DATA FOR POLICY PROCEEDINGS PAPER

Social justice considerations in developing and deploying AI in Africa

Getachew Hailemariam Mengesha[1], Elefelious Getachew Belay[2] (ORCID) and Rachel Adams[3] (ORCID)

[1]School of Information Science, Addis Ababa University, Addis Ababa, Ethiopia
[2]School of Information Technology and Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia
[3]Research ICT Africa, Cape Town, South Africa
Corresponding author: Elefelious Getachew Belay; Email: elefelious.getachew@aau.edu.et

Abstract

In the literature, there are polarized views regarding the capabilities of technology to embed societal values. One aisle of the debate contends that technical artifacts are value-neutral since values are not peculiar to inanimate objects. Scholars on the other side of the aisle argue that technologies tend to be value-laden. With the call to embed ethical values in technology, this article explores how AI and other adjacent technologies are designed and developed to foster social justice. Drawing insights from prior studies, this paper identifies seven African moral values considered central to actualizing social justice; of these, two stand out—respect for diversity and ethnic neutrality. By introducing use case analysis along with the Discovery, Translation, and Verification (DTV) framework and validating via Focus Group Discussion, this study revealed novel findings: first, ethical value analysis is best carried out alongside software system analysis. Second, to embed ethics in technology, interdisciplinary expertise is required. Third, the DTV approach combined with the software engineering methodology provides a promising way to embed moral values in technology. Against this backdrop, the two highlighted ethical values—respect for diversity and ethnic neutrality—help ground the pursuit of social justice.

1. Introduction

This study intends to draw concepts from social justice and moral theorization and explores how modern technologies such as artificial intelligence (AI) need to be designed and implemented to foster social justice by reducing scenarios of discrimination, prejudice, and bias and eventually upholding basic human rights. Emerging AI technologies can interact with the environment and make routine decisions. There has been a growing concern regarding the ethical soundness of decisions being made by these autonomous entities. Furthermore, it is unclear who should be accountable for the consequences of decisions made by artificial agents: the designer, the current owner, the end-user, and so forth. To address this paucity in the existent literature, the study attempts to draw explanatory insights from social justice and moral theorization that would inform how socially good and ethically appropriate AI-based technologies will be designed and developed.

In his seminal work, Rawls (1971) theorized justice as fairness and proffers principles to foster social justice. Similarly, Justice has been conceptualized as the foremost virtue of social institutions (Jost and

Kay, 2010). With this underpinning notion, Jost and Kay (2010) described social justice as a state of affairs where (a) benefits and burdens in society are dispersed in line with some allocation principles; (b) procedures, norms, and rules that govern political and other forms of decision-making preserve the basic rights, liberties, and entitlements of individuals and groups; and (c) human beings should be treated with dignity and respect, not only by authorities but also by other relevant social actors, including fellow citizens. Jost and Kay further stated that the definition they have put forth aligns well with distributive, procedural, and interactional justice. These scholars also argued that social justice is a manifestation of a fair social system, in contrast to an unfair system marked by prevalent exploitation, tyranny, oppression, prejudice, discrimination, and similar issues. The term "justice" is often used to characterize fairness, deserving, and entitlement. As many people might think, justice is not only about self-interest and one's gain but also fair treatment and one's value within the group. With this in mind, various studies have been conducted to understand the racial and gendered implications of AI, which constitutes a major ethical concern (Benjamin, 2019). AI systems and tools tend to impact social groups through algorithmic bias and data imbalance (Adams, 2021; Buolamwini and Gebru, 2018). To minimize the adverse social impacts of AI-based technologies, various regulatory guidelines have been developed (Poel, 2020). However, these guidelines are predominately rooted in Western culture and embrace principles such as autonomy, nonmaleficence, fairness, transparency, explainability, and accountability. Many of these ethical principles may not fit the African context as African countries have unique cultures and moral values. Generic ethical guidelines developed considering the Western way of life and social systems would not fit the African context. Therefore, there is a need to conduct a thorough study of African culture to identify relevant moral values that would guide the design, development, and adoption of AI-based technologies. Moreover, Munn (2023) contended that a range of ethical principles formulated thus far tend to be opaque, ambiguous, and incoherent. As a result, could not guide the process of designing and developing ethically appropriate AI-based technologies.

