A Multistakeholder Framework for Responsible AI in Educational Platforms:

Addressing the African Context

Report to funder:
The Centre for Intellectual Property and Information Technology Law (CIPIT), Strathmore University

Strathmore University

Centre for Intellectual Property and Information Technology Law

Acronyms & Abbreviations

AI	Artificial Intelligence
AIED	Artificial Intelligence in Education
HRIA	Human Rights Impact Assessments

Glossary

Algorithm – Instructions or rules to be followed by a computer during problem-solving or data-processing.

Bandwidth – The maximum amount of data transmitted over a wired or wireless internet connection over a certain period of time.

Chatbot – A computer application designed to simulate human text or speech during a conversation with a user, typically with the intention of providing answers to user queries.

Cyberattack – The unauthorised access of a computer system or network with the intention of compromising the integrity of the system or network.

Data Breach – The unauthorised disclosure or loss of personal information that had been stored digitally.

Telepresence – The use of technology to interact with users virtually regardless of their physical location.

User Interface – The point at which a user interacts with technology, typically through a display, keyboard, mouse or assistive tools.

Table of Contents

About this Report

This report presents a comprehensive framework for the responsible use of AI in educational platforms, specifically designed for the African context. Developed through collaboration with stakeholders across ten countries, the framework ensures diverse perspectives and addresses regional needs. It goes beyond high-level principles, offering concrete recommendations for stakeholders involved in designing and deploying responsible AI-powered platforms for education. The framework rests upon six core principles: human-centred approach, inclusiveness and fairness, safety & security, privacy & data protection, transparency and accountability. Each principle comes with tailored actions for different stakeholders.

Methodology

To achieve our research objectives, we employed a three-phase methodology.

In the first phase, we conducted a comprehensive review of existing research on digital learning platforms and how artificial intelligence (AI) is currently being used in education. This foundation of knowledge then informed the second phase, where we summarised a list of relevant principles and recommended actions for an AI-powered learning platform. Finally, the third phase focused on evaluating the suitability of the developed framework. To do so, we engaged with various stakeholders, including experts in digital learning platforms and AI, local actors in tech hubs, government officials, and education professionals.

Through joint multi-stakeholder dialogue sessions, we gathered valuable feedback to assess the framework's feasibility and practical effectiveness in the African context.

AI in Education

Artificial intelligence (AI) is significantly transforming the educational landscape, presenting both promising opportunities and challenges. Its applications extend across both academic and administrative domains, offering the potential to personalise learning, enhance accessibility, and improve administrative efficiency. However, it is crucial to address ethical concerns relating to data privacy, algorithmic bias, and equitable access.

On the academic side, AI-powered platforms have the potential to dynamically adapt to individual student needs and learning styles, delivering tailored lessons and assessments. This personalised approach may optimise learning outcomes by catering to individual strengths and weaknesses. Moreover, AI can bridge geographical and physical barriers through the use of tools like telepresence robots, granting access to education for students with disabilities or residing in remote locations. This augmented accessibility may significantly promote inclusivity within the educational system.

Apart from pedagogy, AI streamlines administrative processes at educational institutions. Automating tasks like grading assessments and scheduling classes can free up valuable time for teachers to focus on personalised interactions with students, resulting in a more efficient and effective learning environment for both educators and learners. Additionally, AI-powered learning analytics provide valuable insights into student progress, offering data-driven information that can inform more effective teaching strategies and optimise learning experiences.

Despite these promising benefits, implementing AI in education requires careful consideration of several ethical challenges. Safeguarding student data privacy is paramount and necessitates robust measures against unauthorised access and misuse. Furthermore, AI algorithms are susceptible to bias, potentially perpetuating existing societal inequalities if not rigorously monitored and mitigated. Additionally, ensuring equitable access to AI-powered education remains crucial, as resource limitations and infrastructure disparities could exacerbate existing educational divides.

Furthermore, effectively integrating AI into the educational ecosystem requires careful consideration of local contexts and diverse learning styles. Implementing generic solutions without adaptation may hinder successful implementation, emphasising the need for culturally appropriate and context-specific approaches.

The Framework

Overview

The goal of this project is to design a Responsible AI framework for the education sector with a focus on the African context.

The framework consists of two parts:

1. A set of principles selected by the multistakeholder community to govern the use of AI in education. The principles are drawn from AI ethics principles and education ethics principles and interpreted specifically for the use of AI in educational platforms.
2. Actions that stakeholders can take to operationalize the principles in various stages of the use of AI in education. For each principle, the framework recommends a set of actions that stakeholders can take to implement the principles.

