



Resource Allocation for Trustworthy Artificial Intelligence Projects in African Context

Abiola Joseph Azeez, *Elnathan Tiokou*,
and Edmund Terem Ugar

INTRODUCTION

Can AI be trustworthy? What does it mean to have a trustworthy AI design in Africa? In this chapter, we answer the first question in the affirmative. Our central claim is that, even though trust and trustworthiness

A. J. Azeez (✉)

Canadian Robotics and Artificial Intelligence Ethical Design Laboratory,
University of Ottawa, Ottawa, ON, Canada

e-mail: aazeez@uottawa.ca

E. Tiokou

Polytechnique Montréal, Montreal, QC, Canada

e-mail: elnathan.tiokou@polymtl.ca

E. T. Ugar

Department of Philosophy, Centre for Africa-China Studies, University of
Johannesburg, Johannesburg, South Africa

e-mail: edmundu@uj.ac.za

always apply to agents that have the capacity for moral responsibility, trust/trustworthiness can also be extended to AI. However, our use of the term trust in AI differs from the trust we attribute to moral agents—beings capable of responsibility/accountability. We situate the trust in AI within the framework of institutional trust, that is, trust in the institution where the technology is designed.

The need for trustworthy AI design is a serious contention within the literature on AI ethics and government policies and frameworks on AI because of the role AI plays in our current social milieu (Agarwal & Mishra, 2021; Benjamins et al., 2019; Dignum, 2019; Mikalef et al., 2022; Peters et al., 2020; Schiff et al., 2020). The subject of trustworthy AI has received enormous contributions in the last two decades. For example, in 2022, Google Scholar had 11,000 hits on the search on the topic of “trustworthy AI” (Slosser et al., 2023). Many ethical AI guidelines and white papers that have emerged recently have trust as their central topic of discussion (e.g., Gunning et al., 2019; HLEG, 2019; Leslie, 2019; OECD, 2017; United Nations, 2021). While the central or paradigmatic nature of trust is deeply interpersonal, the worry for most discussions on trust is whether we can use trust relationships to describe our relationship with artificial intelligence. The main problem is that, in most instances, rather than a rationally appropriate motive, trust is always considered on the grounds of emotion, intuitions, and the personal history a “trustor” has with a “trustee”.

Some theorists proposed the notion of reliability, rather than trust, as that which can be suitably applied to artificial intelligence. These theorists argue that we can rely on AI rather than trust the system (Dur'an & Fromanek, 2018; Nickel et al., 2010; Ryan, 2020). The rationale behind their argument is that we can only trust humans because of their moral capacity for accountability, while technology, on the contrary, can only be relied on given its lack of praise- or blameworthiness. However, the problem with the above view is that technologies such as AI are not merely technical artefacts. AI technologies are socio-technical artefacts that interact differently with humans and are built in ways that fit into the rules of human society (Ugar, 2023a). Furthermore, AI affects our lives differently from other technologies in areas like medicine and health-care, education, transportation and others (Benk et al., 2022; van de Poel, 2020). As a result, it is crucial that we do not conceptualise the relationship between humans and AI in terms of reliance as applicable to other technical artefacts since AI is more advanced.

In this chapter, we contend that while it is plausible that we cannot trust AI systems because of the moral and emotive implications of trust, relying on AI also does not capture the entirety of the relationship humans share with AI, given their social nature. Thus, we argue that we can trust AI, but from the perspective of institutional trust. While we explain later the meaning of institutional trust, in simple terms, it is the kind of trust that can be applied to non-human subjects within the framework of an institution with natural persons (Misztal, 1996). We adumbrate that for AI to be trustworthy, it must have natural persons who understand the capabilities and limitations of the systems and supervise their functioning. These natural persons ought to monitor overreliance on the system and understand and interpret the systems to approve or override decisions that do not align with the values of human society. Natural persons are entrusted with the role of interfering with the performance of this high-risk AI system, and they form part of the institutional web of trustworthiness.

However, one may ask: what does it then entail to trust an AI system? When we trust an AI, what exactly are we trusting? Are we trusting the technical artefacts, the institution, the rule or the decision makers? We argue that the trust relationship with AI encompasses the artefacts, the institution, the rule makers, and the decision makers as a unit. In line with the above contention, and to address the second question posited at the beginning of this introduction, we underscore that a trustworthy AI design in Africa is one whose institutions are based in Africa, including those providing technical infrastructure for design, algorithm, data storage, data generation, and so on.

African scholars contend that prevailing AI technologies exhibit Western bias that is unsuitable for the African milieu (Azeez & Adeate, 2020; Eke et al., 2023; Wakunuma et al., 2022). In addition, it is argued that because Africans have a historical trajectory that is embedded in colonialism and neo-colonialism, the inadequacy of funding from African governments not only undermines the perception of AI trustworthiness in the African context but also exposes the African population to the manifestation of technological colonialism. In response, this chapter argues that African governmental entities must adopt a proactive strategy by providing an enabling environment for AI development and innovation to thrive. This enabling environment includes the provision of adequate financial resources to fund AI projects within the continent. For example, the estimated global economic value of AI according to PWC is reported

to reach \$15.7 trillion in 2030. Even though AI is at its early stage, the financial benefits of AI are as follows: \$7.0 trillion for China, \$3.7 trillion for Northern America, \$1.8 trillion for Northern Europe, and \$1.2 trillion for Africa. Africa sits at the bottom of the economic gains from AI. Furthermore, in 2022, the US spent \$47.7 billion in AI investments, China spent \$13.4 billion, and Germany spent almost a billion euros; while in 2021, the European Union Council targeted an annual investment of €20 billion to AI, while Africa spent \$2.0 billion in 2021 and \$3.0 billion in 2022. Additionally, despite having low investment in AI, some of Africa's AI investments come from Western investors, such as the Google and IBM innovation labs and the Nigerian Kudi AI, which Silicon Valley funds. Given the low financial investment we have highlighted above, we aim to (i) draw on a comparative analysis of global AI advancement projects in Europe, the United States, China, and Africa, our claim is that to develop trustworthy social technologies that resemble the African worldview, the issue of funding deficit must be addressed to advance trustworthy AI research which prioritises setting up trustworthy AI design parameters in the African context and (ii) show how this funding gap obscures and shapes the conversation on design parameters and variables of a trustworthy AI with an African outlook—that is, we show how issues like funding biases can obscure a trustworthy AI design from Africa for Africans.

