

AI Ethical Framework: A Government-Centric Tool Using Generative AI

Lalla Aicha Koné¹, Anna Ouskova Leonteva², Mamadou Tourad Diallo³, Ahmedou Haouba⁴, Pierre COLLET⁵

Research Unit of Scientific Computing-Computer Science and Data Science,

University of Nouakchott, Nouakchott, Mauritania¹

ICUBE Laboratory, Strasbourg University, Strasbourg, France^{1,2}

Research Unit of Scientific Computing-Computer Science and Data Science,

University of Nouakchott, Nouakchott, Mauritania^{3,4}

Institute of Technology for Innovation in Health and Wellness

Faculty of Engineering, Université Andrés Bello, Viña del Mar, Valparaiso, Chile⁵

Abstract—Artificial Intelligence (AI) is transforming industries and societies globally. To fully harness this advancement, it is crucial for countries to integrate AI across different domains. Moral relativism in AI ethics suggests that as ethical norms vary significantly across societies, frameworks guiding AI development should be context-specific, reflecting the values, norms, and beliefs of the cultures where these technologies are deployed. To address this challenge, we introduce an intuitive, generative AI based solution that could help governments establish local ethical principles for AI software and ensure adherence to these standards. We propose two web applications: one for government use and another for software developers. The government-centric application dynamically calibrates ethical weights across domains such as the economy, education, and healthcare according to sociocultural context. By using LLMs, this application enables the creation of a tailored ethical blueprint for each domain or context, helping each country or region better define its core values. For developers, we propose a diagnostic application that actively checks software, assessing its alignment with the ethical principles established by the government. This feedback allows developers to recalibrate their AI applications, ensuring they are both efficient and ethically suitable for the intended area of use. In summary, this paper presents a tool utilizing LLMs to adapt software development to the ethical and cultural principles of a specific society.

Keywords—AI ethics; Gen AI; LLMs; moral relativism; ethical norms; adaptive ethical framework

I. INTRODUCTION

Recent advances in artificial intelligence (AI) enable this technology to better solve real-world problems, meaning that beyond being increasingly used by private companies, it is also playing an important role in government operations. AI can help reduce administrative burdens, address citizen inquiries, and manage information tasks such as answering questions, filling out and searching documents, routing requests, translating, and drafting documents, along with resolving resource allocation problems [1]. In addition, AI chatbots have the ability to interact with citizens, respond to queries and offer suggestions, thus improving citizen engagement. Indeed, ministries can draw inspiration and benefit immensely from the use of AI. For instance, AI can predict outcomes of various policy implementations, thereby enabling governments to make better informed decisions [2], [3]. For example, a recent survey of public sector professionals in the UK revealed that 22%

of participants actively use generative AI systems in their work[4].

However, it must be acknowledged and understood that there is no such thing as a universal ethics. Ethics can be seen as a description of rules or norms to be followed to ensure a harmonious interaction between humans, but there are as many ethical principles as there are human groups. What is considered as an ethical behaviour or decision in the US will be different from what is considered as ethical in the EU or in China, Russia, India, North Africa or sub-Saharan Africa. Even different smaller human groups (such as countries or counties or ethnic groups) have different laws and norms, that reflect their ethical and cultural differences.

Generative AI, a subfield of AI, has shown considerable growth. It is now applied for generating many kinds of new content, *e.g.* text, images, music and video [5], [6]. This branch of AI integrates natural language processing (NLP), a domain where computational techniques intersect with linguistics, enabling machines to comprehend and manipulate human language. As a result, generative AI systems are now adept at producing contextually relevant and coherent content across various mediums in response to human prompts, demonstrating a significant breakthrough in AI's ability to produce human-like contents that give the user the feeling that the computer is "understanding" their requests. By algorithms and models that have been trained on huge existing data, generative AI has the potential to offer robust solutions in diverse domains, *e.g.*

- Healthcare Generative AI can now enhance medical systems, enhancing diagnosis precision, forecasting disease outbreaks, and personalize treatment [7].
- Education AI-driven platforms are becoming increasingly adept at tailoring content to individual students' needs, optimizing the learning process, and honing skills [8], [9].
- Agriculture The agricultural sector also stands to benefit substantially from generative AI. Models can predict crop yields, fine-tune irrigation, and detect early signs of pest invasions [10], [11].

Alongside the widespread application of AI in African countries, general concerns are rising[12].AI technologies'

ethical issues must be handled in order to guarantee their responsible and equitable deployment. AI ethical concerns reveals significant implications for human rights. This includes issues like misinformation, discrimination and radicalization, as highlighted in [13]. Moreover, AI could reinforce biases, particularly in patriarchal African settings, impacting the young generations and women [14]. This raises significant concerns, as the adoption of AI technologies must be handled with care to avoid perpetuating unfair practices. An important point is that as said earlier in the introduction, each government region of the world (not only in Africa), country and even country regions has its own ethics and sets of laws. Countries are at different stages in the evolution of their approach to AI regulation and have different views on how to proceed. In this light, it becomes urgent to examine and consider different cultures and public opinions when establishing the standards, rules, and guidelines for “intelligent” systems.

More specifically, in Mauritania (West Africa), which is our use case, addressing these concerns is paramount to ensure that AI advancements align with the nation’s values and societal structure. Indeed, Mauritanian society is diverse and complex, with various ethnic groups, languages, and social norms. Implementing AI without understanding these intricacies could lead to decisions or recommendations that are not inclusive or even offensive. Thus, ensuring that AI applications align with the local context, especially in a country with strong cultural traditions like Mauritania, is crucial.

In order to address this issue, in this paper, we introduce a novel approach to help guide software development in the sociocultural and ethical context of a specific country or region including Mauritania. Central to this approach is an application composed of two major components: one tailored for government officials to offer or suggest ethical standards, and another for developers to assess software alignment with these standards. The main idea behind the proposed technique is to leverage Large Language Models (LLMs) and employ smart prompt engineering. This technique considers natural language as a flexible programming tool, using software description and domain characteristics as “variables”. Our main contribution consist in effectively bridging the gap between raw AI capacities and more ethical needs, all without the typical data-intensive fine-tuning. Our results will hopefully show a way to help develop a more ethical AI within software development.

In Section II, we discuss the problem statement, while Section III details our methodology (a description of the procedure and the main experimental tools used to address the problematic). Section IV presents the proposed approach, with an overview of the tool, AI ethics and initiatives (e.g. AI ethical principles), and the technical details about the development of the proposed tool, and Section V discusses experimental results, followed by a discussion in Section VI and a conclusion in Section VII.