Although some attempts are underway to translate AI ethical principles into practices, nailing them down to algorithmic rulesets and embedding them in AI systems remains a daunting venture. This study is motivated to explore how AI systems will be designed and developed embedding ethical values. Data for the study were collected using Focus Group Discussion (FGD), Key Informant Interview (KII), and document analysis. A series of FGD and KII sessions were conducted with directors and employees of the Artificial Intelligence Institute (AII) and AI system developers in Addis Ababa, Ethiopia to elicit relevant facts regarding AI-based systems adoption, development, implementation, and use. The data collected is used to substantiate and support the process of identification, analysis, and implementation of moral values in AI systems. The empirical findings obtained in Ethiopia were substantiated by similar studies conducted in South Africa, Uganda, Nigeria, and Rwanda sponsored by Research ICT Africa (RIA) and by the Human Sciences Research Council (HSRC).

A considerable debate is underway surrounding the capability of technological artifacts to carry values. Scholars such as Pitt (2013) argue that technological artifacts are value-neutral as values are peculiar to people that inanimate objects do not possess or embody. Conversely, several scholars such as Poel and Kroes (2013) posit that technologies are value-laden. These scholars further suggest the process through which values would be identified, translated, validated, and embedded in AI systems. Nevertheless, standard methodologies and best practices of value-oriented design are scanty.

The other issue regarding embedding value in AI emanates from varying philosophical and psychological perspectives of moral values, which makes conceptualization difficult for technical system developers. This compels system builders to either attempt to blend varied perspectives or declare from the outset the philosophical foundation they would adopt in the course of investigating the phenomena related to moral values. To this end, scholars have been conducting an extensive review of philosophical and social psychology literature to better understand the contextual meanings of moral values. The Merriam-Webster dictionary defines moral values as "Principles of right and wrong in behavior." Similarly, Moore and Asay (2013) considered value as guiding principles of thought and behavior. This conception assumes that behavioral practices and habits would be validated by society. When a society embraces a set of values, members of the society are expected to act and judge in line with these values.

This notion of moral value tends to be akin to the philosophical perspectives of value that lay high regard to uncovering relevant moral values, moral reasoning, and judgment.

Social Psychologists such as Kohlberg (1984) emphasized how moral values evolve and shape human behavior through a long process of observation and social interaction. In his endeavor to provide explanatory analysis, Kohlberg proposes a stage model where humans develop a set of morals as they mature socially and intellectually. He also contends that a person's sense of justice and judgment-making ability about what is considered good or bad evolve due to changes in cognitive abilities. By and large, the psychological perspectives of morals have established a logical connection between a person's beliefs, values, attitudes, and behavior.

Scholars who believe that technological artifacts including AI systems can embed value have identified different types of values associated with the artifact. For instance, van de Poel and Kroes (2013) have identified two types of values: intrinsic and final value or value for its own sake and also called instrumental value. The intrinsic refers to values that are associated with the object's intrinsic properties. Van de Poel and Kroes (2013) conclude that designers need to anticipate the actual use and the actual realization of values in the course of designing and developing an artifact. Recognizing the vitality of morale value considerations in AI-based systems design and development Manders-Huits (2011), Rahman (2018), the current study primarily intends to address the following questions. What prominent moral values shape Africa, particularly sub-Saharan African people's thoughts and behavior? 2. What methodological approaches would help to translate ethical moral values into AI system design features?

The remaining part of the article is structured as follows. First, a brief account of relevant Ethiopian and African values will be provided along with narratives on how these values shape individuals, thoughts, actions, and the overall social reality. Second, a review of methods and approaches for embedding values in AI systems will be provided. Third, critical reflection will be given on the methodological approaches discussed in the paper. Fourth, the paper ends with concluding remarks.