In other words, an African trustworthy AI, or an AI system that is trustworthy in Africa, is one that is designed in Africa, designed within African institutions, has the African Union as a regulatory body, and African values, norms, and lived experiences shape its policies and frameworks. To put this in simple terms, AI's trustworthy design in Africa must have an African agency at the centre of the design. Our claim is informed by the notion of institutional trust, which we find prize-worthy and which can be applied to the trust relationship we have with AI systems.

Three main reasons shape the novelty and significance of our argument to support the yardstick for a trustworthy design of AI systems in Africa, which is centred around African agencies. First, most of the current institutions responsible for AI designs deployed in Africa are not in Africa, and these systems are not shaped by policies informed by the cultural norms, ethos, worldviews, and ethics emerging from Africa. Second, given African histories of colonialism, neo-colonialism, and techno-colonialism, Africans ought to be wary of designs emerging from elsewhere. Furthermore, current designs of AI have shown evidence of biases and discrimination

against Africans, from misrepresentation of black people by facial recognition technology to discrimination in recidivism tools in the US. Lastly, given the previous reasons, on the one hand, and the view that the best version of trust that applies to AI has to be institutional trust, on the other hand, we submit that Africans must rely on their institutions that care for the well-being of Africans, to guide the design of a trustworthy AI.

We divide this paper into three main sections. In the first section, we delve into the philosophical analysis of trust and distinguishing trust from reliability. Additionally, we provide clear reasons why trust is important and why the institutional notion of trust can be applied to AI. The second section shows some of the impediments in Africa that are stumbling blocks to the design of AI in the continent to achieve trustworthiness. One of the problems that we identify is financial constraints. The third section argues that there is no alternative towards trustworthy AI design in Africa. We show what trustworthy AI ought to look like in Africa, encompassing African ontology, along with exuding the relationality and human-centeredness of the African person, and why only Africans can achieve this trustworthy design. Additionally, we make recommendations for AI designers and policymakers on the best approach to designing a trustworthy AI ecosystem in Africa, highlighting a nuanced perspective on channelling financial resources to advance AI projects in Africa, encompassing dormant fund utilisation, corporate social responsibility, partnerships, and community-driven initiatives towards fostering a trustworthy AI framework rooted in the African ethos.

A PHILOSOPHICAL ANALYSIS OF THE CONCEPT OF TRUST

What does trust mean? To understand if we can have a trusting relationship with AI, we must expose the meaning of trust. Trust is conceived as the relationship a party has with another party on the account that the former is willing to rely on the latter. The former party, that is one who trusts, is generally known as the trustor, while the party who is trusted is known as the trustee (Jones, 1996). Trust is usually based on the intuitions and histories of past experiences between the trustor and their trustee (Hardin, 2002; Hardwig, 1991). However, prerequisites of trust, like past experiences, are not generalisable to warrant trust or distrust objectively. I can trust Apple to produce their iPhone 16 Pro

Max with exquisite camera features because, based on my past experiences, Apple produces iPhones with the best camera qualities, and Apple has been reliable in this regard in the past. However, my experience with Apple cannot be a yardstick for anyone else to trust Apple since experiences are not the same. Given this, one can then claim that trust is subjective; that is, a trust relationship between a trustor and a trustee is based on the trustor's subjective experience with the trustee. However, it is pertinent to clarify from the onset that the notion of trust that we advance here is not based on subjective estimates, a notion that is ubiquitous in traditional accounts of trust (Bauer, 2019). Our notion of trust is a shared subjectivity. For example, the shared experiences of colonialism, neo-colonialism, capitalism, and techno-colonialism in Africa.

Annette Baier (1986) contends that trust is a concept loaded with normative attitudes. Baier identifies goodwill as an integral component of trust. In Baier's view, a trustor can rely on the trustee because of the trustee's goodwill. When a party expects goodwill from another, and the goodwill becomes non-existent, the trusting party might feel betrayed by the trustee. However, such a dynamic can only exist within the human community because it would be absurd to expect goodwill from non-human entities or for humans to feel betrayed if non-human entities do not meet the demands of goodwill. On this account, Baier believes that trust is an exclusively human attitude (also see Hawley, 2014; Holton, 1994; Jones, 1996). For us not to fall into the trap of ambiguities, it is pertinent that when we use concepts like "trust" in human-AI relationships, we clarify how trust is construed. It is absurd to think that non-human beings can be the object of trust. Why so? Non-human entities cannot be held accountable or responsible for their actions in the absence of goodwill, nor can they commit to upholding certain relationships with humans (Hawley, 2014; Holton, 1994; Ryan, 2020).

As a concept, trust is a mental attitude (Hardin, 2006; Jones, 1996; Sztompka, 1999). When discussing trust, we invoke underlying concepts like reason, emotion, and volitional/behavioural choice. As a result, trust relationships involve a trustor believing in a trustee's agency, feeling safe in their hands, and voluntarily placing their trust in the trustee (Baier, 2013). These three concepts must go concomitantly for a trust relationship to be established. Because of the critical role of agency in a trust relationship, one wonders if non-human entities like AI can be trusted, as expatiated above.