II. PROBLEM STATEMENT

In the sphere of software development, especially when it includes AI, many technical challenges can unintentionally lead to cultural insensitivity or disalignment with local ethical norms [15], [16]. These challenges are especially pronounced

in Mauritania, given its rich tapestry of traditions and values. Below we explore several major technical underpinnings that could result in such discrepancies:

- Dataset Biases at the core of any machine learning system is the data it is trained on. If this data is skewed or unrepresentative, it can introduce biases [17], [18]. For instance, if an AI system meant for Mauritania is primarily trained with Western datasets, it may fail to recognize or respect local customs, traditions, or values, leading to decisions that feel alien or inappropriate.
- Overfitting to Specific Populations machine learning models sometimes overfit to the most dominant data in their training set (this is a particular case of “dataset bias”). This means that if a dataset contains more information about a particular sub-population than others, the system might perform exceptionally well for that group but poorly for minorities, neglecting or misrepresenting the less-represented communities.
- Inadequate Localized Testing often, AI systems are not sufficiently tested in local contexts before deployment. Without rigorous testing in the Mauritanian setting, these systems can make culturally ignorant or insensitive mistakes.
- Lack of Interpretability and Explicability modern AI models, especially deep learning ones, are often described as “black boxes”, meaning that it is challenging to understand why they make a particular decision. Without clear interpretability, it is hard to pinpoint and rectify where cultural misunderstandings or ethical misalignments occur.

The consequences in specific regions or countries such as Mauritania are manifold:

- Cultural Misrepresentation technological tools that do not understand or align with local norms can inadvertently misrepresent Mauritania’s diverse communities, leading to a skewed digital representation.
- Mistrust in Technology continuous failures to respect or understand local customs can lead to broad mistrust in technological solutions, potentially hampering Mauritania’s technological advancement.
- Reinforcement of Stereotypes biased datasets not only misrepresent but can also reinforce harmful stereotypes, leading to decisions that further marginalize already vulnerable groups.

Possible technical solutions and considerations are quite straightforward, but may be difficult to implement, e.g.

- Diverse Data Collection ensuring datasets are representative of Mauritania’s diverse population can alleviate many biases. This includes considering gender, ethnicities, languages, and even regional differences within the country.
- Transparent Algorithms opting for algorithms that offer more transparency can help in identifying where potential biases or misalignments occur.

- Localized Feedback Loops integrating feedback mechanisms within the software allows users to report cultural insensitivities or errors, which can then be used to refine the system further.

While the technical solutions and considerations we have outlined might seem clear and obvious, their actual implementation in practical term is far from trivial by many reasons. Local ethical principles are not mere lists of do's and don'ts. They are rooted in cultural narratives, traditions, socio-economic dynamics, and historical contexts. Capturing these nuances algorithmically is challenging. Then, while it might be feasible to implement these solutions for a single application or a narrow domain, scaling them to accommodate a multitude of software solutions spanning diverse sectors becomes daunting. Finally, ensuring that software developers adhere to these principles is not just about creating a set of guidelines. Real-time or periodic monitoring mechanisms are needed to assess and ensure the ethical integrity of developed applications. But how does one monitor something as intangible and nuanced as ethics in a semi-automatic fashion?

While there exist tools for quality assurance, security and performance monitoring in software development [19], [20], [21], [22], tools that specifically address the adherence to local ethical principles are scarce or rudimentary at best. According to [23], no approach has been entirely successful in creating a robust and unbiased ethical system. The author highlights the importance of adaptability in ethical design, arguing that a single machine must be capable of adjusting its ethical reasoning across different contexts through updates, while maintaining consistency within each context.

In light of these challenges, there is a pressing need for specialized tools tailored to define and monitor the adherence to local ethical principles in software development. Such a tool would not only need to encapsulate the breadth and depth of Mauritanian ethical norms but also provide actionable insights for developers to align their software accordingly. It could also be used elsewhere in the world.

III. METHODOLOGY

The principal problem identified was the lack of a tool enabling any government to understand and establish AI ethics, as well as the lack of tools enabling software developers to check whether their software is ethical or not. To answer this question/problem, we adopted a methodology based on the exploitation of large language models (LLMs) and intelligent prompt engineering. The natural language is considered as a flexible programming tool.

Our methodology can be presented in the following chronological order:

- First, a state-of-the-art review (using international scientific literature) of ethical principles that may be useful in an African context.
- Then develop two web applications:
 - One to help governments formulate and choose their AI ethics, based on the ethical principles chosen above. The ethical propositions elaborated thanks to this application could then

be published by the institution for more transparency and to help developers align with the requests.

- A second application for developers to check whether their software is in line with government ethics.

The prompts used to question the LLM contain variables (e.g. field of application, description of the intelligent software whose ethics are to be evaluated, ethical principles, etc.).

IV. PROPOSED APPROACH

A. AI Ethics and Initiatives

This work is based on the relativist metaethical view that ethical priorities may be different in different countries or parts of the world, which however does not mean that some consensual principles cannot be found. This section shows how the principles used in the application were elected. Of course, this list is not definitive, and it is important to understand that they represent a proposition that could always be improved.

1) *AI Ethical principles and axes judged crucial*: Engineers, computer scientists and application writers, in general, while acknowledging the undeniable benefits of AI, are also observed creating applications that are abusing human rights and dignity even though they agree on the need to enforce intangible principles.

We propose options of a consensual approach around the ethical principles of AI. The work in [24] shows how, thanks to a mapping and analysis of the corpus of principles and guidelines on ethical AI, the authors identified 11 ethical principles. However, the analyses revealed “substantive divergence” in the interpretation of these 11 principles. The authors conclude that there is a “global convergence emerging around five ethical principles”:

- transparency
- justice and fairness,
- non-maleficence,
- responsibility and
- privacy

These principles are referenced in over half of the sources. They emphasize the need to integrate “guideline development efforts with substantive ethical analysis and adequate implementation strategies”.

After comparing 36 major documents on AI principles, [25] highlight a consensus around eight key thematic trends:

- Privacy
- Accountability,
- Safety and Security,
- Transparency and Explainability,
- Fairness and Non-discrimination,
- Human Control of Technology,
- Professional Responsibility, and

- Promotion of Human Values.

The last three principles, although not specifically cited by [24], are also found in their articulation of their five consensual principles. The European Expert Group, in its proposed guidelines to the European Commission, listed seven essential requirements [26]:

- Human factor and human control
- technical robustness and security,
- privacy and data governance,
- transparency,
- diversity, non-discrimination, core equity,
- societal and environmental well-being,
- responsibility.

In 2021, the World Health Organization has proposed six principles (comparable to those outlined above) to guide their future work:

- human autonomy
- human well-being and safety and the public interest,
- transparency, explainability and intelligibility,
- responsibility and accountability,
- inclusion and equity, and
- responsiveness and sustainability.

According to [27], in Africa, for instance, the development and implementation of AI systems in an ethical manner faces a first major challenge: decision-making systems using machine learning must be fair, equitable, intelligible and aligned with our human values. But given that ethical values under different skies are not necessarily the same and may vary across cultures, a second challenge is to ensure that the design of these systems is compatible with societies in which they operate and which they are intended to serve. The framework defined in ¹ identifies five guiding principles, echoed by its white paper [26], for a trustworthy AI. Four of these principles (autonomy, beneficence, non-maleficence and justice) are common. The fifth principle, explainability, is specific to AI.

- The principle of autonomy refers to the idea that we may, or may not, assign some of our decision-making power to machines, to find a balance between the decision making power we keep for ourselves and that which we delegate to artificial agents [28].
- Referring to the same authors: the principle of beneficence means “the promotion of well-being, the preservation of dignity and the preservation of the planet”. In other words, the development of AI that benefits humanity.
- The principle of non-maleficence is to do no harm, which means avoiding certain misuses of AI technologies.

