DS Journal of Artificial Intelligence and Robotics
Volume 3 Issue 4, 44-64, Oct - Dec 2025
ISSN: 2583-9926 / https://doi.org/10.59232/AIR-V3I4P101

Review Article

Ethical AI and Privacy: Strategies for Effective Governance

Taban Habibu[1,2*], Obudra Ceasar[1*]

[1]*Computer and Information Science, Faculty of Technoscience, Muni University, P.O. Box 725, Arua, Uganda.*
[2]*Department of Computer Science, Faculty of Science, Islamic University in Uganda, P.O. Box 7244, Kampala, Uganda.*

[1*]hamitech2019@gmail.com

Received: 15 October 2025; Revised: 26 November 2025; Accepted: 17 December 2025; Published: 30 December 2025

Abstract - This study systematically maps how ethical principles of Artificial Intelligence (AI) and privacy-preserving techniques are embedded in governance frameworks at global, continental, and national levels. It offers a multi-layered comparison of normative, policy, and technical instruments, showing how ethical AI ideals are translated or fail to translate into operational AI governance practices. Using a mixed-methods empirical design, the study analyses 132 documents following a structured coding framework to examine ethical principles, enforceability, and the operationalization of privacy-preserving techniques. The findings show that most frameworks emphasize privacy, transparency, and accountability, yet only a fraction convert these commitments into verifiable technical or regulatory requirements. High ethical values are widespread, but the operational depth of tools such as audits, logging systems, data-protection assessments, and advanced privacy-preserving techniques remains uneven, particularly in resource-constrained settings. Binding instruments demonstrate stronger implementation pathways than advisory values. The study demonstrates persistent gaps between principles and practices and underscores the need for governance models that integrate enforceable controls with locally grounded ethical considerations. It concludes that trustworthy AI in diverse contexts requires verifiable governance through risk-tiered audits, mandatory integration of privacy-enhancing techniques, and community-aligned ethical frameworks.

Keywords - AI ethics, AI governance, Artificial Intelligence, Privacy-preserving techniques, Data privacy.

1. Introduction

Artificial Intelligence (AI) refers to the tangible, real-world capability of non-human machines or artificial entities to perform tasks, solve problems, communicate, interact, and act logically, as with biological humans [1]. Ethical AI is concerned with ensuring that the design, deployment, and governance of AI systems conform to moral principles that respect human rights, protect individuals and communities, and align with societal values and the rule of law. In this context, Privacy refers to an individual's right to control and handle their data throughout the AI life cycle [2-5]. AI governance refers to the structured set of frameworks, policies, operational procedures, and technological instruments that direct the design and deployment of artificial intelligence systems in alignment with organizational principles, strategies, objectives, and overarching organizational missions. A multi-stakeholder approach that elaborates on elements to be governed, when and how (frameworks, tools, policies, or models), at the team-level, organizational-level, industry-level, national, and international levels can ensure trustworthy AI [6]. AI integration to improve economic, social, and governmental systems is on the rise, as it enhances prediction, automation, decision-making, and natural language processing across domains such as finance, security, education, transport, health, and public administration. It has been estimated that AI will contribute up to USD 15.7 trillion to the Gross Domestic Product (GDP) of the world by 2030 [7]. This rapid scaling is data-intensive, which increases privacy-related risks, including re-identification, discrimination, intrusive profiling, unexplained decision-making, and the misuse of sensitive information [2, 3, 8]. This raises very critical ethical questions, like whether the ethical AI principles that exist in frameworks are being considered during the AI governance processes. Do the frameworks have adequate details for operationalization for ethical AI? Empirical and systematic studies have paid little attention to answering these questions, although there is a great deal of literature on the effectiveness of the isolated privacy-preserving techniques, with less focus on their inclusion in frameworks and whether the details in these frameworks are adequate for users to easily understand and integrate into the AI lifecycle.

Several AI governance frameworks from Intergovernmental, governmental, and corporate entities have been issued. Most of these are Strategies and policies that are just advisory, as very few are legally binding. They provide information on AI ethical principles such as data protection, accountability, privacy, non-discrimination, explainability, human oversight, transparency, safety, and fairness as well. Literature shows that a range of advanced privacy-preserving techniques such as federated learning, differential privacy, secure computation among have emerged but limited literature have assessed the extent to which these have been included int the various frameworks [9-18]. The existing literature shows an increase in these frameworks, but substantial uncertainty remains about the extent to which these commitments are incorporated into practice for AI governance mechanisms and technical practices, and limited literature has been dedicated to this topic. Also, most of these literatures are not tailored to link how resource-constrained entities are ensuring ethical AI practices.

Particularly, studies that use systematic methods to compare how ethical AI and privacy are embedded across global, regional, and national frameworks, how often privacy-preserving techniques are referenced or required, and whether gaps exist between ambitious strategies and practical application, especially in low- and middle-income states, are limited [9, 15]. This review addresses these gaps by systematically mapping how ethical AI principles and privacy-preserving techniques are embedded in AI governance frameworks at global, continental (African), and national levels, through an analysis of a corpus of normative, policy, and technical documents and the isolation of challenges that hinder the practice of these frameworks. It has consequently offered comparative insight into the differences between ethical AI principles and privacy-preserving techniques in application by technical and regulatory implementers, and those that remain aspirational in resource-constrained contexts, and offers recommendations to align for improved ethical AI governance.

2. Literature Review

This section synthesizes prior work in the following areas: (i) Ethical AI frameworks and principles, (ii) Global and regional AI governance frameworks, (iii) Technical privacy-preserving techniques, (iv) AI governance challenges and implementation gaps, (v) Comparative governance landscape, (vi) Theoretical and conceptual foundations, and (vii) Gaps motivating this review.

2.1. AI Ethical Governance Frameworks at the Corporation Level

Several ethical AI concerns exist, including bias, fairness, discrimination, privacy, and data protection. Bias may originate from data being non-representative. This, in most cases, is unintended but emanates from the methods of data collection, social or institutional, or system structures. For example, data collection may favor a particular group, resulting in biased data, which also makes the model produce unfair results, leading to inequality and discrimination, among others, that may increase the disparity. Typically, if health data is collected from a hospital where high-income earners are the sole clients, their data is different from that of low-income earners [13]. Fairness, another ethical AI concern, emphasizes that AI must be beneficial to all, from developers to final users. This is aimed at reducing the inequalities that exist. So, the AI systems developed should be able to detect disparities across race, socioeconomic status, geography, and gender. Whoever takes responsibility for the decision, prediction, etc., made by an AI model is an ethical issue. The developer and the user must take responsibility for both the good and negative impacts of the use of AI. Since AI lacks human emotions, ethically, the user may be alienated and impersonalized. So, users need not replace human service providers, for example, in healthcare, empathy can be provided by humans. Explainability is a very important aspect of AI ethics. The developers and users of AI must understand why the AI has made such a decision or prediction. In medical AI ethics, it is a must for clinicians to be able to explain to patients why such a diagnosis and treatment plan is so, to have shared decision-making. In case harm is caused because of a decision or prediction made by AI, who is liable? In AI ethics, transparency focuses on the users of AI technology understanding why the AI model made such a decision. This is aimed at making the user trust so that it is not misused. To this effect, techniques such as the Shapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanation (LIME), Class Activation Maps (CAM), and Attention-based explanations, among others, are now employed to understand the decisions of AI models.

AI ethics also looks at the privacy of the data used for training and the data collected by the AI model as it is operating. Such sensitive data ought to be collected with a clear consent process, deidentification of the data, and compliance with the legal standards. When collecting data for AI model training or using AI for solving problems of people, they should be fully informed about it and the risks, especially in medicine. So, the AI technology being applied should adhere to this ethical consideration of consent. Liability is also one of the ethical issues. If there are errors caused by AI, is it the AI developer, the user, or the institution that is applying AI? It is critical to ensure that these issues are handled [13]. An example of such an occurrence is the scamming of $243 000 by criminals using a cloned voice of the CEO of an energy company [19]. In this scenario, the ethical question to be answered is, who is to take responsibility or provide answers? Is it the AI systems developer, deployer, operator, user, or government? Who will carry out reparation for the losses incurred? The answers to these questions are multifaceted; ethics is core, but legal considerations are too [20]. Lastly but not least, accountability in AI ethics involves understanding the limitations of AI so that users can apply their expertise, like determining the errors and biases, to make sure that the AI is accurate. The performance of the AI must be monitored over time to ensure adaptation to the emerging AI ethics [13].