In the AI trustworthy literature, some theorists argue that instead of thinking about AI along the lines of trust, we should rather think about these technologies along the line of reliability (see, for example, Dur'an and Fromanek, 2018; Nickel et al., 2010; Ryan, 2020). Reliability is an alternative to trust or a weak form of trust (Baier, 1986). To put reliability as a weak form of trust into perspective, theorists like Katherine Hawley argue that trust converges with reliability in practical terms because a trust relationship involves a practical reliance on the trustee (Hawley, 2012).

However, the point of divergence between trust and reliance is that reliance is just part of a structural composition of trust on the account that a trustor relying on a trustee does not necessarily require any prior attitude of trust. Furthermore, trust has a moral quality that is not present in reliance (Lagerspetz, 2015). For example, I can rely on my cat to chase out mice from my house. However, my cat does not owe any moral duty to chase out mice from my house, and I cannot hold my cat accountable if it does not chase out mice. In other words, I rely on my cat for practical reasons without any moral obligations attached. While we can rely on animals like dogs and cats or inanimate objects like technology, trust can only be directed to objects to which we can attribute moral responsibility and agency in the trust relationship. Thus, saying a “trustworthy AI” will be missing the point. If we are to introduce the concept of “trustworthiness” to AI, there has to be trust in the AI ecosystem involving humans as agents responsible for praise or blame (Coeckelbergh, 2012). We will return to the above point shortly. Let us briefly engage with the concept of reliance as it applies to AI systems.

The literature on interpersonal trust uses reliance as a weaker version of trust that can be used to describe human-non-human (AI) relationships. For instance, theorists like Dur'an and Formanek (2018) describe human-AI relationships as computational *reliabilism*, while Ryan (2020) believes that such a relationship is based on rational trust, and Nickel and colleagues (2010) argue that it is a thin notion of trust. Even though these concepts are not interchangeable, they have some similarities that warrant that we can use them interchangeably in this chapter. Thus, we group them all into one notion: reliance. We use Ryan's notion of rational trust to expand on them.

According to Ryan (2020), AI cannot be trusted because they do not possess the capacity to be trusted. Ryan alludes that trust requires attribution of some affective and/or normative reasons to act; since AI does not possess any emotive quality or cannot be held responsible for its action, it

follows that AI cannot be an object we can attribute trust to (Ryan, 2020, p. 2). Ryan concludes that the only legitimate trust that can be placed in AI should be rational. Ryan exposes rational trust as a situation in which a trustor depends on the trustee regarding certain specific actions, irrespective of whether the trustee is motivated to carry out such actions. As a result, Ryan underscores that the most rational thing to do in human-AI relationships is for the former to rely on the latter rather than trust the latter.

Additionally, Ryan (2020, p. 13) holds that for trustor A to have a trust relationship with trustee B, B must be a moral agent who can recognise the trust A has for them. In this sense, a trustee qualifies to be trusted because they hold moral agency and can be subject to praise or blame. What Ryan introduces here is the requirement of full moral agency for the relationship of trust to take place (2020, p. 10). His rationale for the above is based on the premise that if an entity cannot be held responsible for their actions, it follows that they do not qualify as trustees.

However, we argue that there are exceptions to Ryan's argument. There are cases where one cannot trace the responsibility to a trustee, given that they cannot be responsible. Nonetheless, the institution where the trustee emerges from can be held responsible on behalf of the trustee. For example, I can trust my fourteen-year-old daughter to be responsible and disciplined at school. But if she burns down her school due to indiscipline, I would be held responsible as the parent. Here, my fourteen-year-old teenage daughter is a human being with sound faculties, yet she cannot be directly held accountable for some of her actions. My daughter and her family form the institutional web that collectively takes responsibility for her actions. This is what some theorists term institutional trust. Institutional trust allows us to trust objects, like AI, that are incapable of being held accountable for their actions.

Trust is imperative in human-AI relationships because of what AI represents in our social milieu (Cheng et al., 2021; Clarke, 2019). There are two broad reasons why we need trustworthy AI design and use. First, AI technology is different from other technologies. AI is a socio-technical system that is designed with essentially three building blocks: (I) technical artefacts—they are designed to perform specific technical functions with intentionality (Ugar, 2023a; van de Poel, 2020); (II) agency—the systems exercise some form of agency because they can adapt to their environment, possess semi-autonomy, and can interact with their users (Ugar, 2023a; van de Poel, 2020); (III) rules—AI technologies are designed to

follow the rules and social norms of their designers (Ugar, 2023a; van de Poel, 2020). In line with the above view, some thinkers ascribe a weak form of moral agency to AI systems (see Floridi & Sanders, 2004; Ugar, 2023a; van de Poel, 2020).

The second reason for a trustworthy design of AI is because of the role AI plays in various domains like healthcare, justice and criminal system, education, transportation, etc. We have designated various sensitive and intricate roles to AI in the abovementioned domains. In line with the above, the ethics guidelines of the European Union, published in 2019, underscore that trust is an essential and necessary condition for the successful implementation of AI in society (HLEG, 2019). Trustworthiness remains integral to most ethical guidelines and white papers published for AI in society (see Gunning et al., 2019; HLEG, 2019; Leslie, 2019). Furthermore, before we can speak of trustworthy AI in Africa, given how contentious trust is, we must establish how trust should be viewed when it comes to AI. As clearly stated, we place trustworthiness in AI within the scope of institutional trust. We argue that trustworthy AI is not based on trust in AI as a social tool, but a chain that encompasses the institution and agents capable of responsibility from where the technology emerges. Thus, our notion of trustworthy AI is based on institutional trust (for explanations of institutional trust, see entries like Bachmann, 2020; Foley, 2001; Gambetta, 1988; Jones, 2004; Lahno, 2001; Lehrer, 1999; Pettit, 1995).