- Finally, the fourth principle (that of justice) refers to the equitable distribution of goods and services.
- Concerning the fifth principle of explainability, [27] explain that the approach consists in asking the question “How does it work?” (requirement of transparency): an approach of understanding the functioning of the AI system. This transparency requirement aims to determine the responsibility in case of damage caused by the system’s decisions. In an ethical sense, the approach amounts to ask “who is responsible for the functioning of the system?” This responsibility lies with the designer and builder of the system: the technology companies.

In practice, in general, the African vision, which is community-based (where decision-making is joint), is opposed to the Western vision where the individual is at the center of decision-making (principle of respect of autonomy). [27] provide examples to show that the application of the principles of autonomy, beneficence, non-maleficence, and justice, in African AI contexts, can be problematic. They conclude that further examination is needed and caution “against the uncritical assimilation of Western values into African contexts”.

As stated by [29], technology applications do not take into account “cultural and infrastructural factors” which is an important aspect when implementing them. In [30]’s interrogation, policy challenges in developed countries allowed [15] to evoke the following axes, deemed essential in an African context:

- Equity, partiality and responsibility.
- Loss of jobs and tax revenue through automation.
- Cultural and linguistic diversity.
- Surveillance and loss of privacy.
- Democracy and political self-determination.

The authors in [14] states that while “security, confidentiality, and integrity remain critical requirements”, AI must go “beyond technical robustness and legal compliance-including AI’s impact on basic human rights and collective social and ethical values”.

Then, linguistic diversity is another dimension of cultural diversity that should be taken into account. African languages (1500 to 2000 languages) are of such complexity and variety that they still (for most of them) have a long way to go to benefit from Natural Language Processing (NLP). NLP is a branch of AI. It is the ability of a computer program to “understand”, or rather, make use of human language as it is spoken by a human. Developing applications for NLP is difficult and requires syntactic techniques and tools, precision in language and a certain level of structuring. Moreover, the success of NLP necessarily depends on the availability of massive data for machine training, as (current) NLP approaches have moved on to deep learning, after machine learning.

Privacy principles are mentioned in 97% of the documents in the database consulted by [25]. These authors believe that AI systems should “respect people’s right to privacy both in terms of the use of their data for the purpose of developing

¹<https://rm.coe.int/cahai-2020-08-fin-fr-mantelero-binding-instrument-report-completed/16809eed6d>

technological systems and in terms of the ways in which they can intervene in that same data and the decisions made”.

Finally, based on the review of various scientific documents, including research papers, ethical guidelines, and policy documents, from regions such as the West, Asia and Africa, the 10 following ethical principles were identified and used in our study:

- Transparency
- Responsibility
- Non-Maleficence
- Equity
- Privacy and Confidentiality
- Societal and Environmental Well-being
- Human Autonomy
- Reactivity and reliability
- Cultural and Linguistic Diversity
- Democracy and Political Self-Determination

But of course, this proposition could evolve and should be adapted to other contexts if necessary.

B. Technical Details

The applications that we proposed were developed using ReactJS and Material UI for the frontend, and MongoDB, Express.js, and Node.js for the backend.

- ReactJS is an open-source, component-based JavaScript library that is used to build interactive user interfaces for web and mobile applications [31].
- Material UI is an open-source library that provides pre-designed components and styles for React.
- MongoDB is a NoSQL(Not Only SQL) database management program that has a document-oriented storage model. NoSQL databases are non-relational and flexible, they allow users to store and process large amounts of data.
- Node.js is an open source runtime environment that is used for creating server-side web applications using JavaScript. It uses an asynchronous event-driven model [32]. Express.js is a Node.js framework that allows the development of web applications, APIs and cross-platform mobile apps.

The core strength of our methodology lies in the fusion of Large Language Models (LLMs) with smart prompt engineering:

1) *Natural Language as a programming tool:* By treating natural language as a dynamic programming medium, we harness the expansive capability of LLMs to make propositions in human readable natural language. Trained on extensive datasets, LLMs excel in tasks such as content creation and question-answering [33]. A key feature of LLMs is the attention mechanism, which improves their ability to generate contextually appropriate responses. These models are typically

based on the “Transformer architecture,” a neural network framework optimized for language tasks [34]. In this study, we used two OpenAI models: GPT-3.5-turbo to build a chatbot and GPT-3.5-turbo-instruct to propose specific questions to developers, as well as to establish ethical standards and identify possible existing correlations between ethical principles. These two models are accessible via the OpenAI API (<https://platform.openai.com/>). OpenAI’s GPT-3 is an autoregressive language model family capable of performing human-like text completion tasks. The GPT-3.5-turbo version has 175 billion parameters and was trained on a variety of permitted and public documents [35].

GPT-3.5-turbo was used to build our chatbot because it is a chatty model that will provide responses beyond what is specifically asked. The instruct model, however, is much more terse and concise, performing exactly as instructed. The /completions endpoint (used for GPT-3.5-turbo-instruct) provides the completion for a single prompt and takes a single string as input, whereas the /chat/completions endpoint (used for GPT-3.5-turbo) responds to a given dialog.

These models have a parameter called temperature, which is playing a vital role in our prompts. It is an important setting as it influences the variability of the generated responses. As the temperature approaches zero, the model will become deterministic and repetitive. With this setting, we can fine-tune the model according to the desired level of creativity [36].

2) *Smart prompt engineering:* Instead of traditional fine-tuning, which requires vast datasets and often masks the decision-making process, smart prompt engineering benefit from LLM’s vast knowledge while guiding it with precision. This ensures our tool remains adaptable and explicable. For instance, given a prompt, the LLM crafts questions that probe the software’s ethical alignments, drawing from the software and domain descriptions. The type of prompt that is used in this study is Zero-shot prompts. In prompt engineering, “Zero-shot” is when we give a task to a language model like GPT-3 for instance, without any prior examples or training specific to that task. Essentially, the model have to accomplish the task based entirely on its pre-existing knowledge and understanding, which it acquired during its initial training phase. As opposed to “few-shot” or “one-shot” learning, where the model is given one or a few examples to guide its response. Zero-shot learning tests the model’s ability to generalize.

Formula for calculating the percentage of compliance with AI ethics. In order to verify if the software developer did respect the AI ethics established by the government, the Eq. (1) was proposed. this formula is designed to calculate the percentage of compliance with the ethics and subsequently verify if the developer has met the minimum percentage value established by the government. The formula is based on the weightings that was assigned by the government to the ten ethical principles and the developer’s responses to the five questions per principle.

$$P = \frac{\sum_{i=1}^{10} \frac{R_i W_i}{5}}{\sum_{i=1}^{10} W_i} \cdot 100 \quad (1)$$

Where W_i is the weighting of the i -th ethical principle. R_i is the developer's response to the i -th ethical principle.

In the formula, each response of the developer is divided by 5 in order to normalize the answer to a scale from 0 to 1, where 0 means no positive response (complete non-compliance) and 1 means the developer answered "yes?" to all questions (complete compliance). Then these normalized values are multiplied by the respective weights assigned by the government, and then the weighted average is calculated.

C. Application Overview

Our application (see Fig. 1) is specifically tailored to bridge the ethical divide in software development, offering tools to both government officials and software developers. The main idea behind this application is to enable the government, on one hand, to establish a list of ethical principles and assign a weighting to each principle. These weightings are specific to each ministry or domain. They could be elaborated by ethical committees for each domain, sub-population or context.

