the point of recognising their interconnectedness and interaction to produce viable knowledge claims. The nature of interconnectedness wired in African thoughts and serving as a justification of its beliefs disavow it from engaging in sensory experience alone as a medium of knowledge acquisition. Objects of sensation are subject to reflection, first by the individual perceiver, and rely on other epistemic agents for validation. The above affirms why Chimakonam and Ogbonnaya (2021: 179) define African epistemology as a subject that "concerns itself with a different approach or method to the study of the meaning, nature, scope, sources, theories, limitations and veracity of knowledge and non-knowledge". As perceived the role of joint action perhaps makes knowledge in an African setting GenAI data-worthy, for the fact that knowledge would undergo multiple verification tests until it is collected—this assures of the possibility of gathering good data for building a GenAI tailored to the African audience.

Central to the question of justification in knowledge is truth. The truth condition in African epistemology is based on the level of trust we have in the other not only as capable agents of knowledge but also trustworthy, that is they must possess both epistemic and moral virtues. Trust is built in the other because we recognise their capacity to deliver on the expectations we have of them. Trust is built because we depend on the role of the other in the actualisation of some of our life goals and needs or require that to help in the development of the tool and process that leads to the actualisation of those life interests. Mutual dependence is expected in this context to form how we think of trustworthiness. It therefore becomes doubtful that a non-human element can strike a deal for a relationship where the interest of both parties would be clearly stated and understood by both parties with the assurance that each one sees the other as trustworthy. It suffices, therefore, that a reasonable background exists to see trust as human property and trustworthiness as what is only meaningful within the human cycle. In the same manner, when we consume knowledge from AI, we must be sure it meets the required conditions for knowledge. Also, and more, importantly, we must be sure they are important trusted epistemic agents in the exercise of knowledge validation and production. However, crucial to this study is whether AI tailored for African audiences is trained on good African data, specifically GenAI for African audiences.

Data gathering challenges in Africa for Generative AI: The narrative

The significance of data in the process of building GenAI technology has risen to an unparalleled height, data are a critical infrastructure necessary to build AI systems, that is, GenAI (Sambasivan et al., 2021; Oubibi et al., 2022; Dhoni, 2023). Global conversations around this process recommend the use of large amounts of data to ensure output accuracy (ibid). While data gathering is a universal challenge but minimal in certain non-African regions such as the US, Canada, EU, and China, it is a more deep-seated issue in the African context for various reasons subsumed under the lack of a structured data ecosystem (Ade-Ibijola and Okonkwo, 2023), along with its specific implications. This section will explore some of these implications, including data cascades—e.g., lack of data quality (Sambasivan et al., 2021), unavailability of digital data (Oubibi et al., 2022), cost-induced data hoarding (Anane-Sarpong et al., 2018; Abebe et al., 2021), datasets inaccessibility (Okolo et al., 2023), anecdotal data (Gwagwa et al., 2020), the inability of the government to guarantee data security (Arakpogun et al., 2021). This section aims to narrate the formidable challenge of amassing a substantial volume of data, which encompasses esoteric information, requisite for the development of a GenAI model, tailored for African society. In addition, this section proposes, given existing challenges, a minimally substitutive approach to establish a foundational framework for data collection, rooted in trust norms that are culturally resonant, and that have the potential to engender trust dynamics between data providers and collectors.

Research has shown that the lack of a structured data ecosystem can have significant consequences, as highlighted by Ade-Ibijola and Okonkwo (2023), who emphasise the importance of data reflecting the demographic variables of the targeted population for the success of AI systems. Describing this phenomenon, Ade-Ibijola and Okonkwo (2023) capture this: “An AI will fail if the data that is used to train the AI system does not reflect the demographic variables in the targeted population. A Chatbot system, for example, requires comprehensive information about its operations to provide correct responses to users; if the information requested by the user is not in the data bank, the system will fail. Data shortages in Africa are well known in the context of development, where high-quality data are essential indicators of growth relating to the Sustainable Development Goals (SDGs) and a key input for the development of modern technologies.” (2023, 106–7). Responding to Ade-Ibijola and Okonkwo’s submission, this paper argues that limiting the lack of a structured data ecosystem in the African context to a mixture of these constitutive factors: data mismatch with demographic variables, inadequate information, and data shortages in Africa, might result in unavoidable implications in African contexts, such as oversimplification of reality variables, real-world deviations, unrealistic assumptions, risk of biases, and a potential limited predictive power. This paper holds this different view given that notable factors are excluded from the constitutive elements of a structured data ecosystem, including the absence of a robust data accessibility, and sharing policy, the dearth of AI talents in African society and so on. Nevertheless, data mismatch with demographic variables in African contexts is a major issue that cannot be overlooked because it contains significant insights that can positively impact efforts in addressing the lack of a structured data ecosystem in Africa.