To understand trust within the context of AI, we must understand the institution that designs the AI. Given that trust is a mental state, they must be an actor to whom the trust is directed towards. This is because of the feasibility of ascribing expectations and responsibility to the trustee (Mollering, 2006). Institutional trust is directed at a system or institutions based on their competencies, honesty, and ability to live up to their commitments. There is an interplay between trust and institution, in the case of AI, for the trust to be effective. For us to trust AIs and the institutions that design them, the institutions must be effective in acting as objects of our trust (Fuglsang & Jagd, 2015, p. 22).

When we trust an institution, it means that we have confidence in the institutional functioning, what it represents, and its performance over time (Mollering, 2006). For an institution to foster trust, it must ensure that the inbuilt rules, routines, and roles that form the basis of an institutional setting are trustworthy. When there is institutional trust, the trustor can see the institutional safeguards measures, decisions, and actions given

their level of transparency. When a technology is trusted, it is not the technology itself that is trusted, but the institution where the technology is designed. The extent to which a trustor can rely on a particular AI is influenced or affected by the level of trust the trustor has in the institution where the AI is designed. For instance, when an institution produces a technology that is flawed egregiously, it leads to trustors undermining the institution's trust (Bekker, 2021). To this point, we ask whether Africans can trust AI technology emerging from other climes.

We argue that besides the obvious reasons of colonialism and the after-maths of colonialism, which have thwarted the trust Africans have in institutions from other climes, especially Euro-America, current designs of AI have undermined the trust relationship between Africa and institutions in the global north due to evidence of biases and discriminations towards Africans and people of colour by their AI designs. For example, facial recognition software designed by these institutions has categorised Africans and people of colour as apes (McCullom, 2017) and has poor recognition of blacks and people of colour (Buolamwini & Gebru, 2018). Additionally, software used for recidivism in the US is biased and discriminatory towards black people, especially African Americans and people of colour (Angwin et al., 2016; Forrest, 2021; Greene, 2023). There were other instances of hand dryers not recognising black people and people of colour during the COVID-19 pandemic (Kennard, 2022). These are clear instances where trust for institutions from other locales by Africans has been undermined. Given the importance of trust in AI and the role that AI plays in our current technologically-driven ecosystem, in the next section, we argue that Africans can circumvent the abovementioned problems of bias and discrimination if they can be agents of their AI designs.

TOWARDS AN AFRICAN AGENCY IN A TRUSTWORTHY AI DESIGN

In the previous section, we argued that the best form of trust which captures AI is the notion of institutional trust, that is, trusting the institution which the AI is designed from. Additionally, we pointed out that it is imperative for AI to be designed in Africa to make them trustworthy. Moving forward, we show how Africans can be at the centre of their AI design and the implications this may have in their AI ecosystem.

One of the most important prerequisites for achieving a trustworthy AI design in Africa is funding. Making a case for the imperative of Africans funding their AI projects cannot be isolated from current issues bewildering efforts to create AI systems tailored for the African audience, with emphasis on several critical factors. This section delves into factors necessitating the imperative of crafting trustworthy AI solutions tailored for the African context, encompassing the examination of the three avenues for funding: (A) funding the design locally, (B) sourcing funds elsewhere, and (C) collaboration and partnerships, and the analysis of pros and cons of these avenues. The significance of Africans financing their AI projects could be considered parallel to the efforts of African-born AI Ethics scholars who are engaged in crafting AI solutions tailored to the African context. Some scholars frame the issue of crafting trustworthy AI solutions tailored for the African context from various perspectives such as misalignment between second-wave AI and afro-existential (Azeez & 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 & Okonkwo, 2023), and technological colonialism (Ugar, 2023b). This chapter acknowledges the issue of funding deficit as a challenge to financing trustworthy AI projects, in line with the notion of institutional trust, tailored to the African audience and demands embracing AI intervention for specific issues on the African continent.

The Onto-Existential Factor

Focusing on the claim of misalignment between second-wave AI and Afro-existential norms, Azeez and Adeate's (2020) analysis could situate the imperative of funding trustworthy AI projects tailored for Africa within an onto-normative context. Making their case, they argued that second-wave AI trends do not reflect the African norms of existence being factored into ordering algorithmic patterns that set up AI systems and programmes, and thus, AI cannot be ingrained within the trustworthiness framework in Africa. The reason for this, they submit, is that Afro-existential practices unsettle with the individualist principle which underlines second-wave AI and therefore, a conversation around the

development and application of communal interpretation of AI is important. Justifying this claim, they further argue that Western ethical patterns, which inform the features of second-wave AI, such as statistical patterns, smart algorithms, specialised hardware, and big data sets, emerge from individualist notions. Proposing a way forward, they submitted that codifying communal values of African society into machines and other forms of robotics should narrow down the conversation of global context to Afro-ontological intelligence. Achieving this, we submit, is a step forward in building a robust institution of trust in Africa as well as bringing African values into the conversation on AI designs in the world. For us, Azeez and Adeate make a valid argument, however, leaving out the case of financing trustworthy AI projects by codifying communal values of African society, which we find prize-worthy in building institutional trust, into machines and other forms of robotics.

The Cultural Factor

We acknowledge the foregoing challenge, tracing it to the analysis provided by Segun (2021), in a comparatively expansive work, where this issue of misalignment is framed within the African cultural context. Segun's analysis touches on key issues such as artificial moral agency, patience, personhood, social robotics, and the principle of explicability, focusing on how the African worldview influences how they are understood in the context of AI tailored for African society. Focusing on the critical issue, Segun presents the argument that the Afro-ethical system is a collectivist system and its normative principles, especially the construal of what makes a right or wrong action, rest heavily on a collectivist disposition, demonstrating the view that an AI-tailored for African must imbibe the principles of collectivism. The collectivist society ensures that from birth people are integrated into cohesive in-groups that prioritise harmony, loyalty, and mutual respect, emphasising the defining principle of relationality. The goal is to ensure that AI as an agent has an important place in the decision-making matrix. Like Azeez and Adeate, Segun reckons that an AI tailored for an African audience is far-fetched when the power to create is not with African builders. We engage further with this point shortly.