Due to the profound negative impacts of these eventualities, which make AI unethical, several big corporations have stepped up intentional acts to combat them. This has shown several frameworks being developed and implemented right from the development phase, via deployment to post-deployment. The core goals emphasize AI that is fair, non-discriminative, and safe to use, with highly reduced biases against minorities. Strategies to ensure that the decisions are within acceptable standards and accountability mechanisms are also set in case they go beyond the goodness of intentions. Some of these companies went up to ensure that sustainability, public awareness, and multi-stakeholder engagement strategies are also well catered for in these frameworks [21, 22]. For example, the Ethical Application of AI (EAAI) framework developed by the American Council for Technology-Industry Advisory Council (ACT-IAC) advises AI system evaluation for ethical compliance based on the principles of bias, fairness, transparency, responsibility, and interpretability.

This guidance, though non-legislative, provides a good ground for building, deploying, and monitoring AI systems for trustworthiness [23]. Amongst the corporates, Facebook has developed the Fairness Flow Toolkit. Facebook understands whether their models are not performing within the set ethical limits by comparing the parameters with how far they differ from the training data and using statistics as well. Correction measures are then employed accordingly in case of variations. So, fairness is ensured right from development up to after deployment [22]. Google, one of the key players in the AI industry, is also performing its role of ensuring ethical AI practices using its model cards framework, which provides detailed information on the AI models, ranging from the architecture, via training data, performance metrics, biases, and responsible AI techniques used. This makes it very easy to evaluate compliance with ethical and responsible AI practices, very simple, practical, and transparent. This has been adopted by other players like HuggingFace, International Business Machines Corporation (IBM), Kaggle, and AWS SageMaker. Because of the transparent data available, engaging state bodies for compliance is very easy when using the model cards [24]. On top of the model cards, IBM has moved further to develop the AI

Factsheets. Though like model cards, it has set out a set of questions in line with ethical AI. By answering all the questions, the AI system will be aligned with the purpose of capturing ethical practices in AI. Other areas captured in the Factsheets include the domain of application, how it is used, and what safety requirements are in place, among others. Google's AI development is guided by their threefold AI Principles of Bold innovation, Responsible development and deployment, and Collaborative progress, together [25].

As they intend to advance AI development, the main driving motivations is to improve the standards of living though innovations that addresses the realistic societal challenges as negative impacts are significant kept lower through putting human in the loop direct the design, development and deployment reevaluations such that they adhere to social accountability, the law at all levels, security for privacy and other aspects of human life, and reduce chances of unforeseen risks. Aware that these require efforts from various stakeholders in academia, industry, governments, and even civil society to make AI more ethical, all parties will be engaged. This framework recognizes not only that AI governance is bound to change over time, but it has also gone ahead to place emphasis on its adaptability while ensuring responsible AI that promotes society without hampering innovations with intentional investment into research for AI safety and security right from design to use.

Microsoft Corporation also has a Frontier Governance Framework, which directs the management of risks based on evidence. This is aimed at high risks that may not only affect the national interest but also the general public at large. It seeks a secure and trustworthy AI life cycle for all models and systems of AI developed and deployed. It has been set to be revised often to cater to the rapid advancement of AI. It is intended to be used in an integrated manner, both within small and large high-risk AI models and by outside developers and deployers. It details the evaluation and assessment as evidence for action to mitigate the impacts that can prove negative. The categorization of risks is low, medium, high, and critical. Low and medium-risk AI models are allowed to be deployed if they satisfy the requirements of responsible AI, but those categorized as high and critical-risk require review. Under this frontier governance, the domains for focus are Chemical, Biological, Radiological, and Nuclear (CBRN) weapons, Offensive cyberoperations, and Advanced autonomy. All these areas are critical as ethical breaches in them can cause very serious challenges to humans and their environment [26].

As AI governance is multi-stakeholder in nature due to its complexity and ever-changing, advancing nature, partnerships are required to learn from each other and, in a flexible manner, quickly adapt to these changes. Mitigating the risks for the frontier models requires that access to model weights be restricted from unauthorized insiders and outsiders. This is bolstered by encrypting the model’s weights and the provision of many security layers, plus engaging third parties to further safeguard the model’s weights. This frontier governance framework, as compared with others, is focused on very critical risks, both current and emerging. Threats that can affect the whole of humanity on a large scale. It cuts across health, environment, security, economy, and politics. It showcases Microsoft’s proactive approach to ensuring responsible and trustworthy AI for all. The rest of the frameworks are general in that their focus is not limited to the frontier models. But Microsoft is well-rounded in that its responsible AI program also includes a set of Responsible AI standards and Principles that guide its operations, even for non-frontier models that are similar to those of other corporations.

Intel Corporation has also set up its Responsible AI (RAI) principles, which are guided by the respect for human rights at every stage of the AI cycle. A similar approach extends to the principle of allowing humans to play an oversight role in the model design, training, and output evaluation, and device corrective measures accordingly in case of deviations. The frameworks advise providing relevant information about the model, especially by developers, such as what the model can best do, its possible risks, rigorous model details pertaining to training and testing, descriptions of the characteristics of the training and testing sets, and the possible biases that exist. Not only this, but Intel also commits to security and safety through the design principles of its technology and development, ensuring safe, secure, and reliable AI. Since AI relies on huge volumes of data, they acknowledge this and commit to practices that safeguard privacy throughout the lifecycle. Further commitments are made to include all stakeholders to benefit and ensure that the environment is protected. Intel's RAI is quite unique in explicitly outlining its commitment to drive AI solutions, both hardware and software, that will reduce carbon and waste footprints, which universally benefit the globe [27]. The intentions of the companies are similar, but their approaches differ greatly. Although the reviewed frameworks for ethical AI governance from corporations are not exhaustive, it can be noted that these are big corporations that are not so much constrained resource-wise.

The above frameworks, once used diligently, will significantly improve how AI applications comply with ethical and governance principles. But Concept drift can affect ethical AI. For example, economic changes may affect health-seeking habits, with people typically going to pay for health services rather than free ones, when they become wealthier. Even the patterns in diseases shift, for example, undernourished kids are found among the poor as over-nourished kids among the wealthy, and so, a model trained on data when such people were poorer may become biased. This demands AI models that are adaptive as well as the methods of their development. Among other areas for better ethical AI and AI governance are the security of models from, for example, adversarial attacks. It also provides clear directions on what is needed for audits [28].

Virginia Dignum also emphasized that even if AI development and deployment comply with ethical AI standards, users must also fall in line, as the use impacts the behaviour of individuals and society is at the core. This is particularly important in maintaining the privacy of the users. Not only this, but in performance, testing for compliance is recommended to be done by both the developer and a third party. This ensures transparency and safety [21, 29]. No one ethical AI framework is sufficient to address the ethical demands universally. This is because, despite ethical principles being universal to a greater extent, they are highly influenced by cultural and social interactions that differ from one community to another.

It is imperative, therefore, that developers must be conscious of the specific ethical demands of where such AI is applied and modify its parameters accordingly. For users, a similar approach must be used to adhere to the specific societal ethical demands. It is through such intentional applications of the frameworks that AI will be safe and trustworthy for everyone. Despite the presence of these frameworks, ethical breaches still occur in these corporations, ranging from discriminatory advertisement algorithms in the areas of gender and race, among others [30, 31]. The use of these frameworks is limited by some factors that need to be identified, and it seems like such breaches may be even more pronounced in companies that do not have such regulations, which need to be investigated.

2.2. Global and Regional AI Governance Framework

AI governance needs strong AI ethical frameworks Greytok et al., [29] for examples recommends the following for the achievement of responsible AI; fairness which comes if the training data is unbiased, transparency in terms of the data collected and their use, misuse of AI systems to generating either abusive, misleading or even toxic content has to be mitigated through architectural design and human oversight; aligning AI systems operations to human values and ethical principles would ensure better governance and finally, strict adherence to data laws, especially regarding its usage and sharing with third parties would reduce privacy and data loss.

In 2019, Jobin, Ienca, and Vayena [32] mapped documents that showcased how institutions respond to AI governance. Following this were several instruments that have either normative or regulatory force. Notably: (a) The UNESCO Recommendations on the Ethics of AI (2021), which marked the first normative instrument adopted by up to 193 Member States. Transparency, accountability, fairness, human rights, and privacy are the key ethical AI concerns addressed [33], (b) The European Union Artificial Intelligence Act (2024), which is a wide-ranging regulation focusing on risks. It squarely grouped the use of AI based on the level of risk and went ahead to provide obligations for high-risk systems, plus how to enforce it [10], (c) The OECD AI Principles (2019, updated 2024) were also formulated with five-fold principles, which are high-level based on values with also five recommendations, all aimed at the promotion of trustworthy AI.

These principles are meant to be adopted for national policies and international co-operation. These principles are, (i) Inclusive growth, sustainable development, and wellbeing, (ii) Respect for the rule of law, human rights, and democratic values including fairness and privacy, (iii) Transparency and explainability, (iv) Robustness, security, and safety, and (v) Accountability. The National policies and International Co-operation for Trustworthy AI are, (a) Investing in AI research and development, (b) Fostering an inclusive AI-enabling eco-system, (c) Shaping an enabling interoperable governance and policy environment for AI, (d) Building human capacity and preparing for labour market transformation, and (e) International co-operation for trustworthy AI [11]. and (d) NIST AI Risk Management Framework (AI RMF, 2023) was developed to foster trustworthy AI.