The issue of data mismatches with demographic variables in African contexts has been articulated diversely in prior research endeavours, such as data cascades. Data cascades, in Sambasivan et al. (2021), are defined as compounding events causing negative, downstream effects from data issues, resulting in technical debt over time (2021, 2). What is worth noting, as drawn out by this paper, is that when problems with data accumulate and keep causing more problems, it leads to a kind of ‘technical debt’ that gets bigger over time. It is like a snowball effect, where one issue leads to another, and if they are not addressed, they can become more difficult to fix. Common causes of data cascades emerge from a larger problem of broken data practices, methodologies, and incentives in the field of AI (ibid), and manifest as poor record keeping, that is, some African countries lack proper record storage of dataset points such as date of birth, materials dated references and so on, and inconsistent data entry practices which often make cumbersome the process of sharing information across networks.

This foregoing issue implicates the unavailability of digital data (Oubibi et al., 2022) and aligns with the negative consequence of inadequate information (Ade-Ibijola and Okonkwo, 2023). In response to the

problem of inadequate information, only a few African countries are addressing the unavailability of digital data. These handful of countries are approaching digital data by setting up centralised data platforms and drawing insights from open data and big data infrastructures to build a localised data platform for African society. They establish these infrastructures by adopting principles and fundamental practices that govern the efficient establishment and citizen-centric delivery of data services. In Morocco, for example, the government has launched its portal, data.gov.ma containing hundreds of files on the various public departments. These data are sent to all users, researchers, investors, students, journalists, and other governments, and these data are freely available to everyone without restrictions. Saddled with the governance of this national operation is the Moroccan Data Protection Authority (CNDP) governed by multiple texts, and the objective of these texts and rules is the control of collecting, processing, and storing personal and sensitive data to protect privacy and accessibility and limit the divulgation and the use of these data in the whole data processing operations (Oubibi et al., 2022). What is worth noting is that most African countries do not have a robust data protection authority that is making a deliberate effort to digitise data for easy accessibility for technology projects, that is, the inability of the government to guarantee data security (Arakpogun et al., 2021). In turn, this has encouraged the participation of private enterprises and data entrepreneurs, and consequently, created a breeding ground for cost-induced data hoarding (Anane-Sarpong et al., 2018; Abebe et al., 2021) and in some cases, shared unless on merchandise grounds.

In the African context, data hoarding is commonly expressed as “You cannot collect data using your resources and put it on open access” (Anane-Sarpong et al., 2018). There is no doubt that data sharing contributes the largest, most efficient source of scientific data, but is fraught with contextual challenges, which makes stakeholders, particularly those in under-resourced contexts hesitant or slow to share. This, in other words, means that sharing data is a substantial and highly effective source of scientific data. However, some people, especially in places with limited resources, are cautious or take their time when it comes to sharing data because of the challenges involved, possessing the intrinsic potential of delaying the commencement of data-driven technology projects. In most cases, this results in a lack of dataset accessibility (Okolo et al., 2023) to African researchers and the relevance of this data to African problems in domains such as voice/text recognition and so on. This issue is further exacerbated despite initiatives like the Inclusive Images Challenge from Google aiming to improve the representation of imagery from the Global South but face the challenge of fully representing the vast diversity within the African continent. This stresses the importance of local communities within the African continent being involved in the creation, sharing and use of datasets, where anecdotal data (Gwagwa et al., 2020) are common.

In general, anecdotal data refer to information, narratives, or accounts that are based on personal experiences, individual observations, or specific, often unverified, incidents rather than systematic research or comprehensive data collection. In the African context, anecdotal accounts rely on storytelling and oral traditions, some of which are esoteric forms of knowledge. These forms of knowledge, inaccessible to the public, are prevalent in many African cultures, and they include stories, legends, and first-hand experiences that convey knowledge, cultural practices, or historical events. In making data-driven decisions for Generative AI projects in Africa, it is essential to consider the limitations of anecdotal data, and when necessary, complement it with more systematic and empirical research to obtain a more complete picture of a given situation. It is worth noting that anecdotal data are not representative of broader trends or provide a comprehensive understanding of complex issues; however, they emerge from valuable narratives, which help preserve cultural heritage and traditions. It is to this end that this paper proposes, in the absence of serious government intervention in establishing a structured data ecosystem, a minimal approach to establishing a foundational framework for data collection that has the potential to engender trust dynamics between data providers and collectors, which would be rooted in trust norms that are culturally resonant with the African people. This proposal aims to engage with individuals on a cultural level and advance technological innovation as extensively as possible, ultimately leading to the compressive integration of the state apparatus into the data-driven technology ecosystem in African society.