2. Account of African/Ethiopian moral values

This section aims to explore salient moral values that shape people's thoughts and actions in sub-Saharan African countries. Africa has a unique culture and philosophy. Ethical principles that emanate from African culture are supposed to be helpful for moral formation. Although a single African worldview is none existent, still African conceptions and ideas are shared by many sub-Saharan Africans (Behrens, 2013). Among other things, high regard given to a community (the oneness of the community) appears the most salient African moral perspective (Behrens, 2013). This is quite the opposite of Western culture, which emphasizes individualism. The author cites the most popular African adage "A person is a person through other persons" or "I am because we are" as expressing the essence of an African philosophy, which stresses being in a relationship with others. In this context, personhood implies being moralized or a genuine human being such that to be a person in the true sense is to exhibit good character (Metz, 2010). Emphasizing the high regard given to social relations, Metz (2010) stated that in African culture harmonious or communal relationships are valued for their own sake, not as a means to some other values such as pleasure.

The high regard given to relations and communitarianism is now being theorized with the conception of "ubuntu." Metz (2010) described the principles of Ubuntu as follows:

> *An action is right just insofar as it is a way of living harmoniously or prizing communal relationships, ones in which people identify with each other and exhibit solidarity with one another; otherwise, an action is wrong.*

The emphasis on community, identifying with others, solidarity, and caring makes ubuntu a relational ethics that prizes harmonious relationships (Behrens, 2013).

African people have many commonalities and diverse cultures. The diversity in Africa, particularly, is based on culture, ethnicity, gender, and religion. These days diversity is considered as an instrument for an

appreciation of individual differences. It is now becoming imperative to regard diversity as a means to ensure inclusion in society's socioeconomic system (Appiah et al., 2018). The authors further state that it is necessary to properly manage diversity to mitigate deleterious ethnic tension and conflicts. The polarized and poorly managed, ethnic-based political contest and violence have led to untold human suffering and genocide in Rwanda. Belgian colonizers divided Rwandans by requiring all Rwandans to carry identity cards that classified people by their ethnicity and this made it easy for perpetrators to identify and kill people brutally (IWGIA, 2023). Recognition of diversity is believed supportive of instituting distributive justice through the allocation of social goods across a range of autonomous spheres (Flanagan et al., 2019). During the FGD session, a respondent stated that "City administrations 'in Ethiopia' conducted household survey of dwellers including their ethnic information." When violence occurred in the cities in October 2018 and June 2020, respectively, dwellers were profiled and attacked because of their ethnicity. Strong legislation and enforcing mechanisms need to be in place that outlaw the gathering of ethnic-related data routinely to provide government services. The future AI and machine learning systems should be constrained from conducting classification or prediction based on ethnicity politics-oriented datasets.

Ethiopia is one of the African ancient countries with a diverse culture. As the majority of Ethiopian people is religious, Ethiopian ethics is mainly drawn from the holy books such as the Bible and Quran (Tessema, 2021). The Ethiopian culture played a critical role in terms of building strong relationships among citizens. The indigenous political structures of the past such as the "Gada" system of the Oromo people are the source of moral life in some parts of Ethiopian society. Among other things, the Ethiopian culture lays high emphasis on peaceful coexistence and humanity. In line with this, Tessema (2021) identified the following ethical virtues:

- Peaceful coexistence
- Humanity
- Brotherhood
- Tolerance
- Patriotism and sacrificial love for the country
- Harmony
- Cooperation
- Communality
- Mutual respect
- Respect for diversity

Before the communist dictatorial regime came to power in 1974, Ethiopian children used to learn virtues and right conduct from the holy books, families, neighbors, and school settings. The modern education system adopted in Ethiopia is Western-oriented, which departs from local virtues and needs. Efforts are now underway to revitalize indigenous Ethiopian knowledge and values through curriculum development and Ethics education. The indigenous Ethiopian ethics anchors on peaceful coexistence that places high emphasis on mutual respect and brotherhood. The indigenous ethics that endured for generations are being eroded due to internal and external challenges such as globalization, political conspiracy in the country, and money politics.

In general, the Ethiopian traditional ethical virtues identified by Tessema (2021) are fully in line with the contextualized African set of ethical principles identified by Gaffley et al., (2022). Table 1 presents a list of relevant values identified by these scholars and from the upcoming study we conducted on context-specific African moral values.