The Educational Factor

Scholars such as Ade-Ibijola and Okonkwo (2023) have reckoned that the nature of AI education provided for local African talents may not support the tailoring of AI systems for an African audience. This is owing to the idea that much of the AI education curricula is Western-oriented, which is often focused on the guiding notion of individualism—a contrary worldview to the African collectivist and relational worldview. The implication is creating a development gap between the demands of AI solutions for African problems and the nature of education inculcated by Africa’s local talents. At best, the local talents will only possess the theoretical skills required but lack the practical skills required for the development, implementation, and use of AI applications, unless they work within a Western-based environment. Making the case, Ade-Ibijola and Okonkwo (2023) argued that any project requires the right expertise to succeed, and AI is no exception. We argue that the right skills can only be given if Africa considers investing in an African AI education curriculum to produce AI engineers who can develop AI systems tailored for the African audience.

The Data Counter-Narrative Factor

Another concern that necessitates financing AI projects by Africans in order to gain robust AI institutions within the continent that can be trustworthy is the adverse implication emerging from non-African stakeholders leading the data-sharing conversation in Africa. The implication is that principles alien to African society are exported, silencing the principles driving the values of relationality and collectivism in African society. Assessing this situation, Abebe et al. (2021) made the case that although the datasets are often extracted from African communities, conversations around the challenges of accessing and sharing African data are too often driven by non-African stakeholders. These perspectives frequently employ deficit narratives, often focusing on the lack of education, training, and technological resources in the continent as the leading causes of friction in the data ecosystem. Therefore, they argued that these narratives obfuscate and distort the full complexity of the African data-sharing landscape. It is worth noting that continuing data access and sharing conversations by non-African stakeholders may frustrate research and policy design to alleviate poverty, inequality, and derivative effects in Africa (ibid.). We argue that tailoring AI projects to an African society requires finance to

build a robust data infrastructure, acknowledging that the significance of data in the process of building AI technology has risen to an unparalleled height—data is a critical infrastructure necessary to build AI systems, specifically for the African audience.

The Techno-Colonialism Factor

There is the argument that the West's algorithmic invasion impoverishes the development of local products while also leaving the continent dependent on Western software and infrastructure (Birhane, 2020; Ugar, 2023b). The colonisers' first encounter with Africa in history was characterised by unilateral power and domination over colonised people, as seen in Birhane's (2020) analysis, seizing control of the social, economic, and political spheres by reordering, and reinventing social order in a manner that benefits it. What is worth noting is that traditional colonialism is often spearheaded by political and government forces, and digital colonialism is driven by corporate tech monopolies—both of which are in search of wealth accumulation (ibid.). Sharing the same view, Ugar (2023b) argues that techno-colonialism specifically means the transfer of technology and its values and norms from one locale to another, which has become a serious concern with the advancement of socially disruptive technologies of the Fourth Industrial Revolution (4IR), like artificial intelligence and robots (ibid.). Drawing from Ugar's (2023b) analysis, it is pertinent for Africa to finance their trustworthy AI projects, acknowledging the fact that technologies are not abstractly designed but based on the experiences of human relations in the society where they are designed. This encompasses facts such as technologies that come with the values of their designers and the cultural orientation of the locus from which they are designed (ibid.). We allude that given the histories of technological colonialism and extortion of Africans by foreign tech companies, it is imperative for Africans to build institutions that they can trust to enable them to create their trustworthy AI designs.

The Project-Management Factor

Across the world, AI project owners face three main challenges, but Africa suffers the implications more compared to other regions of the world (Campos-Zabala, 2023). The first challenge has to do with the phenomenon of limited budget, and this refers to low internal funding

powers. In responding to this challenge, countries with clear funding strategies and planning consider a phased approach, starting with smaller, less costly AI projects that can provide quick wins (*ibid.*). Another type of challenge is the issue of limited AI expertise, and this is often tied to the educational reason—the absence of an African-tailored curriculum for African builders, encompassing technical skills, knowledge, certification, soft skills, interdisciplinary skills and so on. In response, non-African countries look for partnerships with universities or consider upskilling current employees (*ibid.*). The third main challenge is perhaps the most difficult to address: high data costs. This is a bigger problem in the African context due to the lack of a robust and structured data ecosystem (Ade-Ibijola & Okonkwo, 2023). Responding to this issue, most countries leverage open-source data, negotiate partnerships for data sharing, or invest in generating their data (*ibid.*). Peculiar to the African context is the absence of a trust framework that can coordinate approaches to gathering data AI for projects in African society. We argue for heavy financial investments into the African social technology space, enabling a healthy procedure of gathering and handling locally sourced data within the African context for AI projects.

EXAMINING AVAILABLE AVENUES FOR FUNDING AFRICA'S TRUSTWORTHY AI PROJECTS: PROS AND CONS

In the African setting, the successful path to trustworthy AI is closely related to the availability and distribution of financial resources. Despite the growing interest in AI in Africa, a critical concern arises when examining the funding landscape for trustworthy AI development, particularly the glaring disparities between developed countries and Africa. We contend that designing trustworthy AI in Africa requires human agency and oversight, technical robustness and safety, privacy and data governance, transparency, etc., which come at a cost, however critical for driving innovation, fostering the growth of AI technologies, and building trustworthy AI institutions and ecosystems in Africa (Díaz-Rodríguez et al., 2023). However, the continent lags in this area, with limited funding available for AI-related projects.

