International Journal of Computer Applications Technology and Research
Volume 13–Issue 08, 01 – 10, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1308.1001

Ethical Considerations in the Development and Deployment of AI-Powered Systems

Asnath Nyachiro
Department of Information Technology
Institute of Computing and Informatics
Technical University of Mombasa
Mombasa, Kenya

Fullgence Mwakondo
Department of Information Technology
Institute of Computing and Informatics
Technical University of Mombasa
Mombasa, Kenya

Abstract: The aim of the study was to give a general account of the ethical principles that need to be infused in the design, development and deployment (DDD) of AI powered systems. The study was guided by two theories, the deontological theory of ethics and the descriptive theory. The study used qualitative survey design and convenient sampling was done to get a sample of 20 experts in various fields related to moral philosophy, AI software engineers, policy makers, end users and legal practitioners. Participants were conveniently selected to cover a wide spectrum of players involved in the AI development and deployment. The data collection focused of collecting qualitative data through focused group discussion and expert interviews that were guided by interview schedules. The findings showed that various stakeholders in the AI industry had concerns related to bias, privacy issues, accountability, transparency and prejudice. The participants highlighted concerns on privacy issues as important. Participants demonstrated concerns over the safety of personal data in AI development, data security challenges and propensity for misuse for misuse of personal information. Participants stressed on need of developing AI systems and testing them to prevent the reprojection of inequalities through the AI outputs. Other emergent themes were accountability, responsibility for AI generated decisions, lack of clear frameworks for handling errors or harm. Transparency was repeatedly mentioned as needing explainability in developing trust in AI products by all stakeholders in ethical their ethical deployment. The research concludes handling of ethical issues in AI development and use necessitates a multidisciplinary approach that combines technological, regulatory and ethical perspectives. The study recommends establishing comprehensive rules and regulations to govern AI development, stressing on fairness and inclusivity in the designing of algorithms alongside formation of functioning mechanisms for oversighting AI industry in achieving accountability and transparency. This research value is that it advances knowledge of AI ethics further by providing new insights and dimensions from the diverse perceptions of AI stakeholders. It provides the impetus for including ethical issues in the whole lifecycle of AI systems which ranges from design, development to deployment. Further investigations need to be done in order to provide a framework for designing legislation that addresses the intricate ethical issues emergent with AI technologies.

Keywords: AI Powered Systems; Personal Data Security; Algorithm Bias; AI Design and Development; AI Trust; Transparency and Accountability

1. INTRODUCTION

Ethical issues are critical in ensuring that Artificial Intelligence (AI) powered systems are not used to violate the rights of data subjects globally. This paper adopts the Council of Europe's definition of AI as "[a] set of sciences, theories and techniques whose purpose is to reproduce by a machine the cognitive abilities of a human being" (OECD 2019). AI involves being cognizant of its impact on the individual and the society. It is from this angle that the subject matter of ethical considerations of AI-powered systems must be factored in the deployment, as well as in the design and development phase of AI applications. Deployment in AI discourses refers to the process of applying machine learning functions to an existing problem-related environment in resolving challenges. This is done through pragmatic decision making based on the available data that is used to train the AI. AI deployment marks the terminal end of the life cycle of machine learning and can be awkward if not well designed or implemented. There is a need to explore and understand the digital technologies and their developments in the wider context of policies, regulations, and legislation. Infusing ethics into the deployment of AI powered systems safeguards human rights and improves issues of trust and accountability in them by data subjects whose data may be used by the AI systems.

Ethical factors play a vital role in ensuring there is transparency and explainability is advancing user trust. AI development requires collaborative efforts between developers, data providers and relevant stakeholders in order to ensure the effective sharing of information leads to problem resolutions while addressing privacy issues. In the recent past, growth of AI has revolutionized industrial operations and technologies, affecting how work is done and outcomes achieved in various fields such as healthcare, transport, education and energy sectors (Carrasco Ramírez & Islam, 2024; Dunsin et al., 2024; Wang'ang'a, 2024). As the functions of AI in various fields continues to increase in scope and intensity, they become vital factors in the way AI works such as the ability to diagnose patients and give prompt feedbacks, the growth of automated self-driving cars alongside other machine learning systems (Bohr and Memarzadeh 2020). According Sullivan, ethics in AI helps in ensuring that such digital technologies and machine learning

"respect human values, avoid undue harm, and act as beneficial force in society… [by encompassing] fairness, privacy, accountability, transparency and human rights" (Sullivan, 2023). In the life cycle of AI systems, the way they are used and the manner in which they are developed should be harmonious with the respect of human rights, the independence of individuals and conform to diversity issues without prejudice or discrimination. The AI systems are created to be at the disposal of serving human interest and not the other way which means they have to be ethically aligned to human values, which must be evident during deployment.

The question of data handling and processing strategies of AI-powered systems requires enforcement of control measures that lead to compliance with data security concerns. This calls for observance of accountability and trust issues, ability to explain the outcome of, and the functioning of the AI systems, privacy, compliance to data governance safety regulations, risk awareness, responsible and bias mitigation. This is under the domain of data governance at both individual and organizational level that is subject to data management policies and relevant legal requirements. The whole way in which AI operates should be addressed with caution since the main goal of developing ethics in AI is to ensure public trust and the use of these technologies (Ouchchy, Coin, and Dubljević 2020). The development and implementation of AI technologies raises deep challenges regarding its influence on society, individual rights, and the future of mankind, ranging from data privacy and prejudice to openness and responsibility (Bartoletti 2019). There are generally six principles that constitute the value systems that are to be taken into account in the development of AI systems, the deployment and their uses. This principles act as the ethical beacons aligning the AI application to the "societal values, respect of human rights, promote fairness, transparency and accountability (Ejaz & Olaoye, 2024). Therefore, the subsequent sections discuss the ethical considerations during the development and deployment of AI. Due to this, during the deployment phase, it is incumbent upon the developers to ensure that transparency and explainability of the AI systems. This by guaranteeing that the said AI systems provide explanations for the decisions they make and actions outputs to both the data users and the data subjects.