The authors contend that these ethical principles fit well with the African sociocultural context. Therefore, grounding the design and development of AI technologies accordingly increases the uptake of the technology (Gaffley et al., 2022). Over the past 30 years, Ethiopia has been embracing ethnic-based federalism. This mode of administration has brought some advantages and unintended consequences as

Table 1. *Relevant African moral values*

#	African moral values	Description
1	Respect for persons	Refers to the right threat people in dignity and respect the right to make decision
2	Beneficence	The principle of creating AI technology that is beneficial to humanity
3	Maleficence	Refers to "do not harm" which implies preventing harms arising from the intent of humans or the unpredicted behavior of machines
4	Harmony	Refers to preserving solidarity and communal relationship
5	Explicability	The need to understand and take into account the decision-making processes of AI
6	Ethnic neutrality	Deal with outlawing routinely collecting and classifying ethnicity-related data
7	Respect for diversity	Refers to ensuring equitable representation of gender, regions, religions, language, etc… in a social system

well. In connection to the high regard given to ethnicity and collective identity, respect for diversity has emerged as an important value to smoothly maintain the Ethiopian enduring value of "living together."

3. Methods and approaches for embedding values

Software systems tend to exhibit social values (Flanagan et al., 2019). In line with this, there has been a growing call to regulate AI embedding ethical values in the systems. However, operationalizing moral values and identifying concrete design factors is no small feat. Various ethical frameworks that would serve as overarching guidelines for designing and developing AI systems are available. Further, scholars have developed detailed systematic approaches, procedures, and methods to embed ethics in AI systems. The consensus noted is that although ethical issues appear daunting to identify and encode, endeavors are underway to develop generic guidelines that chart how socially good AI systems may be designed and developed. In addition, concerted efforts of scholars from around the world are underway to promote the notion of design thinking and to propose methods and techniques that would be used in the course of embedding ethics in AI systems. The presence of ethical guidelines, ethical design thinking, and design considerations undoubtedly foster AI innovations.

Flanagan et al. (2019) stated that the design and development of technologies have been laying high emphasis on instrumental values such as functionality, safety, reliability, and ease of use. Designers are now expected to embrace social and political values as important design considerations to determine the quality and usability of the product. Moreover, Flanagan et al. (2019) noted that designers need to deal with three dimensions. First, they are required to engage in scientific and technical results. Second, they need to absorb salient philosophical reflections on values. Third, they need to correlate the result of the empirical investigation of values to individuals and their society. Finally, the authors proposed DTV (discovery, translation, and verification) as a methodological approach to capture, translate, validate, and embed values in a technological artifact. While suggesting this approach, they regarded values as criteria of excellence in technical design such as functional efficiency, reliability, robustness, elegance, and safety. Therefore, DTV supplements do not replace the well-established design methodologies. Instead, the reciprocal consideration of functional factors that are pretty well identified with the conventional methods of problem analysis, together with values supposed to be identified with the DTV approach is expected to result in designs that are good from the material as well as moral perspectives. In their experimental project, Flanagan et al. (2019) had identified four types of values: value in the definition of the project, instrumental values, designers' values, and users' values.

The DTV framework Flanagan and his associates suggested relates to previous initiatives. Table 2 presents the basic commonalities and distinctions of each method.

Table 2. *Methods and approaches for embedding values*

Methods	Philosophy/ perspectives	Level of stakeholder's participation	Problem analysis techniques	Artifact design process
Participatory design	Co-design (co-determine the development of systems)	High (democratic participation)	Interactive	Iterative and interactive
Value sensitive design (VSD)	Conceptual, analytical, normative	No clear methodology for identifying stakeholders	Unclear	Unclear
Scenario-based design	Solution-first	High	Candidate solutions analysis means clarifying the problem state	Interactive process
Prototyping	Exploratory	High	Interactive	Iterative
Agile development approach	Adaptive	High	Interactive	Incremental
DTV (discovery, translation, verification)	Exploratory	High	Interactive	Iterative

The DTV activities described in Table 2 can be linked with well-established software development approaches such as participatory design, scenario-based design, prototyping, and agile development methodology. The following section attempts to demonstrate how the DTV set of activities is blended with and complements the conventional software development methodologies to capture, analyze, drive, and implement values in AI systems.