Trust norms for Generative AI data collector in African setting: some recommendations

In this section, we examine the intricate concept of epistemic occultism (Fayemi and Azeez, 2021), a term previously introduced in Section 2 as a formidable barrier impeding the acquisition of data crucial for the evolution of GenAI. By drawing upon the contextual backdrop provided by the narrative of a GenAI data gatherer expounded upon in the preceding section, our objective here is to delve deeper into the complexities surrounding epistemic occultism. We endeavour to elucidate the multifaceted challenges it imposes on the process of GenAI data gathering, thereby offering a more nuanced understanding of its impact on the advancement of artificial intelligence. After this analysis, we will present actionable recommendations that are both descriptive, providing insight into the current situation, and normative, suggesting pathways towards mitigating the adverse effects of epistemic occultism on GenAI data acquisition efforts.

At the traditional layer of modern African society, what is the main challenge of an AI Data collector? We argue that it is access to both forms of knowledge in African society, the esoteric and exoteric forms. The problem of access raises several questions, and one of them is thus: Arising from the idea of knowledge validation in African thought and the emphasis it places on truth testimonies of collective agents, to what degree can we accept or trust the knowledge claims of Generative AI? This study acknowledges that there is no justification to accept or trust the output of Generative AI tailored to African settings because esoteric information is excluded. So far, we have mostly focused on the exoteric form of knowledge; it is the type of knowledge suitable to be imparted to the public. Collecting this form of information is not a strenuous endeavour. It is freely available; however, it can be an object of disinformation when adulterated in any case. On the other hand, the esoteric form of knowledge is difficult to come by and inaccessible to the public. It is exclusive to a few and preserved for special occasions. Fayemi and Azeez (2021) have described this form of knowledge as epistemic occlusion, having traced it to the potential misinterpretation of Barry Allen's treatment of the Yoruba people's moral theory of epistemology.

Epistemic occlusion, in Hallen's view, means reserving knowledge to certain individuals who usually belong to close and sacred associations (Fayemi and Azeez, 2021, 93). In the case of Hallen, it alludes to the group of *oniseguns*, who are trusted as custodians of 'imo' based on their representation knowledge and moral embodiment (ibid). They are trusted to direct communities with their trusted divine instruction from the gods. They are also perceived as the ones whose words should be trusted because, by their *persona*, they bridge the spirit world and the world of the living (ibid). Hallen, however, misconstrued the possibility of the *oniseguns* keeping information from the populace by assuming that the populace lacks the spiritual epistemic quality to liaise with divinity. Hallen's view is common across Africa, and it is perhaps a major challenge to AI data collectors. This is because how do you build Generative AI on exoteric data only, excluding esoteric data? This section proposes a few norms to facilitate trust between collectors and custodians of knowledge in African society.

Trustworthiness is increasingly becoming an important factor in ensuring the safe deployment of artificial intelligence (AI) technologies, specifically, given the recent advancements in large language models (LLM), that is, Generative AI, such as ChatGPT, Midjourney, and so on. Aside from this, trustworthiness holds the fabric of the society together and facilitates openness in the society among the people. This is to say that there are norms of trust in African society, rooted in moral and epistemic values, that promote cordiality and relationality among Africans. We believe that such norms, as mentioned in the first section of this study, can help the AI data collector interact with the custodians of esoteric and exoteric data—the community at large and gather adequate information required for building a Generative AI tailored to an African audience.

Recommendations: setting up trust norms

- (1) Understanding Customary Practices in African Society: The Generative AI data collection role must be given to individuals who understand the moral and epistemic constructs of public and

private knowledge of African society. This is to ensure that the individual is positioned to understand the people's norms of doing things. In many African societies, it is customary for individuals to familiarise themselves with the community's ways before establishing relationships. This includes adhering to the normative manners that guide the acquisition of knowledge. Demonstrating such knowledge, in seeking validation from the community, involves an acknowledgement that the individual understands the customary conditions associated with pursuing knowledge in that community. The primary condition is to cultivate a social relationship with the community to secure their validation before undertaking any data-gathering initiative.