Subsequently, developers could, through specific questions for each principle, verify if their software respects the government's ethical norms or preferences of the context in which the application will be used. Indeed, each ministry / domain / context (health, education, etc.) could have its specific weighting of principles, and every software developer would need to comply with the ethics established by the ministry to which their software is linked for getting an approval. The questions posed to the developer are proposed automatically by an LLM (Large Language Model) based on the application domain (health, education, etc.) of their software and the description of their software. Each ministry could review, and then validate or reject the developers' responses. Subsequently, the developer could re-calibrate their application based on feedback from the respective ministry. This iterative process would allow developers to review and refine their software to ensure it finally complies with the government's ethics, based on further feedback and comments. This application ensures scalability because it can be applied across various domains. But, in this paper, only an example about Healthcare was presented. This could be seen as a way for top-down authoritarian governments to enforce their political views, but it must be understood that such guidelines are constitutive of all governments: even in the most liberal democracies, software that would promote racist or paedophilic contents are rightly strongly prohibited, so such ethical enforcements are necessary and beneficial to any society.

The use of an LLM has been identified as a promising method to not only facilitate the generation of appropriate questions for developers but also to help the government better understand and formulate its ethics optimally and to propose standards to be followed at the level of each ministry. The participation of an expert committee in this process is strongly encouraged. This can guarantee consistency and representative ethics within each ministry but also across ministries.

It is designed to work in two complementary parts (Government-Centric Application and Developer Diagnostic Application).

1) *Government-Centric application*: This module allows government officials to dynamically propose and adjust local ethical norms according to a domain-specific context. These guidelines are defined in the form of clear, natural language prompts, helping for transparency in ethical expectations. Below, the list of sections and features:

a) *Section "Chatbot"*: Users can ask questions and request information about AI ethics via a Chatbot. This LLM-based Chatbot aims to guide the government in understanding and formulating ethics within each ministry/domain. The following message was used to provide guidance on how the GPT-3.5-turbo model should behave for the conversation : "You are an intelligent assistant designed to help users understand AI ethics, become familiar with AI ethical principles, and select the most suitable ethical principles for their specific context. Your goal is to provide clear, comprehensive, and practical guidance."

b) *Section "Define an ethic"*: Government officials can record their expectations regarding ethics via a questionnaire. The questionnaire is domain- or ministry- specific. This section contains three steps:

- Selection of the application domain : Health, education, finance, agriculture. This is the domain to which the concerned ministry is linked.
- Weighting of ethical principles (see Fig. 4). In this step, it is possible to consult a proposed list of principles (principles that we choose during the state-of-the-art step), see the description of each, and assign a weighting to each principle according to its importance in the concerned domain. This task of weighting is made easy thanks to a text proposed by the LLM that helps users to find out which principles are essential and which are less important in the context of the specific application domain. Indeed, the importance of ethical principles can vary from one domain to another.
- Specifying a threshold: The minimum value of the "percentage of ethics compliance" that the developer must achieve for their software to conform with established ethics. Each ministry must have its set of thresholds as some domains are more sensitive than others. The compliance threshold has been identified as playing a crucial role in the process of evaluating governmental ethics. It represents the government's will to implement ethical principles and allows civil servants to assess to what extent the software respects established ethical values.
- In the end, a summary of the provided informations is displayed, along with some suggestions:
 - The previously entered weightings of the principles.
 - A list of ethical standards proposed by the LLM that the developer should implement during the software development process.
 - An overview of the correlations between different ethical principles. These correlations are proposed by the LLM as correlations exist between different ethical

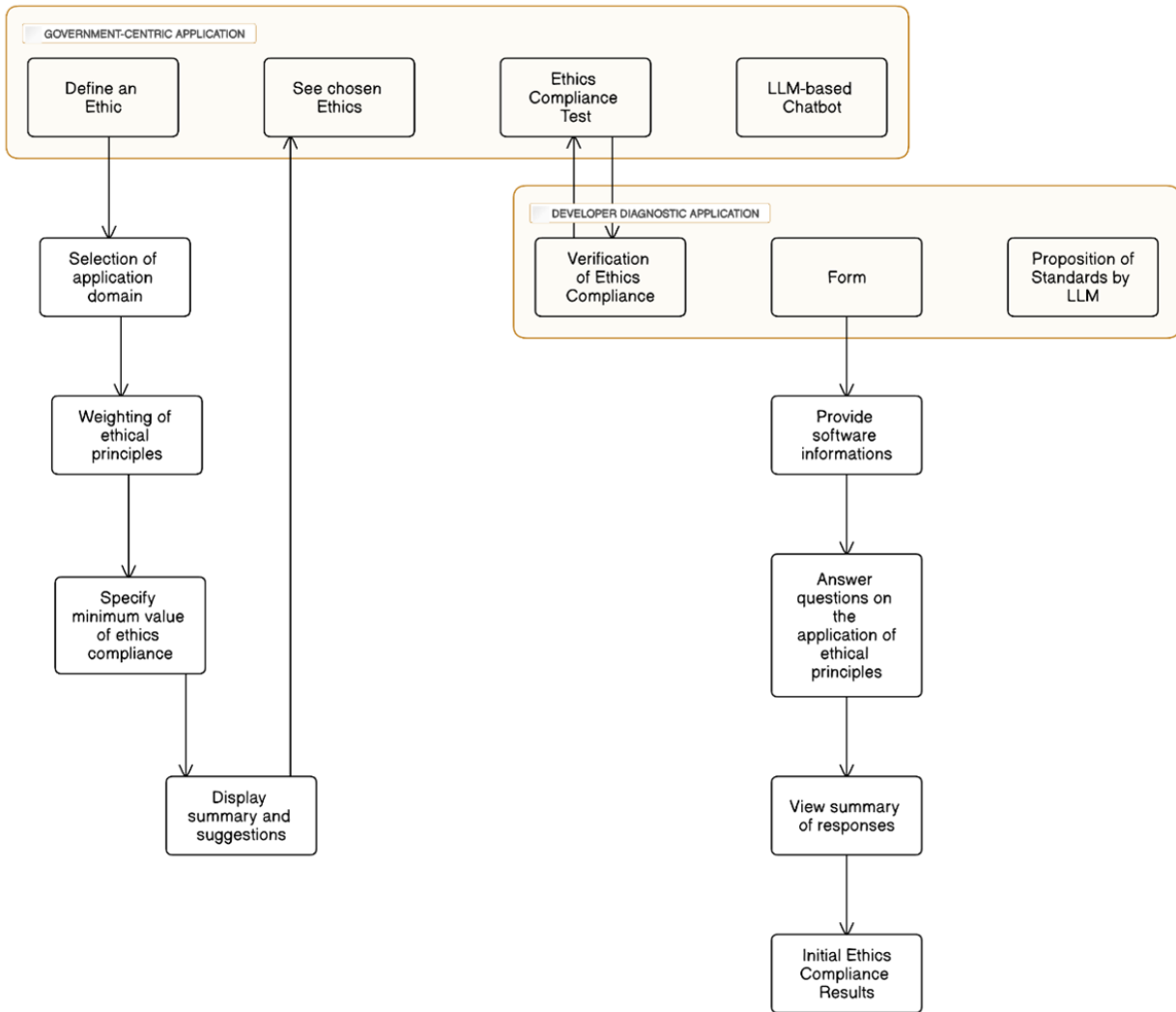


Fig. 1. Flowchart of the proposed approach.