2. THEORETICAL FRAMEWORK

This study is guided by the descriptive theory and the virtue theory of ethics. The descriptive theory is concerned with outlining the characteristics that explain a phenomenon (Chandler 2024). In this way, it will help in providing a structure for describing the ethical issues as they apply within the field of AI deployment. The descriptive theory, in view of the virtue ethics theory, describes the desired conditions upon which AI developers are duty bound to ensure their products conform to moral principles that require the right action to be done always by the AI powered systems (Loeb et al., 2017; Waelen, 2022)). In this way, the deontological theory of ethics also becomes applicable to the study. Secondly, the descriptive theory aids in outlining how the deontological ethics theory applies to AI development and deployment. The deontological ethical theory requires the application of universal moral rules of right or wrong. From this perspective, the universal moral rules of dos and don'ts make everybody duty bound to do the right thing in the development and application of AI systems (Ozone, 2019). The deontological theory imposes moral obligations that must be adhered to against all odds for purposes of achieving the high moral standing. In this case, it will be presumed that AI powered system developers have a duty to design, develop and deploy products that are duty-bound to respect and uphold human values, and protect human rights. Violation of human rights is morally wrong. All humans have a duty to uphold good moral characters and therefore, the implements they construct must be developed and used in accordance to the accepted moral codes in the society.

3.RESEARCH METHODS AND MATERIALS

The study used qualitative survey design to undertake an investigation of the moral issues affecting the development and deployment of AI-powered applications and technologies. to in providing an in-depth understanding of issues related to ethical dimensions of AI systems. This research design was justified because of its ability to explicate in-depth, nuanced and complex insights into the subject of moral conundrums that various interested parties and stakeholders in the AI industry have to face. The target population consisted of 20 participants who were conveniently selected to meet the objectives of the study based on their specialties in the legal field, AI developers and researchers, moral philosophers, policy makers concerned with AI development and utilization and users of AI powered systems in Nairobi, Kenya. The data was conducted by using semi-structured interviews based on a flexible interview guide. The interview outcomes were then recorded, and transcribed with anonymization done to provide confidentiality to the participants. Also, reliance on secondary sources of information was also done where reviews of academic literature, industrial reports, and media sources was done. Analysis of the data was done to using thematic, textual and contextual analyses. Validity was achieved through undertaking a thorough cross-checking of the respondents, and allowing them to review the information they had provided in order to ascertain the accuracy of their information and their interpretations of the data

4. FINDINGS

The analysis of the data showed that the perceptions of the respondents placed a high value on the ethical considerations that must be accounted for in the DDD of the AI systems. The respondents noted that AI has had a tremendous effect in changing industrial processes, revamping and changing economies, and ultimately redefining the human-technology interaction. With this development, AI powered systems have impacted on the nature of the daily human life in which ethical considerations becoming a serious concern in their development and usage. The findings demonstrate ethics is important in AI technologies which is a means of ensuring

that they contribute positively to the society while at the same time, mitigating its adverse effects, limiting potential for harm to the highest degree possible and safeguarding the rights of everyone.

The findings showed that there is need to achieve the ethical standards of fairness in the DDD of AI systems. The study noted that algorithms can be trained to achieve various cognitive competencies such as predicting criminal behavior in vetting of job applicants during the hiring process, and this has the potential of worsening existing biases. This may lead to prejudice and discriminatory practices against vulnerable groups, for example, AI may unfairly discriminate against candidates from a given ethnic or demographic group based on gender, race et cetera if the AI training was constructed with implicit bias in it. To achieve fairness, the respondents were of the view that the algorithms used in AI must be developed in a transparent and accountable manner that is open to scrutiny where possible. The findings showed that the best antidote to this problem would be through frequent auditing of the AI systems and reinforcing avoidance of discrimination, and perpetuation of inequalities in the AI systems.

Privacy was noted to be another important ethical aspect in the DDD of AI systems. In the first place, the training of AI systems is reliant on vast amounts of data in order to undertake its functions effectively. This requires, in many instances, collection of huge amounts of personal data with the risk of storage, analysis and termination of the data after its useful life is over creating risks. This includes data breaches risks, unauthorized access, improper processing and misuse of personal information which can have severe backlash to the data subject, user or data controller. The respondents were unanimous that to control this, the development and enforcement of ethical AI standards and practices involving stringent data protection strategies ensure that user's consent is obtained separately and in a transparent manner. Privacy by design is an ethical principle that also addresses this challenge by incorporating privacy factors as integrated aspects in the system functions. The importance of this is to build and sustain trust in the AI systems in the protection of personal data.