3.1. Value discovery

Discovery involves identifying relevant values to a particular design project (Flanagan et al., 2019). Through this process, a set of values is expected to emerge as a final deliverable. Similar to the conventional system analysis and development methodology in software engineering, discovery involves the use of fact-finding tools and techniques, like interviews, focus group discussions, and document review to elicit diverse values that relate to individuals, institutions, culture, religion, etc.

During the value discovery phase, the participatory design, prototyping, and agile approaches can be used to systematically identify and refine values relevant to the context of the project. For instance, PD may be used as an overarching approach in the process of identifying values. This is because PD advocates the direct involvement of end users as co-designer of a technology they will use (Robertson and Simonsen, 2012). The central tenet focuses on fostering democratic participation and on how to infuse a collaborative design process involving people who would be affected by the technology to be developed. A range of tools and techniques including the classical data collection methods, such as face-to-face meetings, prototypes, mock-ups, facilitated workshops, etc, would be used to complement the participatory design approach.

In software system development, a prototype refers to a working model, simulation, or partial implementation of a system meant to determine user requirements (Carr and Verner, 1997). This

requirement identification capability is considered a salient tool to identify values at individual and societal levels. As prototypes are usually a model or scaled-down version of a real-world phenomenon, people are more likely to comprehend easily the prototype by playing around with it. Furthermore, end users quite often pinpoint limitations in the prototype and also suggest improvements. In this regard, Flanagan et al. (2019) stated that a team in their experimental project called RAPUNSEL succeeded in discovering users' beliefs, preferences, and values using prototyping. This is a typical exploratory approach for identifying what users would like to see from the system under consideration and also suggests how best users' needs will be served. Therefore, through a series of prototyping cycles, hidden requirements, and unexpressed values are expected to be elicited. In addition to prototyping, agile and scenario-based development approaches can be used for discovering values.

3.2. Translation

Translation deals with the process of embedding values in the design of AI-related artifacts. Flanagan et al. (2019) state that translation is further subdivided into two parts: operationalization and implementation respectively. Operationalization involves articulating values in concrete terms whereas implementation deals with transforming values and concepts into corresponding design specifications and lines of code. Designers are expected to work side by side in the value implementation process along with functional requirements and other constraints (Flanagan et al., 2019). In a RAPUNSEL experiment, project values such as cooperation through code sharing and reward systems are translated and implemented using system development tools such as Smart Code Editor II (Flanagan et al., 2019).

Once values are nailed down to variable and parameter levels, a range of conventional software systems development tools and techniques including prototyping and the agile approach would be used to refine the design and to write program codes.

3.3. Verification

Verification refers to the process of assessing the extent to which target values in a given system have been implemented successfully (Flanagan et al., 2019). Unlike the conventional software development system where verification stresses factional efficiency and usability, the inclusion of values focuses on methods such as internal testing carried out by the design team, formal and informal interviews, the use of prototypes, etc.

Flanagan et al. (2019) stated that in a RAPUNSEL project, the agile development method was used to account for the dynamism of the system development with changing requirements. Moreover, the verification process was facilitated through regular meetings with design partners, educators, and industry advisors. When additional values and ideas emerge, designers quickly translate them into factional, technical, and values-oriented parameters.

In a bid to demonstrate how the identified values are translated and verified, we have selected two values listed in Table 1 "Ethnic neutrality" and "Respect for diversity" respectively. In this context, "Ethnic neutrality" is a subset of diversity. Discrimination and gross human rights abuse observed in Rwanda, Ethiopia, and Somalia and all over sub-Saharan Africa compel outlawing routinely collecting and classifying ethnicity-related data. Benjamin (2019) expresses vividly the risk of automating racism rooted in biased datasets and algorithms. The Ethiopian dominant cultural and moral value identified by Tessema (2021) "Abiro Menore" meaning living together in harmony, is a widely shared value throughout the Sub-Sharan Africa and is an integral part of "Respect for diversity." Similarly, "Respect for diversity" stresses the fact that acknowledging individual and societal differences and special needs is at the center of fostering social justice in Africa. Therefore, it is imperative to ensure equitable representation of gender, regions, religions, and languages in a social system. Table 3 and Table 4 show how these values are operationalized, implemented, and verified using the conventional software engineering use case documentation methods.