Funding is the lifeblood of any AI initiative, and securing it requires careful planning and strategy (Campos-Zabala, 2023). As of today, no document can be referenced that contains Africa's strategy for sourcing funds, locally and internationally, in executing AI projects tailored for the

African audience. When embarking on an AI journey, the interconnection between funding, implementing AI initiatives, and measuring their impact is vital (ibid.). In this area today, there is a huge gap between African funders and the rest of the world. Global AI advancement projects in Europe, the United States, and China cannot be compared with Africa in terms of funding and how designers are responding to their people's experiences with AI solutions. Focusing on this as a fact amplifies the call for building trustworthy AI institutions to develop trustworthy social technologies that resemble the African worldview while acknowledging that the issue of funding deficit must be addressed to advance trustworthy AI research, which prioritises setting up trustworthy AI design parameters in the African context.

According to Gerrard et al. (2019) of the World Economic Forum (WEF), the estimated global economic value of AI is \$15.7 trillion in 2030. Even though AI is at its early stage, the financial benefits of AI are as follows: \$7.0 trillion for China, \$3.7 trillion for Northern America, \$1.8 trillion for Northern Europe, and \$1.2 trillion for Africa. Africa sits at the bottom of the economic gains from AI. Furthermore, in 2022, the US spent \$47.7 billion in AI investments, China spent \$13.4 billion, and Germany spent almost a billion euros; while in 2021, the European Union Council targeted an annual investment of €20 billion to AI, while Africa spent \$2.0 billion in 2021 and \$3.0 billion in 2022. Additionally, despite having low investment in AI, some of Africa's AI investments come from Western investors, such as the Google and IBM innovation labs and the Nigerian Kudi AI, which Silicon Valley funds. Given the low financial investment we have highlighted above, we aim to show how this funding gap obscures and shapes the conversation on design parameters and variables of a trustworthy AI with an African outlook.

Funding the Design Locally

This funding option is sometimes described as internal funding, and usually, it is used to describe the funding that comes from within. It may include reinvestment of profits, operating budgets, or internal fundraising efforts (Campos-Zabala, 2023). In most cases, it is preferred for some reasons, such as possessing control for using funds, greater flexibility, and reduced risk in terms of not depending on their parties (ibid.). Considering some worth noting pros, local funding AI design for African audiences creates local empowerment, surrendering control and both

resource and man management to the indigenous users of AI products. Another key advantage is the attainment of cultural relevance. This speaks to the fact that the people can do as they see fit to a project, in most cases, reflective of cultural norms, since the project activities are funded internally. One other advantage that is often mentioned is sustainable development. There is the argument that any monies used are expected to result in a commensurate return to where the monies are taken. Internal funding is released with the idea that it drives development in society. While internal funding has numerous benefits, there are downsides to it. For example, one of the downsides is the problem of limited resources. Most organisations or countries in Africa lack resources to execute their scale of projects, and in most cases, the lack of resources is often measured in terms of funding. Another way that internal funding is measured is to acknowledge the expertise gap. Inadequate internal funding means that the right expertise, which is most of the time expensive, cannot be hired. The risk of isolation is another disadvantage of relying on internal funding. Most AI project initiatives pool together internal and external funding from different sources to expand their access and reach. In cases where there is total reliance on internal refunding, expansion is difficult, leading to risk isolation and denying exposure to global perspectives and best practices.

Sourcing Funds Elsewhere

As the previous funding option, this funding option is sometimes described as external funding, encompassing funding sources which may include government grants and incentives, venture capital and private equity, corporate partnerships and collaborations, crowdfunding, and alternative financing methods (Campos-Zabala, 2023). In essence, funders send the message, “If you want the money, then build trustworthy AI!” (Gardner et al., 2022). As requirements, a compelling business case and a clear demonstration of potential return on investment must be included in the proposal. What is worth thinking about is whether African AI project owners approach potential funders with a business case. Nevertheless, there are some advantages to this funding option. For example, there is access to capital, sometimes referred to as venture capital or private equity (ibid.). Firms provide both funding and expertise. However, they require, in exchange, an equity stake in your business and usually, they ask for a term sheet—a document that outlines the key

terms of an investment agreement. It includes information such as the amount of investment, the valuation of the company, and the terms of repayment. Another advantage is that external funding allows for large-scale projects with the potential for greater impact and innovation. On the other hand, the disadvantages are numerous. There is a case of dependency on external funding, which may make the project susceptible to changes in global economic conditions, geopolitical factors, or the agenda of the funding organisations. Also worth considering is the problem of cultural misalignment, where projects funded from abroad might face challenges in terms of cultural practices, potentially leading to implementation issues. In addition, there is the issue of stringent conditions whereby external funders may impose conditions and expectations that could conflict with the project's original goals, potentially compromising its integrity. Additionally, there might exist stringent control from external funders on what kinds of technologies out to be produced. This problem is similar to what is conceptualised as funding bias in evidence-based medicine (Ugar, 2023c). This form of control may pertain to what kinds of design, policies, and frameworks ought to govern the AI ecosystem in Africa, a similar problem faced in the development of technologies like vaccines and medications in evidence-based medicine (Ugar, 2023c).

Collaboration and Partnerships

Another viable funding source involves collaboration and partnership. It is a usual practice in this type of funding that partners or collaborators provide financial support and potentially access to resources, market opportunities, and expertise (Campos-Zabala, 2023). Identifying companies with an interest in AI and negotiating mutually beneficial terms are critical steps in this process (ibid.). Considering some advantages of this funding option, one is the synergy of resources. Collaboration brings together diverse skills, resources, and perspectives, enhancing the overall strength of the project. Another advantage is risk sharing; it involves partners sharing financial and operational risks, making the project more resilient to challenges. Worth considering is the advantage of global–local balance in the sense that partnership allows for a balance between local knowledge and global expertise, ensuring a more comprehensive and effective project. There are several disadvantages, and one is coordination challenges. It is a matter of fact that managing collaboration can be

complex, especially when dealing with diverse partners with varying priorities and working styles. Another disadvantage is that decision-making may be slower due to the need for consensus among partners, potentially impacting the project timeline. Lastly, the challenge of intellectual property concerns, in which issues related to the ownership and control of intellectual property can arise in collaborative projects, requiring careful negotiation.