From an ethical perspective, the question of trust is built on the elements of transparency and accountability. This is particularly bothersome when the deep learning aspect of AI machines is shrouded in opaqueness in the way they arrive at the decisions which they make. In this, the question of the black box in the way the decision is made is difficult to explain, creating a "black box" phenomenon where the justification, reasoning and derivation of the specific AI decision remains a mystery. The consequence of this is that affected persons reliant on the AI decisions may lack a 'locus standi' in comprehending the basis of how decisions made affecting them were arrived at. When this happens, it undermines trust and reliability in the AI powered systems. This is because they become inexplainable and lack accountability a factor that propels users and stakeholders to fail to know and understand how the decisions derived by the AI are made, hence, makes it difficult for AI developers to be held accountable for the decisions that affect their policies, actions and work et cetera. This is a vital much-needed action that needs to be done in an open and transparent manner to build public confidence in the AI. In addition to that, it facilitates for the redesigning remedial actions of AI products when they are actually operating or deemed to be operating unfairly.

The necessity for ethical standards is made imperative by the fact of the possibility of high chances of AI being applied negatively. This misuse of AI technologies may result in adverse consequences for individuals, the society and the human civilization as is currently known. A good example are AI-powered surveillance systems, which may be utilized to impose authoritarian rule by manipulating the electorate, or used in deploying autonomous weapons in conflict zones with unpredicted results. Accordingly, the respondents were of the view that it is imperative to formulate and promulgate laws that restrict or totally prevent the nefarious application of AI. In this sense, AI will be used in the destruction of the society contrary to deontological theory of ethics that requires the preservation of human lives whether there is war or not by encouraging AI deployment only for the benefit of society.

While the survival of mankind is essential for the existence of future generations, ethical development of AI must proceed and weigh on the short and long-term consequences of these AI-powered technologies to the society. The advancement of AI systems could jeopardize job opportunities for many people and risk their livelihoods as one of their economic rights, through phenomena like job displacements, shifting labor markets, and either implicitly or expressly, establish and sustain new forms of inequity. The concept of the ethical development of AI must consider the ability to resolve the arising challenges from economic and social issues. This will go a long way towards advancing the sustainability of technological advancements that accrue benefits to the society as a whole without prejudicing the rights of vulnerable or disadvantaged groups and individuals.

The findings demonstrate the value of ethics being integrated in the DDD of AI powered systems. There is a serious need to ensure the ethical values of fairness, transparency, accountability, privacy and confidentiality, are inherently present in the DDD of AI powered systems within the broader aspects of socioeconomic development and welfare of society and of individuals. In this way, the element of promoting justice through deployment of Ai and respect of human rights and protecting the same rights should make AI a great tool for innovation and advancement. AI should be used to push the humankind development agenda forward through technology, progress and social harmony.

5.DISCUSSION OF FINDINGS

Data privacy refers to the ability of any individual to have and exercise their capacity in the control of their personal information. In that regard, it is concerned with the issue of data protection of sensitive information that is of a personal nature. AI-powered systems should be designed to protect the private information (personal data) concerning the data subjects. Ethically, it is the duty of the software designer and developer to ensure that the algorithms are trained in a manner that safeguards sensitive information. This traditionally follows the industrial practice of developing AI while adhering to established code of ethics that need to be followed (Gogoll et al., 2021). Secondly, the algorithm should require only information that is relevant for the purposes upon which it was built to address. The rapid advancement of artificial intelligence (AI) technologies has brought unprecedented opportunities and challenges across various sectors, including healthcare, finance, transportation, and education. As AI systems become increasingly integrated into daily life, the ethical considerations surrounding their development and deployment have come to the forefront of public discourse (Sheludko, 2023). This essay explores the ethical dimensions of AI-powered systems, focusing on key issues such as bias and fairness, privacy, accountability, transparency, and the impact on employment. The discussion is framed within the context of contemporary AI advancements and aims to provide a comprehensive understanding of the ethical landscape in AI technology.

The subject of bias and fairness is a dominant pressing issue in the field of ethical dialogues related to AI development and deployment (Singhal et al., 2024). The question of eliminating bias and achieving fairness in AI-powered systems and the decisions they make is crucial in making balanced predictions and fair decisions. Over time, AI is becoming increasingly essential in achieving positive transformation yet it is fraught with challenges, while more others are emerging in relation to the development and deployment of AI. AI-powered systems have the ability to enforce bias and unfair decisions based on the way they have been developed. In that regard, if the algorithms have been constructed in a manner that perpetuates bias and unfair decisions that are prejudicial, it means that the decisions and output of its functioning may as well be unethical.

This may happen where the AI-powered systems mirror the inherent inequalities in the society that may be discriminatory in nature. Discrimination contravenes the United Declaration of Human Rights, which states that all humans are created equal (United Nations, 2024). This is further reflected in the Kenyan national constitutions in Article 47 provides that “every person is equal before the law and has the right to equal protection and equal benefit of the law.” (The Constitution of Kenya, 2011). It is therefore proper that the DDD of AI powered products should reflect this principle in the way it is developed to reflect human values. The Kenya Data Protection Act of 2019 is a big step in the right direction toward securing sensitive information particularly of a personal nature in the age of emerging digital technologies (The Data Protection Act of 2019, 2019). This Act lays emphasis on issues of permission, based on informed consent, the statute emphasizes the need for informed consent and permission, transparency, and the right to privacy, and accountability that must be aligned with the with globally recognized international data protection standards. The legal framework provides a fundamental structure in enabling the ethical application of artificial intelligence (AI). It stipulates that information gathered for AI applications must be handled carefully and responsibly, maintaining strict secrecy. The foundation of the Act, which upholds strict data management procedures, reduces misuse risks, and safeguards individual rights, promotes the moral advancement of AI. Kenya's legal framework offers a strong basis for moral technology practices as AI develops.