Table 3. *Ethnic neutrality*

Value requirement	Ethnic neutrality
Important	Mandatory
Description	The AI system shall flash error message when it encounters a matching between a person's ID, Name, and ethnicity. The application shall disable data elements labeled "Ethn" or "Ethnicity" (and other ethnic-related labels) and shall enforce referential integrity between the cited three data elements in case they are stored in different database tables.
User story/scenario	As a user, I do not want the system to capture or compel me or any users to input ethnicity-related data. If there is a need to place this attribute in the database table, the entry shall be optional to protect my or other users' interests who do not want to disclose their ethnic identity.
Source	Drawn from the value analysis study
Acceptance criteria	The system shall have a basic data entry form or screen that will let users capture biographic data making "Ethnicity" field optional. The system shall disable this data element unless authorization is granted from the system user and/or delegated agency to enable it.

Table 4. *Respect for diversity*

Value requirement	Respect for diversity
Important	Mandatory
Description	The AI system shall capture regional, religious, gender, educational level, language, etc, data elements to maintain balanced data for machine learning.
User story/scenario	As a user, I want to interact with the system using my native language and also, and I would be happy if the system identifies me by my region, nationality, and gender.
Source	Drawn from the value analysis study
Acceptance criteria	The AI system shall capture or analyze regional, gender, language, religious, and educational service representations in order to foster citizens' public good or service access.

The use case description presented in Table 3 can be further refined and decomposed to identify normalized data elements and program logic. To this end, sequence and activity diagrams will be created to clearly depict the inner constitutes of the use case and to facilitate implementation.

4. Discussion

The AI ethics domain has witnessed polarized views and positions regarding the capability of technological artifacts to carry values. Scholars such as Pitt (2013) argued that technological artifacts are value-neutral as values are peculiar to people, which inanimate objects do not possess or embody. Conversely, several scholars such as Poel and Kroes (2013) posited that technologies are value-laden. These scholars further suggest the process through which values would be identified, translated, validated, and embedded in AI systems. This study is founded on the second position and attempts to demonstrate methodological approaches whereby moral values may be embedded in AI systems. The study contends that the process of value identification, translation, and verification needs to be carried out hand in hand with the

conventional process of functional and none functional software requirements, systems design, system testing, deployment, and commissioning.

Over the past ten years, dozens of AI ethical frameworks have been developed by international organizations, government agencies, professional organizations, industries, and educational institutions. These ethical frameworks dwell mainly on higher-level principles and legislative issues. However, they are far from regulating, the design, development, adoption, and use of these technologies (Floridi et al., 2018). Moreover, drilling down these ethical principles to a detailed technical design parameter level appears daunting as moral values do not easily lend themselves to atomic-level decomposition and quantification. This limitation led some authors like Munn (2023) to conclude that the plethora of AI ethics and ethical frameworks developed thus far are useless. Munn raises three issues to justify the position he has held:

- Most ethical principles are incoherent;
- They are isolated and situated in industries and educational settings, which largely ignore ethics;
- Translating abstract principles into technical rulesets is a daunting venture;
- AI ethics initiatives are not only fruitless but also divert resources and focus from more effective activities;
- AI ethical frameworks tend to set normative ideals but lack mechanisms to enforce compliance with values yielding a gap between principles and practice;
- Lack of friction between ethical principles and existing business principles etc.

Although Munn (2023) is fiercely critical of the unfairly high regard given to the formulation of abstract ethical principles, he acknowledges the potential power and danger associated with AI technologies. Munn eventually suggested dealing broadly with AI justice issues as it embraces ethics-related values. He further argued that, in essence, the moral properties of algorithms are not internal to the models themselves but rather a product of the social system within which they are deployed.