They identified the following challenges in the bid to manage risks within the AI lifecycle: (i) Risk measurement – pain points have been identified which include (i) AI risks stemming from third-party data, hardware or software which can be the difference in risk metrics, methodologies used in the AI development, which can be complicated by lack of transparency in disclosure by the developer, and further by users integrate third-party data or system, (ii) Tracking emergent risks by organizations and then devise measurement techniques for them in the context of their occurrence, (iii) Availability of reliable metrics is still a challenge as there is no AI risk, trustworthiness, and applicability measurement methods that are robust and verifiable by consensus, (iv). Risk at different stages of the AI lifecycle. ii) Risk tolerance, iii) Risk prioritization, and iv) Organizational integration and management of risk [9].

2.3. Privacy Enhancing Techniques (PETs)

The privacy of data concerns emanate from the data collected for model training, and more risk comes from Internet of Things (IoT) devices that are powered by AI that collect data, such as industrial sensors, smart home devices, and wearable health devices, among others. The threats to privacy from IoTs are largely due to devices lacking robust security measures, so anyone can access such data and misuse it for identity theft, for example [34]. Solutions to these challenges have been sought through both algorithmic and systems-level privacy-preserving techniques evident in technical literature. These include Differential Privacy (DP), where calibrated noise is a formal way to handle worst-case privacy scenarios. Federated Learning (FL) - it involves keeping the data local, but only the model's updates are combined. FL has been widely used despite its limitations, as its benefits are significant. Secure Multi-party Computation and Homomorphic Encryption, which involves the use of cryptographic protocols that facilitate hiding the raw data, although computations are done jointly.

This technique is hailed but avoided due to its resource intensity, especially in large-scale models. Other technical writings suggest the utilization of advantages conferred by all the advanced PETs in combination with maintaining good control over how the AI systems are accessed to reduce risks. When audit logs and consent mechanisms are also used to track ethical practices alongside this combination, improved privacy preservation is assured [9, 35, 36, 37]. Since these PETs are becoming more achievable because their algorithms are becoming easily accessible, privacy is becoming assured. However, there is limited literature pointing to exactly what retards the technical uptake of these PETs, although a few are pointing to limited computational resources [38].

2.4. AI Governance Challenges and Implementation Gaps

Several ethical AI frameworks have been authored by governments, NGOs, and corporations. Of these, some are legal in nature, while others just provide advice on how to make AIs more ethical. As some companies and other AI players make use of these resources, most do not ensure transparency, fairness, privacy, and accountability, among other principles in the life cycles of AI. Limited financial, technical, and computational resources, plus legal weaknesses, are among the primary saboteurs of implementation. Legal implementations to check whether AI players are complying with the set legal guidelines, or for legal bodies to play oversight roles, as well as the development of appropriate regulations and the delivery of sanctions, are all financially straining, as is the training of expertise to develop and implement PETs and other AI ethical strategies. This challenge has been observed across the globe [39, 40]. The PETs in particular, Differential Privacy (DP), Federated Learning (FL), secure multi-party computation, and the homomorphic encryption are computationally and expert-wise expensive to implement, but very effective though [12-14, 40].

Some gaps also exist amongst the frameworks, as the metrics used to ascertain the levels of AI ethical risks vary. This originates from the relatively different focus of each framework and the various principles of focus. For example, in measuring fairness and bias, metrices from statistics such as Demographic Parity Difference, and Ratio, Disparate Impact Ratio, Statistical Parity, and Grouped Fairness Index are referenced in the NIST AI Risk Management Framework (NIST AI RMF) amongst those listed under Error-based fairness, Individual Fairness, and Proxy Bias [39, 40]. All these gaps highlight the need for an AI governance framework that elaborates the need for a merger in ethical AI implementation within resource-constrained settings and adaptability to specific demands of a context, such that the clearly elaborated principles for practice of ethical AI [40, 41] governance are achievable to a greater extent.

2.5. Comparative AI Governance Landscape

AI governance Frameworks developed at the Intergovernmental level are mainly non-binding, and they provide guidelines and recommendations to ensure the AI lifecycle is guided by ethical norms. In this list are the OECD AI Risk and Trustworthy Framework, which is non-binding but has set out clear principles for ethical AI like fairness, accountability, transparency, and robustness [12], and the UNESCO Recommendations on Ethics of AI, which are fronted by UNESCO, emphasize the need to always observe human rights. As AI is beneficial, this ought to be distributed to all, fairly without discriminating against other humans based on any characteristic, such as race, gender, among others, and most importantly, AI development must be developed following the principles of sustainability, its benefits must not outpace its negative impacts, and this must continue for a considerable time. UNESCO, being a global organisation, expects that its recommendations will be adopted at a higher rate [13, 14]. The Framework Convention on Artificial Intelligence is a convention that binds the member states under the Council of Europe. It aims to make use of AI to make the rights of humans not to be abused, as citizens remain or become more law-abiding, and democratic practices have increased.

This has, like the others, laid down principles for AI governance which include human dignity and individual autonomy, transparency and oversight, accountability and responsibility, equality and non-discrimination, privacy and personal data protection, reliability and safe innovation. This safe innovation is aimed at reducing the negative impacts of AI, already known and unknown, by state members engaging experts throughout the AI lifecycle while advancing AI technology to improve the standard of human life. This is not just a recommendation, it is binding amongst the signatory states [42]. The G7 member countries have adopted the OECD AI principles under the G7 AI Guiding Principles.

In a Statement on by the Center for AI and Digital Policy for the G7 on AI Policy for Democratic Nations in 2025, their member states that also belong to the European Union (EU) to implement the prohibitions made by EU to ensure responsible and ethical AI such as AI systems that are scheming and unfair, those that uses characters of individuals to categorize them, predict their chances of engagement in criminal activities. Not only this, but also databases that derive their facial data from sources other than specific aims from the web or CCTV footage. Others like emotion recognition, except if it is to facilitate the health or safety of the individuals, use of biometric data for discrimination purposes, such as race and similar one, is the use of biometric systems to identify individuals in real-time to be used either by police or any law enforcement bodies. These, if adopted, will improve on the ethics of using AI; however, due to the nature of this framework being just a policy recommendation, its uptake is largely dependent on the willingness and readiness indices of the member states [43]. Global Digital Compact by the United Nations (UN) has also committed to the reduction of AI harms by committing to strengthening data protection and capacity building for member states. Inclusive AI risk management and engaging all the stakeholders. A lot of other commitments have been made to ensure that AI is ethical [44]. These major Intergovernmental frameworks look forward to an AI ecosystem that adheres to good governance mechanisms that make AI more ethical. Although the challenges of ethical breaches still occur.

At regional and governmental levels, several frameworks exist. In the EU, the EU AI Act and EU Digital Services and AI-related standards are in existence. The latter is used in the governance of AI-based services, ranging from the moderation of AI content on social platforms to other applications, including recommendation algorithm systems. The EU AI Act is binding, and it categorizes AI systems by levels of risk, being low risk, medium, and up to high risks, and provides for the detailed requirements for the levels of risk that are acceptable for models to operate. The United States has fronted NIST AI RMF and Federal Executive Orders on AI, for example, 14179, which has set a precedent for fostering AI development by reducing policies that overly restrict, but it also provides for security, which implies that issues like ethical security are considered.

This order is binding. The NIST RMF is recommending it based on risk levels and is detailed enough to allow for adoption [3, 45]. In the US, several nations have developed their own AI frameworks, such as the Colorado AI Law, which is legally binding, plus other regulations that focus on sectors such as those aimed at maintaining human rights. From the Asian world, the China Algorithmic Recommendations Regulation regulates how the recommendation systems must be developed and applied to ensure that they remain transparent with the data they collect, fair, and free from data and privacy breaches. Fake news generation is forbidden as an example of maintaining social values, economic, and national security.

This legally binding regulation requires AI service providers to provide mechanisms to determine if content is legal and positive; otherwise, it must not be published or recommended for public use. Several other provisions of this document ensure that data and privacy are protected for the good of individuals, the general public, and the country as a whole [46]. The People's Republic of China also has two other laws for specific sections of AI. These are the China Generative AI Services Law and the China Deep Synthesis Law. These legally binding instruments all have sections that direct the ethical AI lifecycle [47, 48].

Canada, as a country, has the legally binding Artificial Intelligence and Data Act with major expectations of making the AI lifecycle safe for its citizens without discriminating on any grounds, and in case of breaches, such perpetrators are held accountable. Among other areas of focus is the requirement to provide risk management for harm against vulnerable individuals. The harms are identified to be physical, psychological, or economic in nature. Harms that affect vulnerable groups of people are also prohibited, including biases. Human oversight role in the AI lifecycle is a requirement aimed at a transparent, fair, non-discriminatory, safe, and accountable AI whose output is valid over a range of circumstances in which the models operate. It has also set guidelines to ensure that AI players, such as developers and users, can benchmark and implement mitigation measures against ethically and legally unacceptable impacts, especially from AI models that greatly affect society [49].