principles. They are used to compute a given software's compliance with established ethics. For example, if there is a correlation between the transparency principle and the responsibility principle, and the weighting of transparency is respected but that of the responsibility principle is not, compensation can then take place, allowing the developer's software to be valid even if it is below the minimum value of the responsibility compliance threshold. The way the compensation is done is open to, and can be modified by the user, with the possibility to implement strong thresholds that cannot be compensated for.

c) *Section "See chosen ethics"*: The user can consult all the information they have entered in the "Define an ethic" form, as well as the list of standards and existing correlations between principles.

d) *Section "Ethics Compliance Test"*: Through this interface, each ministry can review developers' responses to

an ethics compliance form (this is a form accessible via the developer module). The ministry validates these responses if they comply with ethics and rejects them if they are not, with a message to the user so that he/she can improve their application before re-submission.

2) *Developer diagnostic application*: This tool serves as an ongoing ethical audit mechanism. By processing software characteristics and responses against the predefined prompts, it provides feedback on potential ethical misalignments. Below the list of sections and features:

a) *Section "Standards"*: The developer can consult the standards to apply during the software development process. This helps developers to acquire a better understanding of ethical issues. This list of standards was previously elaborated using the help of an LLM and validated by the government via the government module.

b) *Section "Form"*: The developer can check whether their software complies with the government-established ethics through a questionnaire. It takes place in three stages:

- First, the developer provides the following information: software name, application domain, and description.
- Then, they must answer “yes / no” questions regarding each ethical principle’s implementation. Five questions are asked for each principle. These questions are proposed by an LLM educated on ethics, based on the software description, application domain, and the principle in question, ensuring a thorough verification of whether ethics are correctly incorporated into the software development process.
- Finally, the developer can view a summary of their responses:
 - A radar chart displaying both the developer’s response results and the government’s weightings. This chart enables a clear comparison between the developer’s performance and government expectations.
 - The number of points obtained for each principle, the name, domain and description of the software. The developer’s response to this questionnaire must be validated solely by an expert committee on AI ethics within the government or the respective ministry.

c) Section “Ethics Compliance”: The developer can review feedback from the ministry to which their software’s domain is attached. And see if their software is approved (meets the government’s expectations in terms of AI ethics) or not.

V. EXPERIMENTAL RESULTS

Prompts used in this study and their outputs:

- Prompt for proposing five questions about the application of each ethical principle (see Fig. 2) : the objective of this prompt is to propose five questions the developer must answer. Each question is about the implementation of each ethical principle. The variables: “descApp” (software description), “domain” (domain of application), “principleName” (name of the ethical principle) are given as an input.

```
const prompt = "You're an expert in Artificial Intelligence ethics. Here's the context: " +
  "the Mauritanian government has a list of ethical principles for Artificial " +
  "Intelligence applied in a given domain. The Mauritanian government wants to " +
  "ensure that any developer of intelligent software respects these principles " +
  "and implements them. To ensure that these principles are put into practice, " +
  "the government wants to set up a system that generates five questions per " +
  "principle. These questions are used to verify, in detail, the implementation " +
  "of each principle. The system takes as input the description of the software, " +
  "the field of application, and an ethical principle of artificial intelligence, " +
  "and gives as output a list of five dichotomous questions on the implementation " +
  "of that ethical principle. Generate a list of five questions, taking into account " +
  "the following information: software description: " + descApp + "; application domain: " +
  domain + "; ethical principle: " + principleName + ". Enclose the list in braces like the " +
  "following JSON format: {\\"questions\\": [{" " +
  "\\"id\\": 1, \\"name\\": \\"q1\\", \\"label\\": \\"question one\\"}, {" " +
  "\\"id\\": 2, \\"name\\": \\"q2\\", \\"label\\": \\"question two\\"}, {" " +
  "\\"id\\": 3, \\"name\\": \\"q3\\", \\"label\\": \\"question three\\"}, {" " +
  "\\"id\\": 4, \\"name\\": \\"q4\\", \\"label\\": \\"question four\\"}, {" " +
  "\\"id\\": 5, \\"name\\": \\"q5\\", \\"label\\": \\"question five\\"}]}";
```

Fig. 2. Prompt that propose questions for developer

descApp = “IBM Watson for Oncology is an AI-powered software designed to assist oncologists in diagnosing and creating personalized treatment plans for

cancer patients. Developed by IBM in collaboration with Memorial Sloan Kettering Cancer Center, this software leverages the power of AI to analyze large volumes of medical literature, patient data, and clinical trial information to provide evidence-based treatment recommendations”, domain/Field = Healthcare, temperature = 0

- Output of the prompt when principal name is Transparency (the output is in the format JSON):

```
"questions":
[ "id": 1, "name": "q1", "label":
"Does the software provide
a clear explanation of how
it arrives at its treatment
recommendations?", "id": 2,
"name": "q2", "label": "Is
the software able to provide
a breakdown of the data and
sources used to generate its
recommendations?", "id": 3,
"name": "q3", "label": "Does the
software disclose any potential
biases or limitations in its
recommendations?", "id": 4,
"name": "q4", "label": "Is
the software able to explain
the reasoning behind its
recommendations in a way that is
understandable to non-technical
users?", "id": 5, "name": "q5",
"label": "Does the software
provide a mechanism for users to
access and review the data used
to generate its recommendations?"
]
```

- Output of the prompt when principal name is Responsibility:

```
"questions":
[ "id": 1, "name": "q1",
"label": "Does the software
accurately and reliably analyze
large volumes of medical
literature, patient data, and
clinical trial information to
provide evidence-based treatment
recommendations?", "id": 2,
"name": "q2", "label": "Does the
software take into account the
individual needs and preferences
of each patient when providing
treatment recommendations?",
"id": 3, "name": "q3", "label":
"Is the software transparent
in its decision-making process,
providing clear explanations for
its treatment recommendations?",
"id": 4, "name": "q4", "label":
"Does the software have
mechanisms in place to ensure
patient data privacy and
security?", "id": 5, "name":
"q5", "label": "Is the software
regularly monitored and updated
to ensure its accuracy and
effectiveness in providing
treatment recommendations?" ]
```

- Output of the prompt when principal name is Privacy and confidentiality:

```
"questions":
[ "id": 1, "name": "q1",
```



```
"label": "Does the software have a clear and transparent privacy policy that outlines how patient data will be collected, used, and protected?", "id": 2, "name": "q2", "label": "Is the software designed to only collect and use patient data that is necessary for its intended purpose?", "id": 3, "name": "q3", "label": "Does the software have built-in security measures to protect patient data from unauthorized access or breaches?", "id": 4, "name": "q4", "label": "Is the software regularly audited and updated to ensure compliance with privacy and confidentiality regulations?", "id": 5, "name": "q5", "label": "Does the software have a mechanism for patients to access, review, and request changes to their personal data collected by the software?" ]
```