This makes the AI-powered systems to be ethically constructed and lead to greater accountability in regard to promoting, respecting and fulfilling human rights. On the matter of reproducing prejudices unfairly, AI experts opine that where AI systems are trained to reflect historical bias or discriminatory trends that are unfair, the decisions they make may be unethical as that decision is generated from algorithms that accept such bias in the outcomes they generate. To illustrate this, an AI developed to help investigate crime but developed with algorithms that portray a particular of people as more prone to crime perpetuation is likely to misidentify criminals based on that algorithm. The facial recognition technology can perpetuate such a vice by ignoring this principle of non-discrimination in the human society based on prejudice. Improper, illegal and unethical treatment of vulnerable populations may be a consequence unethical AI bias in Kenya.resulting in violations of human rights (Akello, 2022). This may worsen by intensifying pre-existing socioeconomic inequalities. In order to advance the course of justice and prevent discriminatory outcomes in AI systems, it is a must that strong data protection strategies for AI systems be implement and ensure diverse, representative datasets are used in the development.

It is important to locate the sources of bias in the AI powered systems. This can help retrain the AI in addressing this challenge in the monitoring and evaluation processes. One must take note of the fact that AI systems are dependent on large datasets in order to develop a systematic pattern in the decision-making process (Aldoseri et al., 2023). Scholars have provided examples of developing AI systems tools trained on historical data may lead to bias in targeting criminals and may unfairly identify innocent persons as perpetrators of crime. Secondly, during the training phase of the AI development, it may become prone to sampling errors and biases which may result in either real or perceived abuse of human rights (Lane, 2022). A case in point is where identification of criminals may have relied unfairly on dark skinned people which may lead to misidentification of Africans as criminals which by itself may lead to unjust prosecution and conviction of

Africans as criminals. Finally, the question of algorithmic bias may inherently lead to the favouring a particular group to the detriment of others in a manner that is unfair, ethically and/or immorally wrong.

The question of bias in the development of AI systems and in the deployment has far reaching consequences on the individual and the society at large. While addressing it from the criminal justice system dimensions, the paper has already pointed out that it can result in the wrongful arrest of an individual or miscarriage of justice (Karmaza et al. 2021). Artificial intelligence has the ability to profile individuals and generate evidence that may be admissible in a court of law, which in reality may not be representing the true picture due to the inherent bias, that is within the development of the AI system, and reinforces existing inequalities and prejudices. In the field of employment, it may lead to discrimination of particular ethnic groups, or persons from getting employed. In Kenya, the profiling of the Somali people as harbingers of terrorists has made it difficult to process their national identification cards. This has led to them being denied their rights as bona fide citizens of Kenya. There is concern among human rights activists that AI profiling may have contributed to the rounding of Somali ethnic community members in Nairobi and subjecting them to inhuman and degrading treatment at the Kasarani Sports Complex in Nairobi (Amnesty International-Kenya, 2014; Gaumond & Régis, 2023; Olojo et al., 2020). This is similar to what has been reported in the USA (Etienne 2015). However, in the Kenyan case cited above, the role of AI in the profiling has not been clearly established, but if it was there, then it demonstrates the adverse effects of bias in the machine learning and training phase of the AI development. This shows that ethical considerations were not included into the development phase of the programs and therefore produces the bias in the decisions it generates. In the employment field, AI may lead to discriminatory practices.

The bias inherent to AI due to its design and consequences in the implementation of AI powered systems can be mitigated through various approaches. This calls for the need to ensure that that the data used for training the AI is representative of entire population frame of interest to avoid or mitigate the effects of bias in the way they function and arrive at decisions (Akello, 2022). In the business world, how people are accorded credit facilities should try as much as possible to develop a realistic view of their financial status which can be explained how the AI system made its judgement related to their credit applications. Continuous monitoring and evaluation of the AI system through audits may facilitate the identification of bias-related algorithmic programming and correcting them. This goes a long in improving issues of accountability on how the AI powered system works, giving a fairly good account of what is expected. The expected output of the AI must conform to the ethical values that are desired in the society at large. Finally, the use of methods that involve fairness-aware algorithms must be encourages. This proceeds on to the post-processing strategies which can be used in providing confidence that the AI is programmed to subject people to same and equal treatment based on the principle of equality and human dignity.

Every individual has their own personal space beyond which their own private affairs are not merged with the public arena. Privacy is a serious concern that needs to be addressed when developing the AI-powered systems ethically (Egan, 2022). The way this data is collected and used is raising concerns to various stakeholders both the legal fields and from other relevant stakeholders. Privacy is another critical ethical consideration in AI development. As the digital landscape continues to evolve, AI continues to advance fundamental and radical changes in the way work is done in various industries. This creates challenges to the potential dangers in may have on matters of data privacy and security where they are used to collect, manage, store and analyse vast amounts of personal data. This calls for robust development of data protection strategies to safeguard the privacy of various individuals.