The African Union AI Strategy is at the African continental level. It details the peculiar ethical demands of the African social context, emphasizing that every member of the society must be included in the AI-powered economy, and equal benefit should be ensured, plus demand for sovereignty of data at the national level. The data used by AI must be governed by the country's laws and governance systems, either at collection, processing, or even storage. In case of cross-border transfer, the appropriate international laws and ethical guidelines need to be respected so that privacy of the data sources is protected [40, 41]. Nationally in Africa, AI strategies that also specify the ethics of AI have been published by only a few. Mauritius [50], Rwanda [51], the United Republic of Tanzania

[52], the Islamic Republic of Mauritania [53], Egypt [54], Kenya [55], and Tunisia [56] are among the African countries that have dedicated National AI Strategies.

Some other countries have AI guidelines for their governance embedded in initiatives that are broader. Amongst these are countries like South Africa [57, 58] whose AI governance and development guidelines are part of its National Digital and Future Skills Strategy 2020. The one for Ghana is part of the National Digital Transformation Strategy [59], and Ethiopia has the Digital Ethiopia 2025. But they have the AI Strategy framework, but policies are also not yet in place [60]. Uganda also has the National 4IR Strategy, which guides the direction of AI development and is strengthened by the National AI Initiative [61].

A number of frameworks have also been developed as standards at the technical level. They are guided by the principles, and they are used to certify AI systems. Amongst those developed under the International Organisation for Standardization (ISO) are ISO/IEC 42002, which provides guidance on the management of AI systems. The ISO/IEC 23894, which is used to manage the risks in AI, plus others that guide IT, of which AI is a part. AI Risk Standards are good examples for further contextualized and specialized frameworks [62, 63].

The above discussion shows that several AI governance frameworks have been published. They are either legally binding or soft laws that provide guidelines for the responsible AI lifecycle. Not only these, but corporations have also fronted many, as seen in the previous sections. All these show a trend that the high-income countries have set up more and better governance frameworks, as the EU, China, and the US have legally binding laws that do not exist in the low-income countries. For example, there is not even one African country that has a law that specifically directs how AI should be governed, with a few that have set out AI strategies that guide how AI should be developed.

Like the African countries, the countries that do not have laws use existing data and privacy protection laws, which are relatively insufficient for the peculiar risks that come with AI. However, the high-level intergovernmental frameworks have existed for some time, but practical uptake seems limited by certain factors. Even amongst the countries that have legal frameworks, breaches still exist, with only a few players in the AI lifecycle using such frameworks. Amongst the corporations, only the big ones have adopted and adapted these frameworks with details that make these frameworks easier to practice [2, 4, 8, 64].

2.6. Theoretical and Conceptual Foundation

Ethical AI and privacy are grounded in the practice of ethics, governance, and sociotechnical systems. Ethical AI frameworks are based on the perspectives of human rights, deontology, and consequentialism, as well as values such as beneficence, justice, non-maleficence, and autonomy [37, 38, 65]. The ethical AI principles in the various instruments, such as non-discrimination, transparency, privacy, fairness, accountability, respect for human dignity, and explainability, are all informed by these traditions.

AI governance is anchored on the main governance that considers the roles played by institutions, regulation, multi-stakeholder, and risk-based oversight as AI is both a technical artefact that is hinged onto a socio-institutional complex. Frameworks such as the NIST AI RMF look at AI governance as a continuous process. With a focus on risks, it involves risk identification, measurement, mitigation, and monitoring [38]. At the regulatory level, the EU AI Act implements a risk-tiered model. The more risk, the more the obligation, which is precautionary governance [3].

The academia front that responsible and trustworthy AI incorporates not only technical but also social norms, legal, and organizational practices [9, 39]. This implies that ethical AI and privacy-based AI governance require synergistic collaboration amongst the technical, business, legal, and social entities. At national levels, the adoption of the global ethical AI principles requires critical interpretation and then adapting to the realities of the local institutions, infrastructure, and social value contexts, especially in Africa, including Uganda.

2.7. Gaps Motivating this Review

Despite existence of large volume of literature that analyzes the ethical AI principles and frameworks at intergovernmental, governmental and corporate levels for AI governance, they remain focused at the specific levels and are normally conducted with less focus on their development and use in resources constrained settings, so substantial gaps exist in systematic literature that analyze the extent to which the high-level frameworks are incorporated into the national and corporate AI governance frameworks for legal and advisory purposes and what limits their uptake, in the low- and middle-income countries. Further gaps exist in the literature that link the challenges of continued privacy breaches in AI despite a lot of literature showing the presence of highly effective PETs and governance frameworks [9, 35, 36, 55, 66].

This review addresses these gaps by systematically analyzing and synthesizing how Intergovernmental AI governance frameworks are adopted and or adapted for state level AI governance frameworks while comparatively mapping the factors that hampers this process, with particular focus on low- and middle-income countries so as to front an actionable framework that is more context-sensitive and addresses the specific realities of institutions, regulation and infrastructural needs, particularly for low- and middle-income states such as those in Africa that have do not align well with the global and high-level frameworks that are tailored to high-income, strong regulatory and technical capacity settings [2, 4, 8]. This review further assesses the extent to which privacy-preserving techniques and related technical controls are incorporated into these frameworks at national and corporate levels.

3. Materials and Methods

This section provides the details of the procedures used during the survey.

3.1. Research Design

This study adopts a mixed-methods empirical design that combines qualitative and quantitative approaches. First, a systematic content analysis of governance instruments and policy documents was done to identify how ethical AI principles and privacy safeguards are expressed; Secondly, a quantitative audit of coverage and enforceability indicators from public repositories, especially OECD.AI, UNESCO, AU, and AlgorithmWatch documents. Thirdly, expert-informed mapping of privacy-preserving techniques from the technical and academic corpus to policy recommendations was done.

The empirical strategy is selected for its replicability, and the main materials are the public normative documents and policy inventories. The mixed methods design ensured that qualitative insights were captured, as well as validating them quantitatively. The ethical AI content and AI governance frameworks were captured, as well as their practical application and the adequacy of technical guidance.

Conceptual and normative patterns were extracted using systematic content analysis, while the implementation, compliance, and technical specificity were done through quantitative audits, making it more evidence-based. AI ethics and governance are multifaceted in that they are related to policy, law, technology, and even culture. So, using both qualitative and quantitative methods captured both the normative governance frameworks language and the operationalization. Furthermore, expert validation makes the findings validated based on the practitioner's perspective of AI ethics and governance.

3.2. Document Identification, Search Strategy, Collections, and Dataset Construction

Targeted searches were done from websites, policy inventories, and academic databases using internal search functions, and their browsed sections were labelled. This started by searching the official websites of international and regional organizations, including the European Union, UNESCO, NIST, the African Union, ministries, and authorities concerned with national data protection. Policy inventories, especially the OECD.AI and AlgorithmWatch AI Ethics Guidelines Global Inventory, were also searched to identify more AI governance instruments. The labels are "data protection", "privacy enhancing techniques", "ethical principles", "AI governance", "Global", and "Regional".

Academic and technical documents were retrieved from official and major publishers and databases such as SpringerLink, IEEE Xplore, Taylor and Francis, DOAJ, Elsevier, ACM Digital Library, MDPI, Scopus, and Web of Science. The following keyword combinations were used: "ethical AI", "AI governance", "privacy-preserving", "privacy-enhancing technology", "data protection", "privacy protection", "differential privacy", "federated learning", "AI privacy". Refining the searches was improved by the use of Boolean operators such as "ethical AI" AND "governance", "AI" AND "privacy", "ethical AI" AND "privacy", among others.

The following is how the samples for the study were constructed: The documents that have legal or regulatory force, high normative influence, public availability, and are relevant to privacy and or AI ethics and or AI governance were selected. A total of 190 documents were studied, out of which 132 met the inclusion criteria. Of these, 37 were global normative and regional regulatory instruments. For example, EU AI Act (2024), OECD AI Principles (2019, updated 2024), The UNESCO Recommendation on the Ethics of Artificial Intelligence (adopted 2021), NIST AI RMF (2023/2024 versions), EU General Data Protection Regulation (EU GDPR), African Union Continental AI Strategy (African Union, 2024a), and IEEE Guidance/ IEEE standards/Ethically Aligned Design Output among others. Some of the information was collected from policy inventories, meta-analyses such as the Algorithm Watch inventory [15], a review of about 200 documents, the OECD.AI national policy dashboard, United Nations reports, and Reuters coverage of UN advisory recommendations on AI governance and ethics.