RECOMMENDATIONS

As hypothesised, this chapter has demonstrated that trustworthy AI is not based on trusting the technologies but on the institution where the technology emanates. We further argued that Africa can only achieve a trustworthy AI ecosystem when they become agents of their designs; that is, designing their AI technologies within their institutions and crafting policies from their socio-ethical and cultural practices to shape the designs and use of AI. Nonetheless, we are not oblivious to some challenges that may arise in achieving a trustworthy design of AI in Africa. The major problem which we addressed is the issue of funding deficit, which is the bane of trustworthy AI projects tailored for the African audience. Based on the discussions and findings, we provide some novel and actionable recommendations.

- a. Hybrid Funding Approach: We recognise that project interests may be diverse, and it is crucial to be flexible to take advantage of available funding options. Therefore, we propose a funding policy that keeps all three funding options open for easy access.
- b. Innovative Programmes: Launch a Tech Collaborators Fellowship programme that pairs African AI enthusiasts with global mentors, in terms of connecting local innovators with international funding opportunities and creating an avenue for local designers to showcase their commitment to upskill for knowledge transfer, to both help African designer develop local expertise and foster cross-cultural collaboration, injecting fresh perspectives into the tech experience of their people.

- c. Financial Security for Innovative Safeguarding: This involves creating a smart funding charter that transparently records and verifies project goals to help external funds adhere to agreed-upon terms, safeguarding against mission drift. This will encompass project alignment with local needs and a financial safety net, providing stability even if a funding source experiences changes.
- d. Collaborative Governance Framework: This goal is to develop an AI platform that can facilitate real-time decision-making in collaborations that is adaptable to the evolving needs of the project to minimise delays. This governance will encompass defining roles and responsibilities in a visually engaging format—a co-creation compact movement and providing constant communication support for collaborators, thereby mitigating communication gaps, and maintaining engagement throughout the project.
- e. Forming Local AI Pressure Group: The problem of funding deficit in Africa's conversation of trustworthy AI projects has a political undertone. There must be an advocacy group that can pressure the government into channelling resources into more fruitful endeavours in support of widening and developing the tech ecosystem in Africa.

CONCLUSION

This chapter highlighted the importance of designing trustworthy artificial intelligence (AI) systems grounded in the African context. It argues that most current AI designs deployed in Africa are not shaped by policies informed by the cultural norms, ethos, worldviews, and ethics emerging from Africa. The chapter identified financial constraints as one of the impediments to the design of AI in Africa to achieve trustworthiness. It recommended a hybrid funding approach, innovative programmes, financial security for innovative safeguarding, a collaborative governance framework, and forming a local AI pressure group to address the funding deficit. The chapter also emphasised the need for African governments to provide an enabling environment for AI development and innovation to thrive. Overall, the chapter argued that Africans ought to be wary of designs emerging from elsewhere, given African histories of colonialism, neo-colonialism, and techno-colonialism. The chapter recommended that Africans build institutions that care for the well-being of Africans to guide the design of trustworthy AI.

REFERENCES

- Abebe, R., Aruleba, K., Birhane, A., Kingsley, S., Obaido, G., Remy, S. L., Sadagopan, S. (2021, March). *Narratives and counternarratives on data sharing in Africa*. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 329–341).
- Ade-Ibijola, A., & Okonkwo, C. (2023). Artificial intelligence in Africa: Emerging challenges. In *Responsible AI in Africa: Challenges and opportunities* (pp. 101–117). Springer.
- Agarwal, S., & Mishra, S. (2021). *Responsible AI*. Springer.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed: 29 September 2023
- Azeez, A., & Adeate, T. (2020). Second-wave AI and Afro-existential norms. *Filosofia Theoretica: Journal of African Philosophy, Culture and Religions*, 9(3), 49–64.
- Bachmann, R. (2020). Trust and institutions. In D. C. Poff & A. C. Michalos (Eds.), *Encyclopedia of business and professional ethics* (pp. 1–6). Springer. <https://doi.org/10.1007/978-3-319-23514-1>
- Baier, A. (1986). Trust and antitrust. *Ethics*, 96(2), 231–260.
- Baier, A. (2013). What is trust? In D. Archard et al. (Eds.), *Reading Onora O'Neill*. Routledge.
- Bauer, P. C. (2019). Conceptualizing trust and trustworthiness. *Journal of Trust Research*, 9(1), 1–17. <https://doi.org/10.1080/17513072.2019.1573460>
- Bekker, S. (2021). *Fundamental rights in digital welfare states: The case of SyRI in the Netherlands*. Netherlands Year Book of International Law 2019: Yearbooks in International Law: History, Function, and Future.
- Benk, M., Tolmeijer, S., von Wangenheim, F., & Ferrario, A. (2022). The value of measuring trust in AI-a socio-technical system perspective. arXiv preprint [arXiv:2204.13480](https://arxiv.org/abs/2204.13480)
- Benjamins, R., Barbado, A., & Sierra, D. (2019). *Responsible AI by design in practice*. arXiv preprint [arXiv:1909.12838](https://arxiv.org/abs/1909.12838)
- Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed*, 17, 389.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.
- Campos-Zabala, F. J. (2023). Scaling AI. *Grow your business with AI: A first principles approach for scaling artificial intelligence in the enterprise* (pp. 479–507). Apress.
- Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, 71, 1137–1181.

- Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, 35(4), 410–422.
- Coeckelbergh, M. (2012). Can we trust robots? *Ethics and Information Technology*, 14, 53–60.
- Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99, 101896.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way* (Vol. 1). Springer.
- Du'ran, J. M., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines*, 28, 645–666.
- Eke, D. O., Chintu, S. S., & Wakunuma, K. (2023). Towards shaping the future of responsible AI in Africa. In *Responsible AI in Africa. Social and cultural studies of robots and AI* (pp. 169–193). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-08215-3_8
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379.
- Foley, R. (2001). *Intellectual trust in oneself and others*. Cambridge University Press.
- Forrest, K. (2021). *When machines can be judge, jury, and executioner: Justice in the age of artificial intelligence*. World Scientific Publishing Co., Pte. Ltd.
- Fuglsang, L., & Jagd, S. (2015). Making sense of institutional trust in organisation: Bridging institutional context and trust. *Organisation*, 23(1), 23–39.
- Gambetta, D. (1988). *Trust: Making and breaking cooperative relations*. Blackwell.
- Gardner, A., Smith, A. L., Steventon, A., Coughlan, E., & Oldfield, M. (2022). Ethical funding for trustworthy AI: Proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice. *AI and Ethics*, 2, 277–291.
- Gerrard, J., Webster, K., McNaughton, S., & Kukutai, T. (2019). *By 2030, AI will contribute \$15 trillion to the global economy*. World Economic Forum. <https://www.weforum.org/agenda/2019/08/by-2030-ai-will-contribute-15-trillion-to-the-global-economy/>. Accessed: 30 November 2023.
- Greene, C. (2023). AI and the social science: Why all variables are not created equal. *Res Publica*, 29, 303–319. <https://doi.org/10.1007/s11158-022-09544-5>
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37).
- Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.
- Hardin, R. (2006). *Trust*. Polity Press.

- Hardwig, J. (1991). The role of trust in knowledge. *The Journal of Philosophy*, 88(12), 693–708.
- Hawley, K. (2012). *Trust: A very short introduction*. Oxford University Press.
- Hawley, K. (2014). *How to be trustworthy*. Oxford University Press.
- HLEG. (2019). *Ethics guidelines for Trustworthy AI*. European Commission. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>
- Holton, R. (1994). Deciding to trust, coming to believe. *Australasian Journal of Philosophy*, 72(1), 63–76.
- Jones, K. (1996). Trust as an affective attitude. *Ethics*, 107(1), 4–25.
- Jones, K. (2004). Trust and terror. In P. DesAutels & M.U. Walker (Eds.), *Moral Psychology: Feminist Ethics and Social theory* (pp. 3–18). Rowman & Littlefield.
- Kennard, J. G. (2022). ‘Trusting-to’ and ‘trusting-as’: A qualitative account of trustworthiness. *Philosophical Studies*, 179(8), 2509–2539.
- Lagerspetz, O. (2015). *Trust, ethics and human reason*. Bloomsbury Academic.
- Lahno, B. (2001). Institutional trust: A less demanding form of trust? *Revista Latinoamericana de Estudios Avanzados*, 15, 19–58.
- Lehrer, K. (1999). Self-trust: A study of reason, knowledge and autonomy. *Philosophical and Phenomenological Research*, 59(4), 1045–1055.
- Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. Alan Turing Institute. <http://arxiv.org/abs/1906.05684>
- McCullom, R. (2017). *Facial recognition technology is both biased and understudied* [Online]. <https://undark.org/2017/05/17/facial-recognition-technology-biased-understudied/?msclkid=b913148dd10911ec932962fbfdacl591>
- Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and ‘the dark side’ of AI. *European Journal of Information Systems*, 31(3), 257–268.
- Misztal, B. A. (1996). *Trust in modern societies: The search for the bases of social order*. Blackwell.
- Mollering, G. (2006). *Trust: Reasons, routine, reflexivity*. Emerald Group Publishing.
- Nickel, P. J., Franssen, M., & Kroes, P. (2010). Can we make sense of the notion of trustworthy technology? *Knowledge, Technology & Policy*, 23(3), 429–444. <https://doi.org/10.1007/s12130-010-9124-6>
- OECD. (2017). *OECD guidelines on measuring trust*. OECD Publishing. <https://doi.org/10.1787/9789264278219-en>
- Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, 1(1), 34–47.
- Pettit, P. (1995). The cunning of trust. *Philosophy & Public Affairs*, 24(3), 202–225.

- Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering*, 26, 2749–2767.
- Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). *Principles to practices for responsible AI: Closing the gap*. arXiv preprint [arXiv:2006.04707](https://arxiv.org/abs/2006.04707)
- Segun, S. T. (2021). Critically engaging the ethics of AI for a global audience. *Ethics and Information Technology*, 23(2), 99–105.
- Slosser, J. L., Aasa, B., & Olsen, H. P. (2023). Trustworthy AI: A cooperative approach. *Technology and Regulations*, 58–68. <https://doi.org/10.26116/techreg.2023.006>
- Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.
- Ugar, E. T. (2023a). Rethinking remote work, automated technologies, meaningful work and the future of work: Making a case for relationality. *Philosophy and Technology*, 36(32). <https://doi.org/10.1007/s13347-023-00634-7>
- Ugar, E. T. (2023b). 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.
- Ugar, E. T. (2023c). Evidence-based medicine and patient-centred care: A patient's best interest analysis. *Revista Opiniao Filosofica*, 14, 1–18. <https://doi.org/10.36592/opiniaofilosofica.v14.1072>
- United Nations. (2021). *Recommendation on the ethics of artificial intelligence*.
- van De Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines*, 30, 385–409.
- Wakunuma, K., Ogoh, G., Eke, D. O., & Akintoye, S. (2022, May). *Responsible AI, SDGS, and AI governance in Africa*. In 2022 IST-Africa Conference (IST-Africa) (pp. 1–13). IEEE. <https://doi.org/10.23919/IST-Africa56635.2022.9845598>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