- (2) **Establish Technology Groups Supported by Indigenous Funding:** Community members are more likely to feel secure sharing information when they know that the AI data collector belongs to a group established by and for their indigenous community. The objective is to empower indigenous individuals and motivate them to contribute to technology development as a means of giving back to their community.
- (3) **Joint Verification:** This recommendation rests on two relational norms: interconnectedness and togetherness. When the AI data collector integrates into African society, it is anticipated that they approach their relationships with community members as a collaborative partnership. In essence, the individuals within the community are regarded as partners in the endeavour of acquiring knowledge until they willingly participate in the collector's data-gathering efforts. The objective is to foster a perception of mutual partnership, encouraging them to see themselves as collaborators and make themselves available as verifiers during the data collection process.
- (4) **Affirming the Sacredness of Information:** Sacred data are necessary for building robust information-making and producing AI, that is, Generative AI, for the African setting. The data collector should establish transparency with the community, ensuring that they are aware of the collector's reverence for sacred information. In many instances, the data itself may not be considered sacred; rather, those safeguarding esoteric information seek to assess the collector's respect for their cultural practices and determine if they are worthy of receiving such information. An illustrative example is Barry Hallen's experience with the Yoruba people, wherein his respectful approach earned their trust, enabling him to gather valuable research information for his book.
- (5) **Recalibrating Trust Norms in Adapting Technological Integration to Shaping African AI:** To effectively implement a GenAI tailored for the African context, it is crucial to recalibrate trust norms in a way that aligns with contemporary social progress while preserving the core cultural values. This recalibration should involve a thoughtful accommodation of innovative data-gathering methodologies. In navigating the intersection of tradition and technology, an approach that respects the unique norms of African societies, yet embraces the potential of cutting-edge data collection techniques, can foster a harmonious integration of AI into the African setting. This intersection will contribute to how African trust norms could inform AI development and further propel the development of indigenous AI technologies in Africa. Balancing these elements is pivotal for ensuring that the development and deployment of AI technologies not only are culturally sensitive but also contribute positively to the societal progress and values of the diverse African landscape.
- (6) **Bridging Ancient Wisdom by Integrating African Traditions into Data-Friendly Formats:** This complex process involves encoding intricate cultural insights and practices from Africa into structured digital information. It requires carefully cataloguing and organising various forms of indigenous knowledge, such as oral histories, ritual practices, herbal remedies, and cosmological beliefs, into easily accessible databases and digital repositories. By collaborating closely with local communities and scholars, this transformative endeavour aims to preserve and share valuable cultural heritage while utilising modern computational tools for analysis and practical application. By combining ancient wisdom with contemporary data science techniques, this initiative aims to uncover new insights, promote cultural exchange, and empower communities to utilise their heritage in addressing present-day challenges and opportunities.

Conclusion

This study examined the relationship between trust norms in the African moral systems and data gathering for Generative AI (GenAI) within African society. We explored how trust norms impact data collection in the African context and their potential to shape the development of indigenous AI technologies in Africa, specifically, in data gathering, with a focus on the inaccessibility to esoteric forms of data in African setting. To set up the central focus of this paper, we frame unsupportive data-gathering norms as a contributory factor to AI hallucination, focusing on various narrative challenges that beset data gathering for AI technology projects, such as Generative AI. The capability of GenAI is limitless. The main challenge of this paper, in the African context, is data gathering for GenAI projects, exploring in different sections the nature of African knowledge, narrative challenges of data gathering in Africa, and adaptable trust norms to facilitate an easy data gathering process. This paper makes a substantial contribution by examining how the integration of trust culture within the African context, particularly in the domain of Generative AI for African society, can propel the development of indigenous AI technologies in Africa. In addition, we provide four recommendations that could help the GenAI data collector build norms with the community, focusing on building a social interaction with the custodians of esoteric and exoteric data and earning the respect of the community people. Our contribution is significant given this contemporary era, where the significance of data has risen to unparalleled heights, vividly underscored by the inception of OpenAI's ChatGPT, an exemplar of NLMs birthed from the cradle of data-driven AI.

Data availability statement. None.

Acknowledgements. None.

Author contribution. Abiola Joseph Azeez (AA) was responsible for the conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, and validation. AA also contributed to resources, software, and writing both the original draft and subsequent review/editing. Tosin Adeate (TA) contributed to data curation, resources, software, validation, and writing the original draft, specifically the section on African Epistemology. Tosin Adeate (TA) also contributed to reviewing the abstract and introduction. For a clearer appropriation of contributions, please refer to the key terms below. Conceptualization—AA; Data curation—AA, TA; Formal Analysis—AA, TA; Investigation—AA, TA; Methodology—AA; Project administration—AA; Resources—AA, TA; Software—AA, TA; Supervision—AA; Validation—AA, TA; Writing—original draft—AA, TA; Writing—review & editing—AA.