- Prompt for proposing a list of norms to apply when developing an intelligent software (see Fig. 3). The following variables are given as an input: “listPrinciples” = list of ethical principles, “domain” = domain of application, “notations” = the weighting given by the government to ethical principles. Output of the

```
const prompt =
"Act as an expert in Artificial Intelligence ethics. " +
"You are familiar with the regulations governing the implementation " +
"of AI ethical principles in all fields. Here's the context: the " +
"Mauritanian government has a list of AI ethical principles applied " +
"in a given field. The government has assigned a weighting between " +
"1 and 5 to each principle. These weights represent the importance " +
"of these principles for the government. The government then wants " +
"to use these weights to generate two standards per ethical principle " +
"to ensure that any intelligent software developer respects government " +
"ethics. In other words, the software developer must respect and implement " +
"these ethical principles according to their weightings. If an ethical " +
"principle has a high weighting, this means it is considered particularly " +
"crucial. The standards associated with crucial ethical principles should " +
"therefore be more stringent and demanding. Conversely, if a principle " +
"has a low weighting, the corresponding standards may be less stringent. " +
"It should also be noted that these standards will change from one field " +
"to another. Software developers must respect the following ethical principles: " +
listPrinciples + ". They must respect these principles according to the following " +
"weightings: " + JSON.stringify(notations) +
". Generate a list of standards to follow (two standards per principle) that " +
"this developer must respect, given that their software concerns the domain: " +
domaine + ". Take into account that weightings influence the generation of norms by " +
"highlighting the relative importance of ethical principles. Do not mention the word " +
"\norm" in the list you're about to generate, and do not include the question in your " +
"answer. Enclose the list in braces like the following JSON format: " +
"({'principles': [{'name': 'principle's name', 'weight': weight, 'norms': " +
"[\norm one\", \norm two\"]}]});
```

Fig. 3. Prompt that propose a list of norms.

prompt that propose a list of norms:

Transparency:

- Provide clear and accurate information about the AI software and its capabilities to healthcare professionals and patients
- Ensure that the AI software does not hide or manipulate information from healthcare professionals and patients

Responsibility:

- Ensure that the AI software is properly maintained and monitored to prevent any potential harm to patients

- Establish clear lines of responsibility and accountability for the development and use of the AI software

Non-maleficence:

- Conduct thorough risk assessments and testing to identify and mitigate potential harms caused by the AI software
- Implement safeguards and protocols to prevent or minimize potential harms to patients from the AI software

Equity:

- Ensure that the AI software does not perpetuate or amplify existing inequities in the healthcare system
- Regularly assess and address any potential biases or discrimination in the AI software that could impact patient outcomes

Privacy and confidentiality:

- Implement strong data privacy and security measures to protect patient information collected and used by the AI software
- Ensure that the AI software adheres to all relevant privacy laws and regulations

Societal and environmental well-being:

- Consider the potential societal and environmental impacts of the AI software and take steps to mitigate any negative effects
- Regularly evaluate the AI software's overall contribution to the well-being of society and the environment

Human autonomy:

- Ensure that the AI software does not impede or restrict patient autonomy in making healthcare decisions
- Provide options for patients to opt-out or request human intervention in the use of the AI software for their care

Reactivity and durability:

- Regularly monitor and update the AI software to ensure it remains relevant and effective in the rapidly changing healthcare landscape
- Implement a system for receiving and addressing feedback and concerns from healthcare professionals and patients about the AI software

Cultural and linguistic diversity:

- Ensure that the AI software is accessible to and inclusive of individuals from diverse cultures and linguistic backgrounds
- Conduct thorough testing and adaptation of the AI software to account for cultural and linguistic differences in healthcare practices

Democracy and political self-determination:

- Ensure that the development and use of the AI software is aligned with democratic principles and political self-determination
- Ensure transparent decision-making processes and active stakeholder engagement in the development and use of the AI software.

- Prompt for proposing the existing correlations between ethical principles (see Fig. 4): If the percentage of compliance of the developer is below the minimum value of “percentage of compliance with AI ethics” set by the government, the correlations between the ethical principles might become significant. These correlations can help government officials to decide whether to approve or not the developer’s software. For instance, if there’s a correlation between Principle A and Principle B, and the developer has closely followed the guidelines for Principle A but not for Principle B, then adherence to Principle A can compensate the lack in Principle B.

```
const prompt =
"You are an expert in Artificial Intelligence ethics. " +
"You are familiar with the different principles of AI ethics " +
"and understand the correlations that may exist between these principles " +
"depending on the domain of application of AI. " +
"I would like you to help me determine whether there are correlations " +
"between the AI ethical principles I'm about to provide you with. " +
"To identify these correlations, consider the domain of application of AI. " +
"The result should indicate the degrees of correlation—strong, weak, or non-existent— " +
"between the ethical principles. Provide a coherent and detailed analysis " +
"by exploring correlations between the following AI ethical principles in the domain of " +
domaine + " : " + listPrincipes + ". " +
"Analyze all possible combinations of correlations. Your answer should only include " +
"the combinations in which the correlations are \"strong\". " +
"Do not include correlations with the degree \"weak\" or \"non-existent\". " +
"Justify your answer by explaining the reasons behind these correlations. " +
"Enclose your answer in braces using the following JSON format: " +
"{\"correlations\": [{\"key\": \"unique identification key\", \" +
\"principle1\": \"name of the first principle\", \" +
\"principle2\": \"name of the second correlated principle\", \" +
\"reasons\": [\"justification for the correlation\"]}]};
```

Fig. 4. Prompt for the existing correlations between ethical principles.

Output of the prompt for the existing correlations between ethical principles: see Fig. 5.

AI-assisted calculation of correlations

Taking into account the selected domain, here are the combinations of strong correlations existing between ethical principles. :

Transparency--Responsibility
Both principles are essential for ensuring accountability and trust in the healthcare domain. Transparency allows for the responsible use of AI in healthcare, while responsibility ensures that AI is used ethically and with consideration for potential harm.

Non-maleficence--Equity
Both principles prioritize the well-being of individuals and communities. Non-maleficence ensures that AI is not causing harm, while equity promotes fair and just treatment for all individuals in the healthcare domain.

Privacy and confidentiality--Human autonomy
Both principles protect the rights and autonomy of individuals. Privacy and confidentiality ensure that personal information is not misused, while human autonomy allows individuals to make their own decisions about their healthcare.

Societal and environmental well-being--Reactivity and durability
Both principles consider the impact of AI on society and the environment. Societal and environmental well-being promotes the use of AI for the greater good, while reactivity and durability ensure that AI is adaptable and sustainable for long-term use.

Cultural and linguistic diversity--Democracy and political self-determination
Both principles recognize the importance of diversity and inclusivity in the healthcare domain. Cultural and linguistic diversity ensures that AI is sensitive to different cultural backgrounds, while democracy and political self-determination allow for the involvement of diverse voices in decision-making processes.

RETURN CONFIRM

Fig. 5. The summary (correlations between different ethical principles, proposed by the LLM). Because of the translation tool Responsibility became Accountability.