However, with its immense potential come significant concerns regarding data privacy and security. As AI systems increasingly handle sensitive and personal information, ensuring robust data privacy and security measures becomes imperative. There has been a legal case in the European Court of Human Rights of Big Brother Watch et al. vs United Kingdom in which the privacy of the individuals was breached due to mass surveillance and collection of irrelevant data in a non-specific fashion (ECtHR 2014). This was a case where a program called TEMPORA, which was used to collect intelligence information by snooping into private emails of individuals in the fight against terrorism, and sharing of the same data with foreign parties at the exclusion of data subjects (Zalnieriute 2022). The ruling of the court was that such blanket surveillance was illegal and it infringed on the family and privacy rights of individuals. However, various people do know how their personal data is used by AI in financial institutions to arrive at the decisions regarding their loans. A case in point is where one does not understand how online lending platforms such as M-Shwari arrive at the amounts that they qualify to take as loans. Secondly explanations on why or why not where they advanced a credit facility or not is not communicated which may not augur well with ethicists from a moral point of view. This brings to fore the question of explainability. It is ethically wrong to make a decision about an individual based on some personal attributes without availing the reasons behind such a decision. In accordance with the respondents, the data subject has a right to know how AI come up with these decisions yet they have no way of knowing that since the online platforms do not offer a way of getting that explanation. This violates the principle of accountability and explainability.

The issue of data privacy and data manipulation is not a new phenomenon in Kenya, while there is no comparative to the Kenyan context so far of the case of Big Brother Watch et. al. v UK, the subject of manipulating private data by AI for sinister ends is also a question that has affected the political

landscape in Kenya. This is what is referred to as deepfakes, which is a form of misinformation where AI is used to create realistic but false scenarios of people saying or acting in ways they have never done (Njogu & Associate Advocates 2024). In previous general elections in Kenya, Cambridge Analytica was cited as having played a role in the manipulation of voters by capitalizing on collection of their private and personal data, and tailor making political messages such as deepfakes to change their perspectives of political concerns in one way or another unjustifiably (Crabtree, 2018). This brings the question of mass data security of individuals in the weaponization of AI according to the respondents. The respondents were of the view that AI technology may be used as a tool of propagating mistrust. In Kenya, it is illegal to process personal data of anybody without a valid legally acceptable reason, and such data should only be collected by the express permission of the data subject.

The security of personal data in the hands of data controllers and data users need to be obtained in an honest manner and used for purposes that are legal. Where such data may be needed for security issues, there must be a legal justification backed by law to ensure that the security of data of a personal nature is not infringed in disregard to the privacy of the individual. Questions of mass surveillance have shown that it adversely affects questions of data security regarding the individual. A case in point is the TEMPORA program, which was used by UK intelligence service to collect private data of individuals through electronic surveys by circumventing the legal requirements of doing so (Big Brother Watch and Others v. The United Kingdom, 2021). This is a breach of data security through the use of AI-powered systems. The ethical concerns regarding accountability were not inherently present in the process as it was meant to be opaque to the data subjects whose data was collected without their permission, neither was it done in accordance with the law. The respondents noted that this may be ongoing in Kenya where the regulatory policy and legal frameworks may not adequately protect individuals adequately against the State from prying into their private lives due to personal data safety concerns. The Kenya Data Protection Act of 2019 was cited as a case in point where the protection of individuals' rights can be violated because of loopholes in the law as so established.

The key function of data privacy protection issues is related to the safeguarding of all personal information related to any individual from unauthorized access and misuse. Software developers have a duty to develop applications that do not harm others. For example, using deontological ethical theories, it is morally wrong to create false impressions to mislead electorates by changing the information and perception of political personalities to unfairly influence voter decisions in the way they vote. This can be achieved through creation of deep fakes using AI technology. Deepfakes are created by exploiting private information of individuals and manipulating them to create impressions that appear real as a way of misinforming the public. Accordingly, in the question of AI, it is morally wrong to develop an AI application that is intended for purposes of deception. Humanity values the virtue of honesty, and therefore in line with deontological theories of ethics, it becomes imperative that the creators of AIs must do so by focusing on sustainability of virtuous acts as an end in the use of the AI product they create. It is from this perspective that immutability of personal data needs to be safeguarded as a way of countering the unethical use of personal data. AI-powered systems usually make decisions based on the data they have on individuals, and mutated data leads to incorrect decisions based on distorted data input. The matter of prejudice many not have been adequately addressed through safeguarding data immutability which leads to creation of deepfakes. It is the same principle that may be used to profile people without reference to legal requirements of protected groups such as sex or ethnicity (Wachter, 2022). This reinforces racial or gender stereotyping which adversely affects issues related to administration of justice, business relations, employment or affirmative action in Kenya. In relation to explainability, AI-powered systems do should be developed with the ethical principles of "accuracy, fairness, transparency and outcomes in AI-powered decision making" (Markus et al., 2021). This helps meet the ethical threshold of accountability in developing trust in the way the AI operates through an honest transparent manner which helps data users and subjects when deployed into productive purposes they were intended for. Explainability requires the task model to be based on a faithful and interpretable explanation of how decisions are arrived. This must be understandable and uses a

Personal data usually is constituted by personal and sensitive information which is inclusive of, but limited to financial details, biometrics, and medical data amongst others. This means that data users and controllers use of this information may raise a lot of issues that are touching on ethical factors. How this data is collected should in a transparent manner and for lawful purposes, the data must be used in accordance to the stipulated laws of a given country. The use of this data is premised on the platform of accountability and transparency which are ethical values that should guide the deployment of the AI tools, and their usage. Information related to the privacy of the individual is therefore at risk of privacy data breaches. Unsecured AI systems lure cyber hackers who try to access this information and illegally expose, use or make available the data to unauthorized entities. It should be noted that privacy concerns can be violated even when data is anonymized, and may result in financial losses to individuals, reputation risks, theft and other harmful practices that affect an individual adversely

