TrustedTech Africa Ltd | Sprite+

AI TRUST AND SAFETY IN EMERGING TECHNOLOGY ROUNDTABLE 2026:

Key Perspectives, Risks, And Strategic Directions

2026 REPORT

www.trustedtechafrica.com

01 Executive Summary

This report captures insights from a multi-stakeholder discussion on Artificial Intelligence (AI) in Africa, focusing on trust, safety, governance, and emerging risks.

Participants from diverse sectors including law, policy, cybersecurity, child protection, technology, and civil society highlighted critical gaps in AI awareness, regulation, and ethical deployment across the continent.

The discussion emphasizes that while AI presents transformative opportunities, Africa must address foundational challenges such as literacy, governance frameworks, data protection, and societal impact to ensure responsible adoption.

02 Introduction

Artificial Intelligence is rapidly shaping global economies, governance systems, and social structures. In Africa, however, the discourse is still evolving, with varying levels of understanding and implementation across countries.

THIS SESSION BROUGHT TOGETHER PROFESSIONALS TO:

- Share perspectives on AI risks
- Identify under-discussed challenges
- Explore what AI trust and safety should mean within the African context

PARTICIPANT OVERVIEW

The discussion included professionals from:

- Technology and AI ecosystems
- Cybersecurity and data protection
- Legal and policy sectors
- Child protection and digital safety organizations
- Civil society and NGOs
- AI governance and strategy development

This diversity provided a multi-dimensional view of AI risks and opportunities.

03 Key Themes And Discussion Areas

UNDER-DISCUSSED AI RISK AND CHALLENGES IN AFRICA

1. Limited Understanding of AI Beyond Chatbots

A recurring concern was the misconception of AI as merely chatbots or generative tools. Many users lack awareness of AI as a broader system influencing decision making, automation, and analytics. This limits adoption, innovation, and risk awareness

2. AI Literacy Gap

Participants emphasized that AI literacy remains critically low: Users do not understand how AI systems work, there is poor prompting and misuse of tools is widespread. The lack of education leads to misinformation and blind trust in AI outputs

3. Data Privacy and Security Risk

A major concern raised was unsafe data practices: Individuals and organizations upload sensitive data into AI tools and Lack of awareness about cloud storage and data reuse risks has potential for data breaches, reverse engineering of sensitive information and violations of privacy regulations

4. “Shadow AI” in Organizations

An emerging risk identified was unauthorized use of AI tools within organizations:

- Employees using unapproved AI systems
- Processing confidential data through external platforms
- Lack of governance exposing organizations to security breaches and compliance violations

5. Misinformation, Deepfakes, and Synthetic Content

Participants highlighted the growing threat of:

- AI-generated misinformation
- Deepfake images and videos
- Political manipulation and election interference These risks are particularly dangerous in regions with:
- Low digital literacy
- High social media consumption

6. Child Protection and AI Abuse

A critical and sensitive issue discussed was AI-generated abusive content involving children and the difficulty distinguishing real vs. AI-generated material. This increased exposure to harmful content online raises urgent concerns for law enforcement and policy frameworks

7. Lack of AI Governance Frameworks

One of the most significant gaps identified was that only a limited number of African countries have AI strategies and even fewer have actionable governance frameworks, this raises the challenges of :

- Fragmented approaches across countries
- Lack of enforcement mechanisms
- Over-reliance on global frameworks not tailored to African realities

8. Bias and Data Representation Issues

AI systems often reflect:

- Biased datasets
- Underrepresentation of African contexts

This leads to:

- Inaccurate outputs
- Reinforcement of inequalities

9. Mental Health and Psychological Impact

Participants noted:

- Over-reliance on AI for decision-making
- Emotional and psychological effects of AI interactions
- Exposure to harmful or misleading content

10. Environmental Impact of AI

An often overlooked issue raised:

- AI infrastructure requires: Significant energy , water resources and Data center capacity

This poses challenges in regions already facing:

- Energy shortages
- Limited infrastructure
- Lack of clean water

Collective Framing of AI Trust And Safety In The African Context

Participants defined AI Trust and Safety through multiple lenses:

1. Prevention of Harm : Protecting the most vulnerable groups (children, women, marginalized communities) and ensuring AI systems do not cause direct or indirect harm

2. Transparency and Accountability:Users should understand how AI works, how decisions are made and organizations must be accountable for AI outputs.