There were 18 national and non-national AI ethics and governance frameworks that have been purposively selected to determine if their principles and application converge with the global and regional frameworks. National policies and Strategies, especially from selected African countries, were 26. These also comprise Data and Privacy protection policies, especially for countries that do not have policies specifically developed for AI regulation. In this, the following documents are selected: National AI Strategies for Kenya (Kenyan Ministry of Information, 2025), South African Protection of Personal Information Act (POPIA), (2013) (Government of South Africa, 2013, 2013), Nigeria (Federal Ministry of Communication, Innovation and Digital Economy, 2025), Rwandan National AI Policy (Ministry of ICT and Innovation, 2022), and The Egypt National Artificial Intelligence Strategy (2025-2030) Second Edition (Egypt National Artificial Intelligence Strategy (2025-2030) Second Edition, 2025), and National Data Protection Laws as Uganda Data Protection and Privacy Act, Kenya Data Protection Act, and Egyptian Data Protection Law among others.

Academic and technical research papers on Ethical AI principles and Privacy Preserving Techniques with a focus on governance were collected from official websites of publishers and websites of academic journals, and they make up the balance of 51 papers. These include Springer, Association for Computing Machinery (ACM), Elsevier, Directory of Open Access Journals (DOAJ), MDPI, IEEE Xplore, Taylor & Francis, Science Publishing, Web of Science (WoS), and Scopus. The AI governance spectrum is considered in the study, including binding legal instruments, soft law recommendations, and policy guidance. The metadata, i.e., year of publication, issuing authority, jurisdictional scope, and thematic relevance, were logged for replicability. Since academic peer-reviewed articles are included on top of the policy documents, alignment and differences of opinion between academia and regulatory intentions were intended to be detected.

3.3. Analytical Approach and Coding Framework

Systematic content analysis was used to code the documents on a standardized coding sheet based on the following dimensions: Presence of principles, i.e., documents with any or all the ethical principles of explainability or interpretability, transparency, fairness or non-discrimination, data protection or privacy, robustness or safety, liability or accountability, and human oversight will be considered. Enforceability of such policies was considered in that documents having regulatory/binding or not, the ones with penalties and supervisory authority were selected. Documents that explicitly elaborated technical controls like differential privacy, auditing, federated learning, and data minimization were selected.

Papers with operational guidance presence or absence were also to be considered. This guidance includes audit requirements, procurement controls, documentation like model cards or data sheets, and impact assessments of the AI tools. Some papers were also included because of their geographical and socioeconomic focus; for example, policy documents at the global level, as well as high-income and low- and middle-income countries, were selected, especially in Africa, and analyzed. The analytical process involved open coding to identify the ethical AI principles, privacy preservation techniques, and governance attributes. This was followed by summarizing the code frequencies and then analyzing them.

Table 1. Inclusion and exclusion criteria for the corpus

Inclusion Criteria: Documents will be included if they:	**Exclusion Criteria: Documents will be excluded if they:**
- contain information on ethical AI, AI privacy, and AI Governance mechanisms. - Have scope within global, continental, and national. - Have been published or updated officially between 2018 and 2025. - Are publicly available and can be traced to the recognized issuing body, like the government, intergovernmental organisation, and reputable academic publishers. - Have direct influence on policy and practice within the AI governance ecosystem. - Websites of reputable AI companies. - Are in the English language,	- Lack information related to ethical AI, privacy, and AI governance mechanisms. - Are basically commercial, marketing, or advocacy documents lacking normative or empirical grounding. - Are duplicates of the same policy, for example, an older draft, summary, or unofficial translation. - Focus exclusively on AI performance technically, but lacks AI ethics and governance information.

Statistically, to identify major trends at global, continental, and national levels. The researcher coded all the documents, but the coding decisions were discussed on a regular basis with the supervisor to ensure that they aligned with the interpretive judgement. The coding framework was developed iteratively, with some categories generated after reviewing AI ethics and governance. These include AI ethics principles such as transparency, privacy, and accountability; governance attributes such as advisory, binding force, and sanctions; and technical controls such as audit mechanisms and PETs.

Pilot coding was done using the framework, using 10 documents sampled from the corpus of academic, national, African, and global levels. This led to the refining of the categories, potentially reducing overlaps between codes. Consistency was ensured through re-coding of a 10% subsample of the documents later, and their results were compared with the initial coding. Where discrepancies arose, especially in categorization, the affected results were readjusted. The inclusion and exclusion criteria have also been defined (see Table 1). The analysis was done using MS. Excel. Documents were categorized and their frequencies determined, plus those of the principals and PETs.

3.4. Limitations of the Study

The sample of documents selected for this study was only those focusing on ethical AI and privacy to identify the practical strategies for better AI governance, with a further interest in Africa and Uganda, and so the results are not exhaustive but rather indicative. The study relied on publicly available regulatory documents for analysis and the use of secondary databases. Therefore, some policies may be left out, although they are in the geographical locations of the study and within Ethical AI, privacy, and governance. This study could be extended in the future by considering primary data collection by interviews.

3.5. Ethical Considerations

Publicly available regulatory documents with aggregated analysis from other studies were used, and no human subjects were involved in this study; therefore, no need for ethical approval as analysis of documents does not require one. Furthermore, the citations in this study are publicly available, along with the policy documents.

4. Results and Discussion

4.1. Overview of the Analyzed Corpus

This study examined 190 documents, which were screened for relevance to ethical AI, Privacy, AI governance, and their related concepts. Out of these, only 132, which covered global, regional, and national frameworks, and those from technical and scholarly sources, were considered for the study. From these, 37 are frameworks at the global and regional levels, for example, the OECD AI Principles, UNESCO Recommendations on AI Ethics, and the African Union Continental AI Strategy, among others. 18 are from non-African countries that have been selected purposively to check their convergence with the global and regional frameworks. 26 are from African countries, which include national strategies, data protection, and privacy policies. The remaining 51 are combinations of academic papers and technical papers.

Table 2. Frequency and percentages of document groupings included in the study

Document Groups	Frequency	Percentage (%)
Global & Regional	37	28.03
African National	26	19.69
Non-African National	18	13.64
Academic & Technical papers	51	38.64
Total	**132**	**100.00**

These have been included to examine the ethical AI principles and Privacy Preserving Techniques being recommended by academia and technical experts in comparison to their inclusion in the policy documents at various levels for AI governance. The percentages of these groups in the documents are shown below (see Table 2). These individual documents were scanned for their adherence to ethical principles applicable to AI, PETs, and their technical controls and enforcement strategies, as well as evidence of their operationalization in the real world for AI governance.

4.2. Ethical AI Principles Coverage

The frameworks, which are categorized as Global, African National, and Non-African National, have ethical principles, although they have wide differences in terms of depth and implementation. The ethical AI principle that appears highest in these papers is privacy and data protection with 90.9%, followed by transparency at 78.8%, accountability at 72.7%, fairness and discrimination stand in the middle at 69.7% followed by human oversight mentioned at 61.3%, safety and robustness as the second last at 55.3% and lastly data sovereignty with 48.5%. More details are in Table 3 (see Table 3). These are consistent with the findings of Corrêa et al. [9]. As already alluded to, the operationalization of these documents is not uniform, as only 36% of those that have mentioned transparency have clearly spelled out mechanisms for controlling transparency, such as public disclosure obligations, traceability logs, and documentation of models, among others. In privacy, only 38.6% have specified their measurable ways of operationalization, such as differential privacy, encryption, and audits, among others. The European Union Artificial Intelligence Act (2024) provides the clearest, most wide-ranging guidance for operationalization, stipulating the responsibilities of institutions, consequences, and even the assessment of impact. Following suit is the OECD Recommendation on Artificial Intelligence (Revised, 2024).

Conversely, the African Union Continental AI Strategy tends to be more of a recommendation, as it lacks regulations that make it enforceable, but the concepts it entails are very strong, for example, inclusion, data sovereignty, Ubuntu values, among others. This indicates how the operationalization of these principles in AI governance is still a challenge. These findings also agree with Corrêa et al. (2023), who reported only 2% implementation, and they have gone on to identify that the EU AI Act (2024) is more detailed for operationalization, just like other legally binding frameworks.

4.3. Privacy-Preserving Techniques Adoption

The most referenced Privacy Enhancing Technologies (PETs) are Data Protection Impact Assessments (DPIAs) and data minimization, with technical regulations at 62.5 %. The immediate followers appear in 61% of the documents; these include pseudonymization as well as encryption. Logging mechanisms are referenced in 52% of documents, just like audit trails. The more advanced methods of PETs, particularly Differential Privacy, Homomorphic Encryption, Multi-Party Computation, and Federated Learning, appear more often in technical and academic papers than in policy and regulatory papers, at 34%, 17%, and 28%, respectively. It can be argued that although these advanced PETs are common among academia and technical experts, their integration into policy for better AI governance will take time [56, 57].