The development of AI is intrinsically faced with the problem of securing the data all through the data life cycle. To ensure that data is safe, one step is to ensure that they are encrypted and only accessible to the right people with the right access codes. This ensures that the security of the data is enhanced especially when in transit either electronically or manually between various entities. AI models that that are based on machine-learning make predictions and forecasts in

consideration of the patterns they have been trained. This requires storage of large volumes of data which can be vulnerable unlawful access and manipulation. One of the solutions to this challenge is through federated machine learning, where data is stored in various locations, and only share insights only but restrict access to the data itself. Federated machine learning removes the need to move data between two user points (Wen et al., 2023). The ethical use of data requires development of AI systems that either eliminate unauthorized access to personal or private data and bar the inappropriate use of the same. Furthermore, in the development of the AI, it should be trained to collect data accurately of the specific individuals of interest to its functions, which is complete and only enough for that specific aim. The AI should not collect, utilize and process data that do not reflect its mandate in the objectives it was designed to achieve. Due to the reliance on the AI's output in decision-making, AI developers should ensure that the data within and at the disposal of the AI system is accessible to the data subject and owned by the same. The data subject should be allowed the right to exercise their rights in correcting the data, updating or inspecting the data should they have a need to do so.

Ethical considerations play a crucial role in addressing data privacy and security in AI. Organizations must ensure that AI systems are designed and deployed with privacy in mind, following principles such as data minimization, where only the necessary amount of data is collected and used. Additionally, transparency in AI operations is essential. Individuals should be informed about how their data is being used, and consent should be obtained where applicable. The role of governments in the field of data protection is to establish regulatory bodies, policies and regulations that are geared towards ensuring the safety pf personal data held by data users and controllers is safe. In Kenya, the Kenya Data Protection Act of 2019 is intended to protect the personal data of data subjects from infringement that may violate their rights (National Assembly 2019). The same is applicable to the European Union's General Data Protection Regulation that sets the standards and tight controls for data protection, providing the data subject with greater autonomy and control over their personal data. These regulations from various regions of the world are all geared towards improving data privacy and security in the deployment of AI powered systems.

Privacy issues should be infused into the design of the AI systems as standard practice. By doing so, it guarantees that the AI systems offer the best practices right from the initiation of their design, development and subsequent deployment upon product finalization. This allows for the integration of the data protection mechanisms and measures in the design and functioning of the AI as a product (Villegas-Ch & García-Ortiz, 2023). To achieve that, one is required to ensure that regular monitoring and evaluation of the AI system in providing a high guarantee that it complies with data protection and governance regulations. To enhance the question of trust, developers have to engage with the public, potential and current data subjects, data controllers and users of the data through constant checks and regular risk assessment, deploying strong encryption (van Daalen, 2023), and ensuring stringent controls exist in the management and control of access to the data. Ethically, the question of promoting issues of accountability and transparency is the designing, development and deployment of AI systems is important in Kenya (UNESCO, 2024). AI systems must articulate and project robust security checks and measures that are aligned to ethical principles, conform to regulatory stipulations while ensuring that the rights of people to privacy and confidentiality of personal information sustains trust in the AI systems and relevant technologies.

In many instances, the AI systems are dependent on personal data of an individual such as information related to their browsing history, health records and their general trends in social media engagements. Informed consent means that the data subject provides permission for their personal data to be collected and used for the purposes elaborated to them. The data users must be privy to the genuine reasons why the data is being collected, and defined parameters of who will have access rights to that information. This is critical in building transparent data practices that essentially helps in development of trust between the individual, the general public, the data user and/or controllers. It also helps is ascertaining the privacy rights of individuals is upheld within the entire product life cycle of the AI powered systems. It is important to establish and understand the fact that data minimization is a drive towards an ethical obligation of only requiring to collect and use data strictly that is necessary for the certified purposes for which it was collected. One is not required to collect excessive data that poses no significance for adds zero value to the tasks being performed. This also brings into mind the concept of surveillance and control. In this case, AI usage may raise the alarm by questioning the autonomy of the individual's autonomy. While facial surveillance is good for security purposes, it may be used pervasively by monitoring people over and above what is necessary leading to a sense of privacy intrusion to the personal space of a person. Therefore, it becomes imperative for AI developers to ensure that the product the develop and deploy does not have an overarching adverse function that is detrimental to the rights of other individuals.

While accountability and transparency has been discussed, it is important to point out that without these the element of trust in the AI systems will not develop by its end users, data controllers and data users and the individuals whose data is subjected to the AI systems and applications (Gichohi, 2024). By developing AI systems that are transparent and accountable, they portray AI developers who involve ethical issues in the products they create and deploy to the markets. They underpin the theme of responsibility on the part of the developers as a fundamental ethical concern when they are designing such applications and deploying them for consumption to various users. Transparency, in this case,

involves making the AI applications' functioning, processes and decisions it has arrived at understandable to relevant stakeholders' interest or affected by it. This has led to the emergence explainable AI concerned with creating models capable of providing simple but elaborate and interpretable reasons, or justifications, for the predictions they make and how they arrive at the decisions that informs relevant parties of their data processing outcomes.