VI. DISCUSSION

As the industrial revolution allowed humans to delegate hard work to machines, the AI revolution will allow us to delegate to AI more and more white-collar tasks: bureaucratic at first but as AI will improve, it will be given increasing decision-making power. In medicine, for example, it is now accepted that AI is better and faster at spotting developing cancers on mammographies. This is great because it will save time to radiologists, who can then concentrate on more difficult cases.

AI assistants are becoming acting assistants, with www.jace.ai, “Your new AI employee” that will act for you: “Don’t just chat, start acting today”. On March 13th 2024, the Members of the European Parliament (MEP) adopted the Artificial Intelligence Act, setting safeguards on general purpose artificial intelligence, limiting the use of biometric identification systems by law enforcement, banning the use of AI for social scoring and manipulating or exploiting user vulnerabilities, and giving consumers the right to launch complaints and receive meaningful explanations.

All this means that societies are becoming increasingly aware of the risks posed by AI, but companies too: in order to get an “app” validated by Apple for upload on their appStore, the app needs to be “approved” and for this, the developer must follow hundreds of guidelines limiting the app to minimize its access to the contents on the phone or tablet on which it will be used.² For example, developers must: “Use AppStoreSettings to manage a user’s App Store settings. You can set a maximum age rating for apps, deny in-app purchases, and require passwords for purchases.”

This is what this work attempts at proposing: a tool to help governments formulate their policies and guide app developers, all this with the help of Large Language Models.

The transparent design that we propose promotes clarity in interaction, accountability in ethical deviations, and empowers stakeholders with informed decision-making capabilities. These are significant strides toward responsible AI development in a culturally-sensitive domain. The major advantage is its emphasis on explicability:

1) *Transparency by design*: By utilizing natural language as the medium of interaction and guidance, stakeholders can easily understand, adjust, and interpret the ethical guidelines and the feedback generated.

2) *Ethical accountability*: When the system identifies a potential bias or misalignment, it does not just flag it but also provides a contextually relevant explanation, making it easier for developers to pinpoint and rectify the root cause.

3) *Informed decision-making*: As the government and developers interact with the tool, they are consistently informed about how the software aligns with ethical norms, ensuring conscious and informed decisions throughout the development lifecycle.

However, as with all pioneering methods, it has its inherent challenges, which can be presented in terms of the following disadvantages:

²https://developer.apple.com/documentation/managedsettings/managedsettingsstore/appstore?changes=_6

- Bias from Language Models: Even with smart prompt engineering, there is a possibility that biases inherent in the LLM (acquired from the data it was trained on) could influence its feedback.
- Scalability Issues: As governments adjust ethical standards and as the volume and complexity of software applications increase, the system might face scalability issues, potentially slowing down feedback times.
- Lack of Ground Truth: Using natural language as a programming tool lacks a strict “ground truth” in the way traditional programming does. Hence, there is a level of interpretative ambiguity.

We, therefore, try to propose ideas:

- Enhanced Training: The LLM could be trained further using Mauritanian cultural and linguistic data. This would enhance its understanding of local nuances.
- Continuous Feedback Loop: Encourage continuous feedback from users, especially when they notice discrepancies or biases. This would help in refining the system over time.
- Incorporate Domain Experts: Engaging domain experts in sectors like healthcare, finance, etc. can help in fine-tuning the responses and ensuring domain-specific accuracy.

Consider the description of the software to be checked and its specific domain as contextual anchors; by integrating these as prompts, we guide the model’s responses in a manner that is intricately aligned with the software’s functionality and its ethical considerations within that domain. This is akin to “fine-tuning on-the-fly”: the model’s vast knowledge is channeled and constrained by the prompt, ensuring relevance and accuracy without the need for additional training data.

Moreover, this method underscores a significant advantage: it is adaptable. As software evolves or if there are shifts in domain-specific ethical standards, the prompts can be adjusted, thus ensuring that the AI’s outputs remain congruent with the changing landscape. In essence, by leveraging smart prompt engineering, we are making a case that with the right prompts, models like LLM can be “programmed” in real time using natural language, bridging the gap between generalized AI knowledge and specific, localized requirements.

The advantages of the proposed approach compared to some existing methods is that, compared to existing frameworks such as the G7 toolkit³ and the U.S. Intelligence Community’s AI Ethics Framework⁴, which address AI ethics broadly but lack a focus on generative AI, our proposed approach is distinct in its specific adaptation for generative AI tools. This model is dynamic, allowing government bodies to calibrate ethical standards flexibly across domains and adapt to evolving cultural contexts. Furthermore, the integrated government-developer feedback mechanism supports continuous alignment with ethical standards, which is typically absent

³https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/10/g7-toolkit-for-artificial-intelligence-in-the-public-sector_f93fb9fb/421c1244-en.pdf

⁴<https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community>

in other frameworks. By utilizing LLMs, this tool also offers a novel, practical approach for real-time ethical compliance, reducing reliance on resource-intensive manual assessments.

VII. CONCLUSION

We present a possible solution to the challenges of modern software development in a sensitive sociocultural context, leveraging the strengths of large language models, and emphasizing explicability as its cornerstone. At the intersection of AI and ethics, our LLM-based tool serves as a critical bridge, particularly in the Mauritanian context.

Technically, our tool has two major components:

- For the government, it’s a dynamic platform to set the ethical standards they want to promote. By tuning these standards, they can directly shape the AI’s desired behavior, ensuring a permanent alignment with local values, all the while avoiding unintentional amplification of biases. Inserting their ethical wishes and priorities in the tool will force governments to think about them and integrating them in the tool will be equivalent to publishing them, resulting in ethical norms to be officially clear and known by all. So, the tool is there too to help the institution to refine their norms.
- For developers, it is a diagnosis tool. It allows them to check if their software aligns thanks to a set of domain specific questions, generated by LLM. The LLM engineering provides smart prompts to maximize robustness: if biases exist in data or algorithmic design — the tool will detect it via smartly formulated closed questions and answers. This immediate feedback helps developers rectify issues before they escalate into bigger problems.

For Mauritania, unchecked biases could lead to wrong decisions, from reinforcing stereotypes to misinformed policies. However, the inherent ambiguity of natural language as a tool is noteworthy and can be a serious limitation. But by acknowledging it and actively seeking solutions, such as further localized training, continuous feedback loops, and domain-specific expert input. Because it is adaptable and resilient, the process of improving it will increase the awareness, understanding and realization of governmental institutions on the importance of ethics.

Interestingly enough, even if the questions or the “guidance” of the LLM end up being strange, trying to understand their meaning will force users on both sides to think about ethics, which would not be so much the case if this tool did not exist.

The source code attached to this work is available on GitHub, here: <https://github.com/Lalla-Aicha/Ethic-App-developer>, <https://github.com/Lalla-Aicha/EthicApp-Government>. A lot of screenshots could not be added to this paper due to the maximum number of pages.

To further enhance the effectiveness and accessibility of our proposed ethical framework for AI, future work will focus on two key areas: improving the model’s response precision and expanding its linguistic reach.

First, fine-tuning the LLM for conciseness and precision will be essential for achieving more targeted, contextually relevant outputs. By training the model on a curated dataset of ethical guidelines and domain-specific language, we aim to refine its ability to deliver concise and precise recommendations. This focused training will help reduce ambiguity and improve the clarity of responses.