Table 3. Percentage of ethical AI principles in AI governance instruments

Ethical AI Principles	Percentage of Inclusion (%)
Privacy and data protection	90.9
Transparency	78.8
Accountability	72.7
Fairness and Non-discrimination	69.7
Human rights	61.3
Safety and robustness	55.3
Data sovereignty	48.5

4.4. Enforcement and Governance Strength

Legally binding documents in this study make up 42%. They are mainly data/ privacy protection or regulations related to AI. Further inspection has shown that 36% have clear requirements for supervision, and 30% have explicitly outlined sanctions. The soft laws frameworks, major strategies, and guidelines make up 58% of the total corpus. They emphasize the operationalization of sandboxes, capacity building, and ethics cards. They remain at an aspirational level as it is not clear regarding enforcement details. The more binding frameworks have up to 62% adoption of the PETs, which is higher than for the non-binding frameworks. This exemplifies the need for AI governance requiring concerted factors ranging from regulatory authority via the independence of institutions to resources like finance to operationalize ethical and privacy in maintaining practices.

4.5. Gaps between Guidance and Practice

About 39% of the reviewed papers that articulated ethical principles did not include operational and technical controls. 19% stated controls clearly, but excluded requirements to publish them publicly for transparency purposes. On the side of procurement, a staggering 14% have outrightly included compliance clauses regarding enforcement that are based on procurements in their contracts for purchases. Policy intentions are least implemented in jurisdictions with resource constraints. Nations that have stronger institutions, like Egypt, Kenya, and Ghana, are closing the gap as they incorporate sandboxes with oversight regulations in a sector-based approach. Education, finance, and health sectors, which already exhibit strong compliance with regulations, are far ahead in infusing audit practices and PETs.

4.6. Awareness and Compliance of Stakeholders

Stakeholder training is incorporated into 42% of the frameworks, as transparency boards and or public reporting are covered by 20% and pilot programmes or AI sandboxes account for 27%. Correlational studies show that stakeholder awareness and PETs adoption are higher by 18% in nations that are actively engaged in AI sandboxes and other pilot initiatives. This implies that multi-stakeholder engagement for AI governance, coupled with iterative experimentations with local expert training, is core to increasing policy uptake for better AI governance.

The results show convergence of ethical principles at the global level, although their uptake for daily governance is low. The commonest ethical AI principles are privacy, fairness, and accountability, despite a lack of concrete tools to measure them, like audits, traceability systems, PETs, and post-market monitoring. The frameworks that are more binding, such as the European Union's GDPR and its AI Act, tend to be more widely adopted due to their technical controls, whereas those that are mainly advisory, such as the AU Continental AI Strategy, are less widely adopted.

Lack of enforcement and verification mechanisms that make it difficult to translate the ethical AI principles and PETs into practice accounts for up to 40%. Not only this, but also, countries, especially those with low- and middle-income status, have limited numbers of technical experts, funding, and the technological tools required for better AI governance. Data governance across borders is limited by the different standards, how they are transferred, and the sovereignty standards, making it difficult for cross-border players to implement AI governance. Implementation of the PETs is met with a fierce shortage of technical experts' skills and limited infrastructural capacity for the deployment of cutting-edge PETs like Federated Learning, Homomorphic encryption, among others.

Lastly, the AI systems that are owned by the companies keep their operations secret to maintain their businesses, which makes it difficult for them to get audited by a third party to verify their claims. These challenges make it hard to implement ethical AI principles with the PETs. To reduce the gaps in implementation of the ethical AI principles and PETs, the following recommendations are proposed: Governments should include a mandatory audit of AI systems that pose high risks to the public in terms of privacy and other ethical issues. This should provide the mechanisms of operationalization, monitoring, penalties, and redress mechanisms explicitly and make simplified summaries available for public scrutiny. This will increase trust from the public and compliance from players.

AI systems that deal in sensitive data like personal information and are multi-party in that they share such data should be required to embed at least one of the PETs. Differential privacy and federated learning are examples that increase data protection. These should be accompanied by DPIAs to make tracking easy for interventions to be provided. Generating frameworks like the NIST RMF using the sandbox experiments could provide guidelines to ascertain compliance of AI vendors and Agencies, so that they get certified using these standards. Making it easy for the vendors to implement, but also governing bodies to intervene and ensure compliance.

For a vibrant AI governance, fundings need to be secured to strengthen authorities that become capable of carrying out investigations on breaches of AI ethics and privacy, issuing sanctions, and then providing reparation mechanisms. These bodies can ensure all-around AI governance. This is both at the business and government levels.

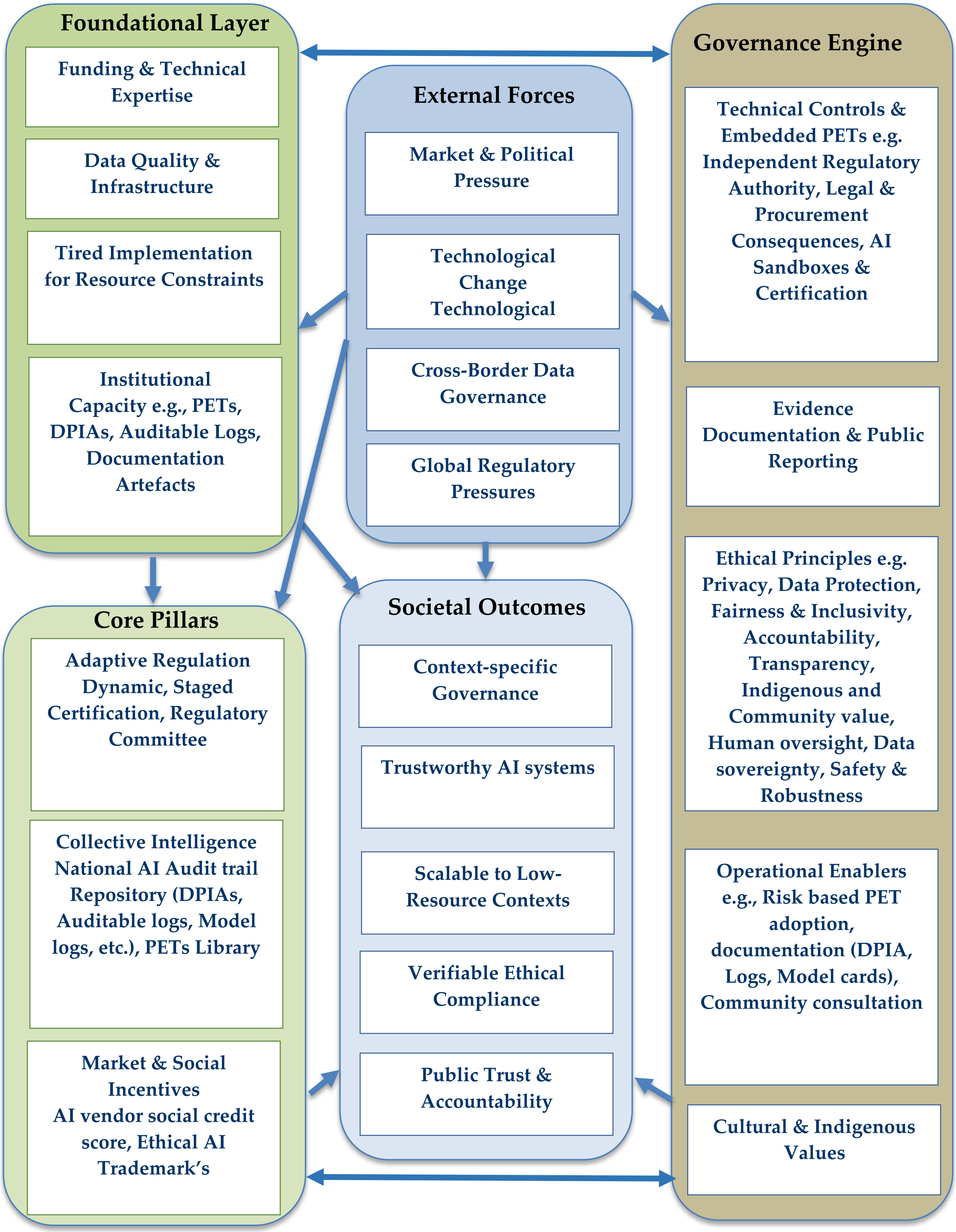
Foundational Layer
Funding & Technical Expertise
Data Quality & Infrastructure
Tired Implementation for Resource Constraints
Institutional Capacity e.g., PETs, DPIAs, Auditable Logs, Documentation Artefacts
External Forces
Market & Political Pressure
Technological Change Technological
Cross-Border Data Governance
Global Regulatory Pressures
Governance Engine
Technical Controls & Embedded PETs e.g. Independent Regulatory Authority, Legal & Procurement Consequences, AI Sandboxes & Certification
Evidence Documentation & Public Reporting
Ethical Principles e.g. Privacy, Data Protection, Fairness & Inclusivity, Accountability, Transparency, Indigenous and Community value, Human oversight, Data sovereignty, Safety & Robustness
Operational Enablers e.g., Risk based PET adoption, documentation (DPIA, Logs, Model cards), Community consultation
Cultural & Indigenous Values
Core Pillars
Adaptive Regulation Dynamic, Staged Certification, Regulatory Committee
Collective Intelligence National AI Audit trail Repository (DPIAs, Auditable logs, Model logs, etc.), PETs Library
Market & Social Incentives AI vendor social credit score, Ethical AI Trademark's
Societal Outcomes
Context-specific Governance
Trustworthy AI systems
Scalable to Low-Resource Contexts
Verifiable Ethical Compliance
Public Trust & Accountability

Fig. 1 Proposed AI governance framework for resource-constrained settings

AI governance stakeholders need to improve accountability through legal and ethical frameworks, incorporating compliance evidence as precursors to successful public and private procurement and contracting. These can include bias reports, incidence logs, and model cards, among others, to increase compliance amongst the AI providers and increase trust. Global corporations, with already established governments and entities, ensure the sharing of information for completeness in addressing issues.