Ethical concerns require elaborate and detailed documentation AI systems development. This will further aid in the evaluation and monitoring phases of product development and product usage. The records must document the data used and typologies, the algorithms that were used to develop and run the product and the mapping of how the algorithm makes decisions in order to enhance accountability and disclosure of this information to external audit and scrutiny. For purposes of developing trust, such documentation on the way the AY systems operate, how they were designed and the data processing it does provides an avenue for developing policies for the adherence to ethical standards as required for legal purposes and building confidence and trust in the systems.

6. CONCLUSION

In conclusion, the development and deployment of AI-systems accumulates a lot of benefits to the society and individuals. While this happens, numerous and diverse challenges have also emerged that pose ethical risk to the development and utilization of AI powered systems. It is crucial that attention be focused towards addressing issues of bias and fairness, accountability, privacy, transparency, legality and accountability in the use of the AI systems and their contribution to the society and to the individual. Ethical frameworks need to developed and followed in generation of sustainable practices, strategies and policies that guide the ethical development and use of AI systems. The continued development and evolution of AI technologies should encourage technological, social and ethical discourses and foster collaboration between developer, government agencies, stakeholders and AI users in leading to ethically AI solutions that are equitable, transparent and beneficial to everybody. It is only through this that ethical innovations of AI solutions can be realized for the common good of everybody

Ethical issues in DDD are centered squarely on principles of fairness, transparency, and accountability. Fairness is essential because there is a dangerous tendency, if unchecked, of AI systems to propagate discriminatory practices and prejudices in the decisions they arrive at. Ethical considerations in this case provide a high measure of assuredness that the AI systems do not drive forward such unethical vices. Instead, developers of AI should actively engage in seeking and isolating these biases in order to provide certainty their AI products are not tools for sustaining or creating new modes of discrimination in line with the UNDHR and the national constitutions of many countries. Accountability on the other is really essentially especially as such AI systems become increasingly autonomous in the way they arrive at decisions independently. In this, determining the persons who are actually responsible for the entirety or part of the decisions generated by AI becomes a serious problem because of the complexities involved. It is on this basis that the European Union came up with comprehensive guidelines that aim to establish accountability for the creators of AI systems, the users, data controllers and other stakeholders in situations where the Ais can make erroneous decisions.

7. RECOMMENDATIONS

The study gives an account of moral dilemmas emerging in the field of AI technology. Therefore, it recommends that addition of detailed case studies outlining actual ethical conundrums that are inherently present throughout the processes of AI development and deployment in order to improve them more. In addition to that, there is necessity of widening the range of the available remedies which was beyond the scope of the study and proposes research be done to improve knowledge in that area.

8. REFERENCES

[1] Akello, J. (2022). *Artificial Intelligence in Kenya*. Paradigm Initiative. https://paradigmhq.org/wp-content/uploads/2022/02/Artificial-Inteligence-in-Kenya-1.pdf

[2] Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, *13*(12), Article 12. https://doi.org/10.3390/app13127082

[3] Amnesty International-Kenya. (2014, May 23). *Kenya: Somalis scapegoated in counter-terror crackdown*. Amnesty International. Retrieved on 12/07/2024 from https://www.amnesty.org/en/latest/press-release/2014/05/kenya-somalis-scapegoated-counter-terror-crackdown/

[4] Bates, E. S. (2011). Counter-Terrorism in International Human Rights Law. In *Elizabeth Stubbins Gate, Richard Goldstone, & Eugine Cotran, Julia A. Hull, Juan E. Méndez, Javaid Rehman (eds) Terrorism and International Law: Accountability, Remedies, and Reform: A Report of the IBA Task Force on Terrorism* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199589180.003.0003

[5] Big Brother Watch and Others v. the United Kingdom, 58170/13, 62322/14, 24960/15 (ECtHR [GC] May 25, 2021). https://hudoc.echr.coe.int/fre?i=001-210077

[6] Big Brother Watch and Others v. the United Kingdom (Communiquée), 58170/13 (ECtHR January 7, 2014). https://hudoc.echr.coe.int/fre?i=001-140713

[7] Bohr, A., & Memarzadeh, K. (2020). Chapter 2—The rise of artificial intelligence in healthcare applications. In A. Bohr & K. Memarzadeh (Eds.), *Artificial Intelligence in Healthcare* (pp. 25–60). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00002-2

[8] Carrasco Ramírez, J., & Islam, M. M. (2024). Utilizing Artificial Intelligence in Real-World Applications. *Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023*, *2*, 14–19. https://doi.org/10.60087/jaigs.v2i1.p19

[9] Crabtree, J. (2018, March 23). *Here's how Cambridge Analytica played a dominant role in Kenya's chaotic 2017 elections*. CNBC. Retrieved from https://www.cnbc.com/2018/03/23/cambridge-analytica-and-its-role-in-kenya-2017-elections.html on 12/07/2024

[10] Dunsin, D., Ghanem, M. C., Ouazzane, K., & Vassilev, V. (2024). A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital forensics and incident response. *Forensic Science International: Digital Investigation*, *48*, 301675. https://doi.org/10.1016/j.fsidi.2023.301675

[11] Egan, M. (2022). Privacy boundaries in digital space: An exercise in responsibilisation. *Information & Communications Technology Law*, *31*(3), 301–318. https://doi.org/10.1080/13600834.2022.2097046

[12] Ejaz, U., & Olaoye, G. (2024). *Ethical Considerations in the Deployment and Regulation of Artificial Intelligence*.