Types of Stakeholders

This framework considers the unique needs and perspectives of three distinct stakeholder groups involved in AI-powered educational platforms: technology providers, regulators, and users.

1. Technology Providers

Examples of technology providers include technology companies, AI developers, platform designers, and educational app creators.

Technology providers can utilise the framework to guide responsible AI development, focusing on minimising risks of bias and ensuring user privacy.

Additionally, the framework can help them design Artificial Intelligence in Education (AIED) platforms that are accessible and inclusive for diverse learners.

2. Regulators

These can be government agencies, education authorities, and ethical oversight bodies.

Regulators can leverage the framework to establish clear guidelines and policies for responsible AIED development and deployment. This includes setting standards for transparency, accountability, and data protection within AIED platforms.

3. Users

Examples of users are educational institutions (schools, universities, etc), teachers, students, and learners of all ages.

Users can utilise the framework to understand their rights and responsibilities within AIED environments. This includes advocating for transparent and ethically designed platforms, questioning potential biases, and seeking appropriate support and guidance when interacting with AI systems.

By tailoring the framework to the specific needs and concerns of each stakeholder group, we can ensure its effectiveness in promoting and shaping responsible, human-centred AIED that benefits all.

The Framework: Principles

Principle of Human-Centred Approach

The principle of human-centred AIED demands that AI technologies in education enhance human capabilities while fostering fundamental rights and allowing for human oversight. This necessitates preserving human agency and control over learning processes, as they remain in control and at the centre of AI implementation.

Principle of Inclusiveness & Fairness

The principle of inclusiveness in AI-empowered educational platforms demands that these platforms be made accessible to all intended users, irrespective of potential barriers. This necessitates recognising and addressing the digital divide, encompassing disparities in infrastructure, equipment, skills, and contextual factors that can exclude vulnerable groups, such as refugees and communities lacking adequate technical resources.

AIED algorithms built on incomplete or skewed data or algorithms risk replicating and perpetuating harmful biases based on factors such as gender, race, ethnicity, and learning needs. The fairness principle for AI-empowered platforms is about not acting in a systematically prejudiced manner for ethnicity, geography, language, age, gender, religion, etc.

Principle of Safety & Security

Ensuring safety and security is critical in AIED platforms. From a security perspective, AIED platforms must be robust against cyberattacks, breaches, and threats. This includes identifying and addressing potential vulnerabilities to misuse, bias, and manipulation. Safety in AIED, on the other hand, focuses on protecting users from unintended harm. This includes safeguarding against unforeseen consequences like addiction, manipulation, or perpetuation of bias, while also ensuring the physical and psychological well-being of users.

Safety and security are inherently intertwined. Robust security measures directly contribute to safety, reducing the likelihood of data breaches that could expose users to harm. Conversely, a safe AIED platform minimises the potential for unintended consequences stemming from security vulnerabilities. This interconnectedness necessitates a proactive approach. Addressing potential vulnerabilities and harms throughout the entire lifecycle of AIED systems prevents security breaches and mitigates safety risks.

Principle of Privacy & Data Protection

Privacy and data protection emerged as a critical ethical concern in the implementation of AIED. Privacy, signifying freedom from unwarranted intrusion, and data protection or confidentiality, ensuring information security, jointly uphold the dignity, rights, and freedom of learners.

Firstly, informed consent is a cornerstone. AIED platforms must acquire well-informed user consent regarding data collection practices, both directly provided and passively gathered. Transparency plays a crucial role here, as users deserve clear explanations on how their data is used, stored, and protected.

Secondly, the confidentiality of user information is indispensable. All data, including both consented and non-consented revelations must be safeguarded and disposed of as per regulations. Moreover, comprehensive risk assessments and strict data governance standards are crucial to minimise potential harm.

The ethical implications extend beyond mere compliance. Developers and educators must actively cultivate trust by ensuring transparency and visibility into AIED functionalities and potential ramifications. Learners deserve to understand how AI affects their learning, careers, and social lives, empowering them to leverage skills while retaining control over their data and digital identities. From obtaining informed consent and guaranteeing confidentiality to fostering transparency and responsible data management, stakeholders must collaboratively safeguard learners' well-being in this increasingly data-driven educational landscape.

Principle of Transparency

Throughout the development and deployment lifecycle, transparency may uphold several meanings. One of the main focuses of this principle is to disclose when AI is being used, such as in a prediction, or recommendation, or when the user is directly interacting with an AI-powered agent, like a chatbot.