Uganda and any other low- and medium-income countries should consider values that are indigenous to their people, for example, inclusiveness and communal accountability in their frameworks to guide the development and deployment of AI in areas like healthcare, justice, agriculture, education, and social services to address the needs of their people. Ensuring data residency and community consultations are done to improve inclusiveness, thereby promoting transparency and trust, hence increasing chances of survival of the AI ecosystems. Some unique biases for such small minorities can easily be eliminated by involvement and using data that are locally collected [36].

Based on the above, an AI governance framework that is easily adaptable for low- and medium-income environments, having its foundation based on numerous funding that can be used for improving the technical expertise at PETs and regulations, ensuring high-quality data that is based on a set national data governance protocol, and improved AI infrastructure that is shared, for example AI sandboxes is proposed. Finally, these data quality needs to be hinged onto ethical AI impacts assessments to reduce biases on minority groups by incorporating review boards and or indigenous knowledge protocols right from data collection, preparation, model training, evaluation, deployment, and after-market monitoring. The governance engine is hinged on technical controls where the PETs adoption is based, as in the EU GDPR [58, 59].

These can be phased in to address resource constraints, with Risks needing to be categorized and the high-risk AI system put under compulsory use of advanced PETs like DP, and medium-risk systems using less advanced PETs like FL, Synthetic data, among others. Having an independent regulatory authority that controls procurement based on vendors adhering to requirements, such as being certified, among others. They also ensure that there is a national repository of audit trails for ease of use by startups. These are grounded in cultural, indigenous, and international values that align well with the ethical AI principles.

All these are affected by the external factors of market and political pressure, technological change, cross-border data governance, and global regulatory pressures, demanding that the expert body involved in the governance cycle should ensure alignment for fitting not only in the local market but also in the global market, where it is required. The frameworks also posit for an adaptive regulation where certification is staged and dynamic to cater to the rapidly changing AI technology. This is systematically illustrated (see Figure 1). This proposed AI Governance Framework is modeled for a Resource-Constrained Environment by allowing for phased and agile implementation.

5. Conclusion

Across global and African contexts, there is a remarkable convergence of ethical AI principles, though implementations vary widely. Privacy and data protection are recognized as one of the most important principles at 90.9 %. Less than half of the policies and guidelines documents have explicitly included how it can be verified, though. Frameworks that are binding, particularly with respect to legal obligations, and supported by stronger funding institutions, show higher adoption of PETs and more pronounced enforcement practices. For the trustworthiness of AI governance at the policy and practice levels, the ambitious declarations in the frameworks need to be translated into more practical, evidence-based guidelines for accountability.

For example, making it compulsory to conduct risk auditing and embedding PETs from design through procurement to deployment. The actualization also requires the regulator to be empowered and fully authorized and to be funded to act accordingly. Future research should use quantitative benchmarking for PET design and deployment, evaluating the quality of AI audits and monitoring incidents after the sale of AI products to ascertain real-world accountability. Lastly, case studies that are more in-depth on specific areas of Generative AI and Autonomous systems would provide a richer understanding of how ethical and privacy mechanisms work effectively despite the fast-changing technology.

Data Availability

The data that was extracted and used for this study can be obtained from the corresponding author on request through his email.

Authors' Contributions

Conceptualization, Taban Habibu; Methodology, Ceasar Obudra; Formal Analysis, Ceasar Obudra; Resources, Ceasar Obudra and Taban Habibu; Data Curation, Ceasar Obudra; Writing – Original Draft Preparation, Obudra Ceasar; Writing – Review & Editing, Taban Habibu and Ceasar Obudra; Visualization, Ceasar Obudra; Supervision, Taban Habibu.

Acknowledgments

The author sincerely thanks Dr. Taban Habibu, supervisor, for his tireless guidance and patience, without which this work would not have been possible.

References

[1] Homero Gil de Zúñiga, Manuel Goyanes, and Timilehin Durotoye, "A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research," *Political Communication*, vol. 41, no. 2, pp. 317-334, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Aastha Pant et al., "Ethics in AI Through the Practitioner's View: A Grounded Theory Literature Review," *Empirical Software Engineering*, vol. 29, no. 3, pp. 1-48, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yuzhou Qian, Keng L. Siau, and Fiona F. Nah, "Societal Impacts of Artificial Intelligence: Ethical, Legal, and Governance Issues," *Societal Impacts*, vol. 3, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Keng Siau, and Weiyu Wang, "Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI," *Journal of Database Management*, vol. 31, no. 2, pp. 74-87, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] N. C. Walker et al., Privacy, Edition 1.0 Research, In AI & Human Rights Index, AI Ethics Lab at Rutgers University, New York, 2025. [Online]. Available: https://aiethicslab.rutgers.edu/glossary/privacy/

[6] Amna Batool, Didar Zowghi, and Muneera Bano, "AI Governance: A Systematic Literature Review," *AI and Ethics*, vol. 5, no. 3, pp. 3265-3279, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Rehema Baguma et al., "Towards an Artificial Intelligence Readiness Index for Africa," *Digital-for-Development: Enabling Transformation, Inclusion and Sustainability Through ICTs: 12th International Development Informatics Association Conference, IDIA 2022*, Mbombela, South Africa, pp. 285-303, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Mariarosaria Taddeo, Alexander Blanchard, and Christopher Thomas, "From AI Ethics Principles to Practices: A Teleological Methodology to Apply AI Ethics Principles in The Defence Domain," *Philosophy and Technology*, vol. 37, no. 1, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Nicholas Kluge Corrêa et al., "Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance," *Patterns*, vol. 4, no. 10, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Regulation (EU) 2024/1689 of the European Parliament and of the Council, European Union, 2024. [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj

[11] *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, National Institute of Standards and Technology, 2023. 10.6028/NIST.AI.100-1. [CrossRef] [Google Scholar] [Publisher Link]

[12] Karen Yeung, "Recommendation of the Council on Artificial Intelligence," *International Legal Materials*, vol. 59, no. 1, pp. 27-34, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] UNESCO, Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2021. [Online]. Available: https://www.ohchr.org/sites/default/files/2022-03/UNESCO.pdf

[14] UNESCO, Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2022. [Online]. Available: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

[15] Lukas Weidener, and Michael Fischer, "Proposing a Principle-based Approach for Teaching AI Ethics in Medical Education," *JMIR Medical Education*, vol. 10, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Lauren Feiner, Dutch Regulator Slaps Clearview AI with $33 Million Fine and Threatens Executive Liability, The Verge, 2024. [Online]. Available: https://www.theverge.com/2024/9/3/24234879/dutch-regulator-gdpr-clearview-ai-fine?