[13] Gaumond, E., & Régis, C. (2023). *Assessing Impacts of AI on Human Rights: It's Not Solely About Privacy and Nondiscrimination*. LAWFARE. Retrieved from https://www.lawfaremedia.org/article/assessing-impacts-of-ai-on-human-rights-it-s-not-solely-about-privacy-and-nondiscrimination on 10/07/2024

[14] Gichohi, L. (2024). Kenya's Path to AI: Launch of Kenya's National AI Strategy Development Process. *KICTANet Think Tank*. Retrieved from https://www.kictanet.or.ke/kenyas-path-to-ai-launch-of-kenyas-national-ai-strategy-development-process/ on 11/07/2024

[15] Gogoll, J., Zuber, N., Kacianka, S., Greger, T., Pretschner, A., & Nida-Rümelin, J. (2021). Ethics in the Software Development Process: From Codes of Conduct to Ethical Deliberation. *Philosophy & Technology*, *34*(4), 1085–1108. https://doi.org/10.1007/s13347-021-00451-w

[16] Karmaza, O., Koroied, S., Makhinchuk, V., Strilko, V., & Iosypenko, S. (2021). Artificial intelligence in justice. *Linguistics and Culture Review*, *5*, 1413–1425. https://doi.org/10.21744/lingcure.v5nS4.1764

[17] Kumar, A. (2024). Exploring Ethical Considerations in AI-driven Autonomous Vehicles: Balancing Safety and Privacy. *Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023*, *2*, 125–138. https://doi.org/10.60087/jaigs.v2i1.p138

[18] Lane, L. (2022). Clarifying Human Rights Standards through Artificial Intelligence Initiatives. *International & Comparative Law Quarterly*, *71*(4), 915–944. https://doi.org/10.1017/S0020589322000380

[19] Loeb, S., Morris, P., Dynarski, S., McFarland, D., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers*. Institute of Education Sciences. https://files.eric.ed.gov/fulltext/ED573325.pdf

[20] Markus, A., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, *113*. https://doi.org/10.1016/j.jbi.2020.103655

[21] Njogu & Associate Advocates. (2024, February 22). *AI & Data Privacy Kenya: Opportunities,Threats & Regulations*. Accessed on 12/07/2024 from https://njoguassociates.com/ai-data-privacy-regulation-in-kenya/

[22] Olojo, A., Gumba, D. E. O., & Daghar, M. (2020). *Somalia, terrorism and Kenya's security dilemma*. ISS Africa. Retrieved from https://issafrica.org/iss-today/somalia-terrorism-and-kenyas-security-dilemma on 8/07/2024

[23] Ozone, T. (2019, May 8). Deontological AI Ethics. *Medium*. https://medium.com/@tim_ozone/deontological-ai-ethics-c8de98211497

[24] Sheludko, M. (2023). *Ethical Aspects of Artificial Intelligence: AI Ethics Explanation*. Software Development Blog. Retrieved from https://lasoft.org/blog/ethical-aspects-of-artificial-intelligence-challenges-and-imperatives/ on 13/07/2024

[25] Singhal, A., Neveditsin, N., Tanveer, H., & Mago, V. (2024). Toward Fairness, Accountability, Transparency, and Ethics in AI for Social Media and Health Care: Scoping Review. *JMIR Medical Informatics*, *12*(1), e50048. https://doi.org/10.2196/50048

[26] Sullivan, M. (2023). *Key principles for ethical AI development*. Transcend. Retrieved from https://transcend.io on 11/7/2024

[27] The Constitution of Kenya, Article 92 44 (2011). Accessed on 05/12/2023. https://www.undp.org/sites/g/files/zskgke326/files/

migration/ke/Popular-Version-of-the--Political-Parties-Act-Eng.pdf

[28] The Data Protection Act of 2019, Pub. L. No. 24 (2019). Retrieved from http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24of2019.pdf on 12/07/2017

[29] UNESCO. (2024, April 2). *Shaping Kenya's AI Future: UNESCO Contributes to National AI Strategy Formulation*. Retrieved from https://www.unesco.org/en/articles/shaping-kenyas-ai-future-unesco-contributes-national-ai-strategy-formulation on 13/07/2024

[30] United Nations. (2024). *Universal Declaration of Human Rights*. United Nations Bulletin; United Nations. Accessed on 27/05/2024 from https://www.un.org/en/about-us/universal-declaration-of-human-rights

[31] van Daalen, O. L. (2023). The right to encryption: Privacy as preventing unlawful access. *Computer Law & Security Review*, *49*, 105804. https://doi.org/10.1016/j.clsr.2023.105804

[32] Villegas-Ch, W., & García-Ortiz, J. (2023). Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. *Electronics*, *12*(18), Article 18. https://doi.org/10.3390/electronics12183786

[33] Wachter, S. (2022). *The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law* (SSRN Scholarly Paper 4099100). https://doi.org/10.2139/ssrn.4099100

[34] Waelen, R. (2022). Why AI Ethics Is a Critical Theory. *Philosophy & Technology*, *35*(1), 9. https://doi.org/10.1007/s13347-022-00507-5

[35] Wang'ang'a, A. W. (2024). Consequences of Artificial Intelligence on Teaching and Learning in Higher Education in Kenya: Literature Review. *East African Journal of Education Studies*, *7*(1), Article 1. https://doi.org/10.37284/eajes.7.1.1718

[36] Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: Challenges and applications. *International Journal of Machine Learning and Cybernetics*, *14*(2), 513–535. https://doi.org/10.1007/s13042-022-01647-y