Transparency also involves data and algorithm transparency. First, data transparency demands clear communication about the type, volume, and origin of data used to train and refine AIED models. Users need to understand how their data is utilised and the extent of their control over its collection and use. Second, algorithm transparency requires explaining the logic and decision-making processes employed by AIED systems. Lastly, regulation and user awareness play crucial roles in ensuring transparency. This allows users to make informed choices, and to increase acceptance and trust in adopting such technologies.

Principle of Accountability

Achieving responsible AIED necessitates embedding the principle of accountability throughout its different stages. This requires acknowledging and assigning responsibility for the actions of each stakeholder involved in its design and use. Implementing clear regulations is crucial, explicitly addressing these responsibilities through transparent auditing, and establishing clear reporting mechanisms for unintended consequences, trade-offs, and potential harms.

The Framework: Actions

Having established the core principles for responsible AI-empowered platforms in education, the question arises: how do we translate these principles into tangible actions?

This section serves as a practical guide, offering a comprehensive collection of concrete steps that stakeholders can take to bring these principles to life.

Building upon the insightful questions raised during our dialogue sessions, we present a detailed matrix of recommended actions, tailored to each stakeholder group involved in the ecosystem.

Actions: Principle of Human-Centred Approach

Question

How can we ensure that AI-enabled learning platforms empower learners and educators, putting them at the centre of the educational experience in a way that is culturally and contextually relevant?

Recommended Actions

Regulators:

- Mandate human rights respect and due diligence: Enact regulations requiring AIED developers and deployers to conduct Human Rights Impact Assessments (HRIAs) and thorough human rights due diligence processes to identify and mitigate potential risks to learner autonomy and agency.
- Support "human in the loop" frameworks: Promote and fund research into, and the adoption of "human in the loop" design principles, ensuring human oversight and intervention in critical decision-making processes within AIED systems.

Technology providers:

- Filter and reduce coercive and manipulative automation: Design AIED systems that minimise automated features that could coerce or manipulate learners, focusing instead on supporting intrinsic motivation and identity development.
- Empower learner control: Integrate features that allow learners to control the type, frequency, and intensity of AI-driven support they receive.
- Prioritise transparency and explainability: Ensure AIED systems are transparent in their operation and reasoning, allowing learners to understand how decisions are made and challenge them if necessary.

Users:

- Integrate AI literacy programs: Develop and integrate AI literacy programs into curricula, equipping learners with the skills to understand, critique, and interact effectively with AIED systems.

Actions: Principle of Inclusiveness and Fairness

Question

What practical steps can be taken to make AI-enabled learning platforms more inclusive and accessible to learners with diverse backgrounds, abilities, and learning styles? Considering the diversity within the African educational landscape, how can we ensure that AI systems promote fairness and do not inadvertently perpetuate biases or inequalities?

Recommended actions

The cornerstone for this principle is the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access and representation through inclusive design and validation processes.

Regulators:

- Identify the infrastructure, equipment, and connectivity gaps within the population, particularly among vulnerable groups.
- Develop targeted programs: Bridge the digital divide by providing subsidised devices, internet access, and digital literacy training programs.
- Support teacher training: Equip educators with the skills to integrate AIED effectively, addressing diverse learning styles and mitigating potential biases.

Technology providers:

- Implement bias detection and mitigation mechanisms: Integrate algorithms within development processes to identify and address potential biases in data and models. Ensure that data used in development includes underrepresented groups
- Offer flexible access options: Design platforms with diverse access points, including offline functionalities and low-bandwidth alternatives.
- Partner with educators and experts: Collaborate with diverse educational stakeholders for ongoing feedback and guidance on inclusivity efforts and throughout the process from testing to needs identification and scaling stages.

Users:

- Provide feedback on existing platforms: Share experiences and suggest improvements to ensure AIED tools cater to the community needs and those of different learners.

Actions: Principle of Safety and Security

Question

How can we ensure that AI-enabled learning platforms do not cause avoidable harm, or unintended consequences, or pose unreasonable safety and security risks throughout their lifecycle?

Recommended actions

Regulators:

- Foster public awareness and education about AI safety and security in education.
- Establish clear guidelines for risk levels and safety testing procedures.
- Define clear standards for security and consent governance.
- Investigate any reported safety incidents.
- Review and update safety regulations as needed, adapting to evolving risks.

Technology providers:

- Invest in research and development of safer and more secure AI technologies for education.
- Conduct thorough risk assessments and prioritise potential situations when AI fails.
- Install safeguards to identify abnormal behaviour and prevent manipulation
- Conduct rigorous testing and validation to ensure the system functions safely and ethically.
- Develop mechanisms for user feedback and incident reporting.
- Continuously monitor system usage and performance, addressing safety issues promptly.
- Conduct risk assessments for new functionalities and scaled deployments.
- Maintain records of data characteristics, to ensure traceability when needed and analyse outcomes

Users:

- Participate in design discussions, highlighting potential risks from an educational perspective.
- Train direct users on responsible AIED usage and potential risks.
- Monitor student interactions with the system for safety concerns.
- Follow established safety guidelines and report any unexpected issues or concerns.