[17] Patrick J. Grother, Mei L. Ngan, and Kayee K. Hanaoka, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, National Institute of Standards and Technology, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Kashmir Hill, *Wrongfully Accused by an Algorithm*, 1st ed., Ethics of Data and Analytics, Auerbach Publications, pp. 138-142, 2022. [Google Scholar] [Publisher Link]

[19] Catherine Stupp, "Fraudsters used AI to Mimic CEO's Voice in Unusual Cybercrime Case: Scams using Artificial Intelligence are a new Challenge for Companies," *The Wall Street Journal: Cybersecurity*, vol. 30, no. 8, 2019. [Google Scholar] [Publisher Link]

[20] Mikaela Pisani, AI Ethical Framework, 2024. [Online]. Available: https://www.rootstrap.com/blog/ai-ethical-framework

[21] Johnson Emily, AI Ethics Frameworks: 10 Essential Resources to Build an Ethical AI Framework, 2025. [Online]. Available: https://www.secureitworld.com/blog/ai-ethics-frameworks-10-essential-resources-to-build-an-ethical-ai-framework/

[22] Adelaide O'Brien et al., Ethical Application of Artificial Intelligence Framework, ACT-IAC Accelerating Government, 2020. [Online]. Available: https://www.actiac.org/documents/act-iac-white-paper-ethical-application-ai-framework

[23] Margaret Mitchell et al., "Model Cards for Model Reporting," *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, New York, NY, United States, pp. 220-229, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] M. Arnold et al., "Factsheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity," *IBM Journal of Research and Development*, vol. 63, no. 4/5, pp. 6: 1-6: 13, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[25] Toju Duke, *AI Principles*, Building Responsible AI Algorithms, Apress, Berkeley, CA, pp. 15-35, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Microsoft, Frontier Governance Framework, pp. 1-15, 2025. [Online]. Available:
https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Frontier-Governance-Framework.pdf

[27] Intel Responsible Artificial Intelligence (RAI) Principles. [Online]. Available:
https://www.intel.com/content/www/us/en/artificial-intelligence/responsible-ai-principles.html

[28] Virginia Dignum, "Responsible Artificial Intelligence-from Principles to Practice: A Keynote at the Web Conf," *ACM SIGIR Forum*, vol. 56, no. 1, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[29] Foundation Models: Opportunities, Risks and Mitigations 2 Foundation Models: Opportunities, Risks and Mitigations | Attribution, IBM, 2026. [Google Scholar] [Publisher Link]

[30] Jacob snow, Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots, American Civil Liberties Union, 2018. [Online]. Available: https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28?

[31] David Tong, How Big Tech's AI 'Neutrality' Enables Ethical Disasters: 3 Cautionary Tales for Business Leaders, David Lee Tong, Singapore, 2025. [Online]. Available: https://davidleetong.com/how-big-techs-ai-neutrality-enables-ethical-disasters-3-cautionary-tales-for-business-leaders/?utm_source=chatgpt.com

[32] Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 388-399, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[33] Center for Data Ethics and Innovation, PETs Adoption Guide, 2025. [Online]. Available: https://cdeiuk.github.io/pets-adoption-guide/what-are-pets/

[34] Taban Habibu, and Ayo P. Julius, "Cybersecurity in the Internet of Things (IoT) – Review," *DS Journal of Cyber Security*, vol. 3, no. 3, pp. 15-38, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Gonzalo Munilla Garrido, Vivek Nair, and Dawn Song, "SoK: Data Privacy in Virtual Reality," *Proceedings on Privacy Enhancing Technologies*, vol. 2024, no. 1, pp. 21-40, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[36] Algorithm Watch, "AI Ethics Guidelines Global Inventory-Algorithm Watch," Algorithm Watch, 2020. [Online]. Available: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/

[37] Qinbin Li et al., "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3347-3366, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[38] Ying Zhao, and Jinjun Chen, "A Survey on Differential Privacy for Unstructured Data Content," *ACM Computing Surveys (CSUR)*, vol. 54, no. 10s, pp. 1-28, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[39] Janak Ghansham Dhokrat et al., "A Framework for Privacy-Preserving Multiparty Computation with Homomorphic Encryption and Zero-Knowledge Proofs," *Informatica*, vol. 48, no. 21, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[40] African Union, Continental Artificial Intelligence Strategy: Harnessing AI for Africa's Development and Prosperity, UNESCO, 2024. [Online]. Available: https://www.iicba.unesco.org/en/africa-education-knowledge-platform/continental-artificial-intelligence-strategy-harnessing-ai-africas-development-and-prosperity

[41] Kenneth Muhangi, "Overview of the Data Protection Regime in Uganda," *Journal of Data Protection and Privacy*, vol. 3, no. 1, pp. 82-92, 2019. [Google Scholar] [Publisher Link]

[42] Marc Rotenberg, "Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Council Eur.)," *International Legal Materials*, vol. 64, no. 3, pp. 859-902, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[43] Center of AI and Digital Policy, Statement on AI Policy for Democratic Nations from the Center of AI and Digital Policy (CAIDP) for the 2025 G7 Summit Meeting, 2025. [Online]. Available: https://www.caidp.org/resources/g7/

[44] Vinton G. Cerf, "The Global Digital Compact," *Communications of the ACM*, vol. 67, no. 10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[45] "AI, Data Governance, and Privacy: Synergies and Areas of International Co-Operation," *OECD*, 2024. [Google Scholar] [Publisher Link]

[46] Regulations on the Management of Internet user Account Information - State Council Departmental Documents - China Government Website, Cyberspace Administration of China, 2022. [Online]. Available: https://www.gov.cn/zhengce/zhengceku/2022-06/28/content_5698179.htm

[47] Michael Sheng, and Hoi Tak Leung, New Generative AI Measures in China, 2023. [Online]. Available: https://www.ashurst.com/en/insights/new-generative-ai-measures-in-china/

[48] Provisions on the Administration of Deep Synthesis Internet Information Services, China Law Translate, 2022. [Online]. Available: https://www.chinalawtranslate.com/en/deep-synthesis/

[49] The Artificial Intelligence and Data Act (AIDA)-Companion Document, Innovation Science and Economic Canada, 2023. [Google Scholar] [Publisher Link]

[50] Mauritius AI Strategy, The OECD.AI Policy Navigator, 2018. [Online]. Available: https://oecd.ai/en/dashboards/policy-initiatives/mauritius-artificial-intelligence-ai-strategy-7829

[51] The National AI Policy - Rwanda, Ministry of ICT and Innovation, 2022. [Online]. Available: https://africadataprotection.org/sources/Artificial_Intelligence_Policy.pdf

[52] The United Republic of Tanzania: Ministry of Communication and Information Technology, Guidelines for the Secure and Ethical use of Artifical Intelligence in Tanzania (Draft Version One), United Republic of Tanzania, 2024. [Online]. Available: https://www.mawasiliano.go.tz/uploads/documents/sw-1749982790-Guidelines for AI ethical USE Guideline MICIT 2025ver.pdf

[53] The National Artificial Intelligence Strategy of Mauritania for 2025-2029, The Ministry of Digital Transformation, Innovation, and Modernization of the Administration, 2024. [Online]. Available: https://dig.watch/resource/the-national-artificial-intelligence-strategy-of-mauritania-for-2025-2029-draft-3

[54] Amr S. Tala et al., Egypt National Artificial Intelligence Strategy, 2nd ed., 2021. [Online]. Available: https://ai.gov.eg/SynchedFiles/en/Resources/AIstrategy%20English%2016-1-2025-1.pdf

[55] Kenya AI Strategy 2025 - 2030, Kenyan Ministry of Information, 2025. [Online]. Available: https://ict.go.ke/sites/default/files/2025-03/Kenya%20AI%20Strategy%202025%20-%202030.pdf

[56] Tunisia - National AI Strategy (2021-2025), National AI Strategy and Roadmap, 2022. [Online]. Available: https://regulations.ai/regulations/tunisia-2021-national-ai-strategy

[57] Part 2: Emerging AI Governance in Africa, AI Governance, 2024. [Online]. Available: https://www.trust.org/toolkit/part-2-emerging-ai-governance-in-africa/

[58] Government Gazette Republic of South Africa, Act, 2013. [Online]. Available: https://www.gov.za/sites/default/files/gcis_document/201409/3706726-11act4of2013protectionofpersonalinforcorrect.pdf

[59] Republic of Ghana National Artificial Intelligence Strategy: 2023-2033, 2022. [Online]. Available: https://www.africadataprotection.org/Ghana-AI-Strat.pdf

[60] Wegene Demisie Jima, Tesfaye Addisu Tarekegn, and Taye Girma Debele, "The Landscape of Artificial Intelligence Implementation in Ethiopia," *Asric Journal on Natural Sciences*, vol. 3, no. 2, pp. 1-209, 2023. [Google Scholar] [Publisher Link]

[61] Uganda's National 4IR Strategy a continental 4IR Hub that Enables a Smart and Connected Ugandan Society, Ministry of ICT and National Guidance, 2021. [Online]. Available: https://ict.go.ug/site/documents/Executive-Summary-Ugandas-National-4IR-Strategy.pdf

[62] Information Technology - Artifical Intelligence - Management System - ISO 42001, International Organisation of Standardisation, 1st ed., 2023. [Online]. Available: https://www.iso.org/standard/42001

[63] Information Technology - Artificial Intelligence - Guidance on Risk Management - ISO/IEC 23894, International Organisation of Standardisation, 1st ed., 2023. [Online]. Available: https://www.iso.org/standard/77304.html

[64] Huw Roberts et al., "Global AI Governance: Barriers and Pathways Forward," International Affairs, vol. 100, no. 3, pp. 1275-1286, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[65] AI Ethics Guidelines Global Inventory, Algorithm Watch, 2019. [Online]. Available: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/

[66] The State of AI in 2023: Generative AI's Breakout Year, 2023. [Online]. Available: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

[67] "IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks--Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016)*, 2021.