Building on the notion transpired by Munn (2023) this study has demonstrated how distributive justice can be enforced and nailed down values to the stage of design considerations by employing a use case analysis method widely used in the software engineering domain. Particularly, in the African context "Respect for diversity" and "Ethnic neutrality" are instrumental to institute justice and reduce discrimination. These two ethical issues can be regarded as an overarching principle to foster justice in African society. African people have gone through long years of subjugation, discrimination, marginalization, and lack of justice. The legacy of colonization coupled with a lack of good governance is visible in the prevailing uneven distribution of resources and economic inequalities. Some African countries are still battling with the historically rooted inequalities. Therefore, social justice is a long-held social problem in Africa. Respect for diversity and ethnic neutrality feature well African way of life and African fatal problems respectively. Efforts to design disruptive technologies like AI need to lay high emphasis on ensuring social justice, which is thought to be an overarching and pressing ethical challenge in the African context. To validate ethical principles and moral values identified in this study, FGD and KII were conducted with officers at the Ethiopian AII and private software companies. Respondents of the study provided in-depth explanations regarding efforts underway to develop ethical guidelines for emerging AI technologies. Concerning the development of AI ethical guidelines, one of the respondents stated the following:

> "The Artificial Intelligence Institute is partnering with a Western company that has ample experience and expertise in developing AI policy. This company is supposed to consult the development of Ethiopian AI Policy, with inputs provided by Ethiopian experts."

It appears the government somehow outsourced part of the AI policy development to a third-party (International Organization). This casts doubt that contextual ethical issues like "respect for diversity" and "ethnic neutrality" would remain unidentified and would not be included in the policy document. Rather,

it would have been good if the institute had organized an interdisciplinary local expert to develop the policy document with the facilitation role of a third-party entity.

Technical experts working in the institute and private AI companies reflected on the ethical stance of young experts engaged in AI system development. One of the respondents has made the following remark:

> "The young experts who engage in the development of AI and related systems are less social as they spend most of their time in the laboratory. Ethical issues are incompressible to them for they were not acquainted with them when they were at school. Therefore, there is a need to revisit the curriculum of computing disciplines and incorporate ethical issues."

The respondent further reflected on the critical role and mandate of government as follows:

> "The government is supposed to be proactive and farsighted. For instance, the Clean Air Act was drafted and enacted in 1968 when pollution was not a major challenge. These days air pollution is a series problem in Addis Ababa. "By the same token government should forecast the future of AI and engage vigorously."

Although AI technologies are at the early stage of adoption in Ethiopia, the health, security, and finance sectors have started adopting these technologies. Particularly, one of the respondents expressed his concern regarding the deployment of the technology in the security system:

> "Currently the National Defence Force has successfully employed a combat drone technology. Unless strict legislative and law enforcing mechanisms are in place, these technologies would fall in the hands of terrorists and corrupt individuals and would cause untold damages."

In general, Ethiopian and other sub-Saharan African countries must foresee the future and prepare themselves to tap the potential of AI technology to improve the prevailing efficiency problems widely observed in the public and private sectors. At the same time, it is high time to investigate ethical and social justice issues to reduce the unintended consequences of AI technologies. Particularly, ethical issues are central to fostering social justice in African society and are salient to dealing with historical inequalities, which tend to be more pronounced by AI unless timely intervention measures are considered.

5. Conclusion

Recognizing polarized views surrounding the capabilities of the technical artifact to carry values, this study has demonstrated how to translate high-level relevant ethical principles into technical design considerations. Building on the notion transpired by Munn (2023), this study has revealed how distributive justice would be realized by embedding salient ethical values into AI systems. Particularly, in the African context "Respect for diversity" and "Ethnic neutrality" are instrumental to instituting justice and reducing discrimination and are suggested to be an overarching principle to foster justice in the society. Therefore, contextual moral values identified in this study such as "Respect for diversity" and "Ethnicity Neutrality" serve as important considerations to nurture social justice. In line with the notion transpired by prior studies in the field, this study regarded value as criteria of excellence in technical design similar to functional efficiency, reliability, robustness, elegance, and safety.