Actions: Principle of Privacy and Data Protection

Question

What robust data protection measures to consider in the context of AI in education, and how can we address concerns related to privacy and consent?

Recommended actions

All stakeholders:

- Regularly review and update internal policies and procedures to reflect evolving legal requirements and technological advancements.
- Develop training programs: Provide comprehensive training to employees on responsible data handling, data security best practices, and ethical considerations surrounding marginalised communities.

Regulators:

- Convene inclusive forums: Organise discussions with developers, user representatives, and experts from marginalised communities (including refugees) to understand their data concerns and guide technology development towards inclusivity and respect.
- Issue clear and adaptable guidelines: Develop flexible regulations that encourage responsible data practices while acknowledging evolving technological threats and allowing innovative solutions.
- Hold platforms accountable: Enforce compliance with data protection laws and actively investigate practices and incidents impacting marginalised communities.
- Support research and development: Fund research projects exploring privacy-enhancing technologies and responsible AI development for marginalised populations.

Developers:

- Participate in inclusive dialogues: Actively engage with diverse stakeholders, including refugees and their advocates, to understand their data privacy needs and incorporate their feedback into design processes.
- Implement privacy-by-design principles: Integrate privacy considerations from the outset of development, minimising data collection, implementing robust security measures, and offering clear control options to users.
- Conduct impact assessments: Regularly assess the potential impact of AI systems on marginalised communities, addressing biases and mitigating risks to privacy and confidentiality.
- Implement breach notification protocols: Establish clear procedures for promptly informing data subjects and relevant authorities in case of security incidents.
- Develop and deploy privacy-preserving technologies: Explore and implement techniques to protect data while enabling AI functions.

Users:

- Self-education: Seek information about data privacy practices of platforms, focusing on their handling of sensitive data.
- Be mindful of consent: Carefully review and selectively grant consent for data collection and use, understanding the potential consequences of each agreement.
- Utilise privacy settings: Take advantage of available platform settings to limit data sharing.
- Report suspicious activity: Raise concerns about potential data breaches or discriminatory practices to platform operators and relevant authorities.

Actions: Principle of Transparency

Question

How can transparency be achieved in the design and operation of AI algorithms within learning platforms, and how might this be communicated effectively to all stakeholders, including students and educators?

Recommended actions

Regulators:

- Establish transparency standards: Mandate clear labelling and disclosure of AI use in educational platforms.
- Define performance communication: Set guidelines for platforms to inform users about the AI system's capabilities and limitations.

Technology providers:

- Integrate clear labelling: Design user interfaces that explicitly indicate when users are interacting with an AI tool, chatbot, or human. Another example would be to be explicit in distinguishing between recommendation and information.
- Communicate performance: Provide easily accessible information about the AI system's capabilities and limitations, using metrics like accuracy rates and error margins. Do testing in realistic conditions to verify how the information is relayed.

Users:

- Select transparent platforms: Choose AIED platforms that prioritise clear labelling, performance communication, and user control over data.
- Monitor transparency practices: Regularly assess AIED platforms used within the institution to ensure adherence to transparency principles.

Actions: Principle of Accountability

Question

What are the key aspects of accountability that should be considered when it comes to the development and deployment of AI in education?

Recommended Actions

Regulators:

- Proactively create frameworks for clear accountability mechanisms, liability concerns, and regular audits to ensure responsible development and deployment of AI in education.

Technology providers:

- Ensure auditing processes that are up to date.
- Provide evidence for engineering best practices.

Users:

- Implement mechanisms for users to report concerns, feedback, and incidents related to AI system usage, ensuring voices are heard and addressed.
- Keep a comprehensive record of the use of the platform and data governance practices.

Conclusion

In this report, we have presented a comprehensive framework that governs the responsible use of AI in educational platforms. The framework was developed through a collaborative multistakeholder approach involving representatives from ten African countries.

This inclusive approach ensures that the framework reflects a diverse range of perspectives and considers the specific needs and challenges of educational systems in different contexts. Moreover, this framework goes beyond high-level principles and guidelines, providing concrete recommendations for actions on the design and development of responsible AI-powered platforms for education.

This study was made possible by a grant provided by the International Development Research Center (IDRC). We thank the organisation for their continued support.

IDRC · CRDI

International Development Research Centre
Centre de recherches pour le développement international

Canada

ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT AFRICA