Provenance. This article is part of the Data for Policy 2024 Proceedings and was accepted in Data & Policy on the strength of the Conference's review process.

Funding statement. None.

Competing interest. None.

References

- Abebe R, Aruleba K, Birhane A, Kingsley S, Obaido G, Remy SL and Sadagopan S (2021) Narratives and counternarratives on data sharing in Africa. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 329–341.
- Adeate T and Sewchurran A (2023) African epistemologies and the decolonial curriculum. *Acta Academica* 55(1), 1–19. DOI: <https://doi.org/10.38140/aa.v55i1.5762>
- Ade-Ibijola A and Okonkwo C (2023). Artificial intelligence in Africa: emerging challenges. In Eke DO, Wakunuma K and Akintoye S (eds), *Responsible AI in Africa: Challenges and Opportunities*. Cham: Springer International Publishing, pp. 101–117.
- Anane-Sarpong E, Wangmo T, Ward CL, Sankoh O, Tanner M and Elger BS (2018) “You cannot collect data using your own resources and put it on open access”: perspectives from Africa about public health data-sharing. *Developing World Bioethics* 18(4), 394–405.
- Arakpogun EO, Elsahn Z, Olan F and Elsahn F (2021) Artificial intelligence in Africa: challenges and opportunities. In *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*, pp. 375–388.
- Azeez A and Adeate T (2020) Second-wave AI and Afro-existential norms. *Filosofia Theoretica: Journal of African Philosophy, Culture and Religions* 9(3), 49–64.
- Bewaji JAI (2007) *An Introduction to the Theory of Knowledge: A Pluricultural Approach*. Ibadan: Hope Publications Press.
- Birhane A (2020) Algorithmic colonization of Africa. *SCRIPTed* 17, 389.

- Chimakonam JO and Ogonnaya LU** (2021) *African Metaphysics, Epistemology, and a New Logic: A Decolonial Approach to Philosophy*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-72445-0>.
- Dancy J** (2005) Problems of epistemology. In T Honderich (ed.), *The Oxford Companion to Philosophy*. Oxford: Oxford University Press, pp. 263–265
- Dhoni PS** (2023) Exploring the synergy between Generative AI, data and analytics in the Modern Age.
- Fayemi A and Azeez A** (2021) Epistemic unfairness in Barry Hallen’s account of agency in Yoruba moral epistemology. *Lagos Notes and Records* 27(1), 76–95.
- Gwagwa A, Kraemer-Mbula E, Rizk N, Rutenberg I and De Beer J** (2020) Artificial Intelligence (AI) deployments in Africa: benefits, challenges and policy dimensions. *The African Journal of Information and Communication* 26, 1–28.
- Mbiti JS** (1970) *African Religions and Philosophies*. Oxford: Heinemann
- Okolo CT, Aruleba K and Obaido G** (2023) Responsible AI in Africa—Challenges and opportunities. In Eke DO, Wakunuma K and Akintoye S (eds.), *Responsible AI in Africa: Challenges and Opportunities*. Cham: Springer International Publishing, pp. 35–64.
- Oubibi M, Zhou Y, Oubibi A, Fute A and Saleem A** (2022, January). The challenges and opportunities for developing the use of data and Artificial Intelligence (AI) in North Africa: case of Morocco. In *International Conference on Digital Technologies and Applications*. Cham: Springer International Publishing, pp. 80–90.
- Pritchard D** (2016) Epistemology. In Pritchard D (ed.), *What Is this Thing Called Philosophy?* London: Routledge. <https://doi.org/10.4324/9780203771006>
- Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh P and Aroyo LM** (2021) Everyone wants to do the model work, not the data work”: data cascades in high-stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–15
- Segun ST** (2021) Critically engaging the ethics of AI for a global audience. *Ethics and Information Technology* 23(2), 99–105
- Srbic D** (2018) (De)generative art: rules-transgressing algorithms. *Electronic Visualisation and the Arts*, 221–222
- Ugar ET** (2023) The fourth industrial revolution, techno-colonialism, and the Sub-Saharan Africa Response. *Filosofia Theoretica: Journal of African Philosophy, Culture and Religions* 12(1), 33–48.
- West P, Lu X, Dziri N, Brahman F, Li L, Hwang JD, ... Choi Y** (2023) The generative AI paradox: “What It Can Create, It May Not Understand”. arXiv preprint [arXiv:2311.00059](https://arxiv.org/abs/2311.00059).