The study has realized relevant findings regarding the systematic process of conducting moral value analysis and modeling. However, it is imperative to acknowledge that the study has limitations in terms of scope and rigor. First, the study revealed only two moral values presumed to be shared across sub-Saharan African countries. Second, although other competing software systems analysis and design methodologies exist, the study used a Use Case Analysis method. While exploring abstract phenomena such as moral values, mixed methodologies tend to produce the desired result.

6. Recommendation

The process of identifying, analyzing, modeling, and embedding moral values in AI systems involves complex sociotechnical inquiries. Therefore, establishing an interdisciplinary team that involves legal professionals, social psychologists, software engineers, and philosophy is needed for sound moral value identification, analysis modeling, and implementation. Comprehensive studies are needed to discover additional African moral values for further deliberation. Moreover, future studies that employ recent software system analysis methodologies, such as simulation and the agile approach are supposed to yield a better result in terms of nailing down abstract values into design parameters.

References

Adams R (2021) Can artificial intelligence be decolonized?. *Interdisciplinary Science Reviews 46*(1–2), 176–197.

Appiah EK, **Arko-Achemfuor A and Adeyeye OP** (2018) Appreciation of diversity and inclusion in Sub-Sahara Africa: the socioeconomic implications. *Cogent Social Sciences 4*(1), 1521058.

Behrens KG (2013) Towards an indigenous African bioethics. *South African Journal of Bioethics and Law 6*(1), 32–35.

Benjamin R (2019) Assessing risk, automating racism. *Science 366*(6464), 421–422.

Buolamwini J and Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77–91). PMLR.

Carr M and Verner J (1997) *Prototyping and Software Development Approaches*. Department of Information Systems, City University of Hong Kong, Hong Kong, 319–338.

Flanagan M, **Howe DC and Nissenbaum H** (2019) *Embodying Values in Technology: Theory and Practice* (p. 322), na.

Floridi L, **Cowls J**, **Beltrametti M**, **Chatila R**, **Chazerand P and Dignum V** (2018) AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines 28*(4), 689–707.

Gaffley M, **Adams R and Shyllon O** (2022) *Artificial intelligence. African Insight: A research summary of the ethical and human rights implications of AI in Africa.*

IWGIA. A Nation without Ethnicity: the Rwandan Reconciliation Model. Available at https://www.iwgia.org/en/news/4556-a-nation-without-ethnicity-the-rwandanreconciliation-model.html. (accessed 05 January 2023).

Jost JT and Kay AC (2010). *Social Justice*: History, Theory, and Research.

Kohlberg L, **Ricks D, and Snarey J** (1984). Childhood development as a predictor of adaptation in adulthood. Genetic Psychology Monographs.

Van de Poel I and Kroes P (2013) Can technology embody values?. In *The Moral Status of Technical Artefacts* (pp. 103–124). Dordrecht: Springer Netherlands.

Pitt JC (2013) Guns don't kill, people kill"; Values in and/or around technologies. In *The Moral Status of Technical Artefacts* (pp. 89–101). Dordrecht: Springer Netherlands.

Van de Poel I (2020) Embedding values in artificial intelligence (AI) systems. *Minds and Machines 30*(3), 385–409.

Manders-Huits N (2011) What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics 17*(2), 271–287.

Metz T (2010) African and Western moral theories in a bioethical context. *Developing World Bioethics 10*(1), 49–58

Moore TJ and Asay SM (2013) Values, attitudes and behaviors: understanding family choices. *Family Resource Management 99*, 110.

Munn L (2023) The uselessness of AI ethics. *AI and Ethics 3*(3), 869–877.

Rahman MM (2018) Effects of moral values on individuals and society and its relation to economic development: an analytical study. *International Journal of Shari'ah and Corporate Governance Research 1*(1), 38–42.

Rawls A (1971) *Theories of Social Justice*. Harvard University Press, Belknap Press.

Robertson T and Simonsen J (2012) *Participatory Design. Routledge International Handbook of Participatory Design* 1, Routledge.

Tessema TA (2021). Ethiopian Ethics or "šine migibari""ābiro menori,"(living with others) as Ethiopian Ethics. *International Journal of Academic Multidisciplinary Research (IJAMR) 5*(11), 93–100.