Second, to support broader accessibility in diverse linguistic contexts, we plan to implement multilingual capabilities. This could involve training the model on datasets in multiple languages or incorporating advanced translation features that allow ethical standards to be accurately reflected in non-English-speaking regions. By expanding the tool's language support, we aim to provide a more inclusive platform, enabling it to be a valuable resource in global settings where ethical norms and practices vary significantly.

Through these advancements, the tool can become both more precise and more widely accessible, contributing to a more ethically aligned implementation of AI across diverse cultural and linguistic landscapes.

REFERENCES

- [1] H. Mehr, H. Ash, and D. Fellow, "Artificial intelligence for citizen services and government," *Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch.*, no. August, pp. 1–12, 2017.
- [2] J. Berryhill, K. K. Heang, R. Clogher, and K. McBride, "Hello, world: Artificial intelligence and its use in the public sector," 2019.
- [3] S. J. Mikhaylov, M. Esteve, and A. Champion, "Artificial intelligence for the public sector: opportunities and challenges of cross-sector collaboration," *Philosophical transactions of the royal society a: mathematical, physical and engineering sciences*, vol. 376, no. 2128, p. 20170357, 2018.
- [4] V. J. Straub and J. Bright, "Unite the study of ai in government: With a shared language and typology," *AI & SOCIETY*, pp. 1–2, 2024.
- [5] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [6] M. D. M. Reddy, M. S. M. Basha, M. M. C. Hari, and M. N. Penchalaiah, "Dall-e: Creating images from text," *UGC Care Group I Journal*, vol. 8, no. 14, pp. 71–75, 2021.
- [7] A. Névóel, H. Dalianis, S. Velupillai, G. Savova, and P. Zweigenbaum, "Clinical natural language processing in languages other than english: opportunities and challenges," *Journal of biomedical semantics*, vol. 9, no. 1, pp. 1–13, 2018.
- [8] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günemann, E. Hüllermeier *et al.*, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and individual differences*, vol. 103, p. 102274, 2023.
- [9] J. C. Young and M. Shishido, "Investigating openai's chatgpt potentials in generating chatbot's dialogue for english as a foreign language learning," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023.
- [10] C. Hernández-Castillo, H. H. Guedea-Noriega, M. Á. Rodríguez-García, and F. García-Sánchez, "Pest recognition using natural language processing," in *International Conference on Technologies and Innovation*. Springer, 2019, pp. 3–16.
- [11] S. Biswas, "Importance of chat gpt in agriculture: According to chat gpt," Available at SSRN 4405391, 2023.
- [12] C. C. Corrigan, S. A. Asakipaam, J. J. Kponyo, and C. Luetge, *AI Ethics in Higher Education: Insights from Africa and Beyond*. Springer Nature, 2023.
- [13] M. Bensalah, "Toward an ethical code of ai and human rights in morocco," *Arribat-International Journal of Human Rights Published by CNDH Morocco*, vol. 1, no. 2, pp. 187–203, 2021.
- [14] A. Gwagwa, "Recommendations on the inclusion sub-saharan africa in global ai ethics," *RANITP Policy Brief*, vol. 2, 2019.
- [15] A. Gwagwa, E. Kraemer-Mbula, N. Rizk, I. Rutenberg, and J. De Beer, "Artificial intelligence (ai) deployments in africa: benefits, challenges and policy dimensions," *The African Journal of Information and Communication*, vol. 26, pp. 1–28, 2020.
- [16] A. Gwagwa, E. Kazim, and A. Hilliard, "The role of the african value of ubuntu in global ai inclusion discourse: A normative ethics perspective," *Patterns*, vol. 3, no. 4, 2022.
- [17] E. Ntoutsis, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdil, M.-E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis *et al.*, "Bias in data-driven artificial intelligence systems—an introductory survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, p. e1356, 2020.
- [18] E. Van Miltenburg, "Stereotyping and bias in the flickr30k dataset," *Proceedings of multimodal corpora: computer vision and language processing (mmc 2016)*, 2016.
- [19] A. Al-Ghamdi, "A survey on software security testing techniques," *Int J Comput Sci Telecommun*, vol. 4, pp. 14–18, 2013.
- [20] R. R. Althar, D. Samanta, S. Purushotham, S. S. Sengar, and C. Hewage, "Design and development of artificial intelligence knowledge processing system for optimizing security of software system," *SN Computer Science*, vol. 4, no. 4, p. 331, 2023.
- [21] S. Qazi, M. S. Memon, A. Ali, and S. Nizamani, "Role of artificial intelligence (ai) tools for assuring quality in software," *Journal of Southwest Jiaotong University*, vol. 57, no. 2, 2022.
- [22] S. Ramchand, S. Shaikh, and I. Alam, "Role of artificial intelligence in software quality assurance," in *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys) Volume 2*. Springer, 2022, pp. 125–136.
- [23] V. Nallur, "Landscape of machine implemented ethics," *Science and engineering ethics*, vol. 26, pp. 2381–2399, 2020.
- [24] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.
- [25] J. Fjeld, N. Achten, H. Hillgoss, A. Nagy, and M. Srikumar, "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for ai," *Berkman Klein Center Research Publication*, no. 2020-1, 2020.
- [26] V. Tiple, "Recommendations on the european commission's white paper on artificial intelligence—a european approach to excellence and trust, com (2020) 65 final (the 'ai white paper')," 2020.
- [27] M. Carman and B. Rosman, "Applying a principle of explicability to ai research in africa: should we do it?" *Ethics and Information Technology*, vol. 23, no. 2, pp. 107–117, 2021.
- [28] L. Floridi, J. Cowsls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi *et al.*, "Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations," *Minds and machines*, vol. 28, pp. 689–707, 2018.
- [29] H. Miller, R. Stirling, Y. Chung, S. Lokanathan, E. Martinho-Truswell, J. New, I. Rutenberg, and F. Scrollini, "Government artificial intelligence readiness index 2019," *London: Oxford Insights*, p. 32, 2019.
- [30] R. Calo, "Artificial intelligence policy: a primer and roadmap," *UCDL Rev.*, vol. 51, p. 399, 2017.
- [31] P. Rawat and A. N. Mahajan, "Reactjs: A modern web development framework," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 11, pp. 698–702, 2020.
- [32] B. Basumatary and N. Agnihotri, "Benefits and challenges of using nodejs," *International Journal of Innovative Research in Computer Science & Technology*, vol. 10, no. 3, pp. 67–70, 2022.
- [33] N. F. Lindemann, "Chatbots, search engines, and the sealing of knowledges," *AI & SOCIETY*, pp. 1–14, 2024.
- [34] J. Yang, H. Jin, R. Tang, X. Han, Q. Feng, H. Jiang, S. Zhong, B. Yin, and X. Hu, "Harnessing the power of llms in practice: A survey on chatgpt and beyond," *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 6, pp. 1–32, 2024.

- [35] J. Huang, D. M. Yang, R. Rong, K. Nezafati, C. Treager, Z. Chi, S. Wang, X. Cheng, Y. Guo, L. J. Klesse *et al.*, "A critical assessment of using chatgpt for extracting structured data from clinical notes," *npj Digital Medicine*, vol. 7, no. 1, p. 106, 2024.
- [36] R. Dulepet, "Generating high-quality "about me" narratives," 2023.