3. Education and Awareness: AI safety begins with user education and digital literacy is foundational to trust

4. Regulation and Governance : The regulations and policies of AI need to be clear and enforceable. There needs to be a regional collaboration in the document of this governance framework.

5. Ethical Development and Deployment : AI systems must be fair, inclusive and context-aware

INTERACTIVE SAFETYMETER SHOWCASE

During the session, the Trusted Tech team conducted a live demonstration of the Safety Meter, showcasing its functionality and practical application in identifying and mitigating potential risks in technology products prior to deployment. Participants were then invited to share feedback on the tool's relevance and usability for developers and tech founders.

Overall, the feedback was overwhelmingly positive, with participants highlighting the Safety Meter as an essential resource for proactively addressing safety, trust, and risk considerations before product launch. Many noted its potential to strengthen responsible innovation practices within the ecosystem.

Notably, representatives from developer-focused organizations such as HTC Academy and the Google Developers Group Abuja expressed strong interest in the tool. They requested follow-up engagements with the Trusted Tech team to facilitate dedicated demonstrations and training sessions for their developer communities, underscoring the growing demand for practical safety tools in the tech development lifecycle.

Deep Dive Risk and Responsibilities

During the discussion and deep-dive session, participants were asked to reflect on a critical question: when AI-related harms occur in African markets, who bears responsibility — builders, investors, civil society, users, or government? This question sparked a robust and nuanced debate, with a particular focus on the roles of builders and investors.

A significant perspective from participants emphasized the role of investors, noting that as primary funders of technological products, they hold considerable influence over what gets built and how it is scaled. Some argued that investors often shape product direction, sometimes prioritizing growth and returns over safety considerations. In such cases, builders may have limited power to fully implement their preferred safety standards, as they must balance ethical responsibilities with financial pressures and expectations tied to funding.

Conversely, another group of participants maintained that builders should bear primary responsibility, as they are the creators of the technology and ultimately decide what is designed, developed, and deployed. From this viewpoint, investors are seen as enablers who act based on what is presented to them, rather than direct drivers of harm.

Beyond the builder–investor dynamic, participants also highlighted the critical role of government and policy frameworks in shaping accountability. It was noted that effective regulation can help mitigate unintended harms by setting standards and boundaries for both builders and investors. Additionally, civil society and users were acknowledged as important stakeholders in identifying, reporting, and responding to harms.

A key theme that emerged from the discussion was the distinction between responsibility and accountability. Participants explored how responsibility for AI harms may be distributed across multiple actors, while accountability mechanisms such as regulation, governance structures, and enforcement determine who is ultimately held answerable. Overall, the debate underscored the complexity of assigning responsibility within the AI ecosystem and the need for a more collaborative, multi-stakeholder approach to ensuring safer technology outcomes.

04 Key Takeaways

In closing, participants were invited to reflect and share one key takeaway from the session. They were encouraged to identify either an assumption that had shifted, a risk they now perceive differently, or a concrete commitment they intend to make toward ensuring that AI systems deployed across Africa over the next five years reflect the depth, responsibility, and solution-oriented nature of the discussions held. These are the key takeaways:

- **AI Literacy is Foundational :**
  Without widespread education, users will continue to misuse AI tools and remain vulnerable to risks.
- **Data Protection is a Critical Risk Area :**
  Organizations and individuals must adopt safer data practices when using AI tools.
- **Vulnerable Groups Require Strong Protection :**
  Children and marginalized populations face disproportionate risks in AI systems.
- **Collaboration is Essential:**
  Stakeholders across sectors must work together to create context-specific AI solutions for Africa.
- **Governance Must Move Beyond Strategy:**
  Africa must transition from policy discussions to actionable frameworks and enforcement.
- **Misinformation is a Growing Threat:**
  AI-generated content poses serious risks to public trust, elections, and societal stability.
- **AI Risks are Multi-Dimensional:**
  From cybersecurity to mental health to environmental impact, AI risks are interconnected.

05 Next Steps

Following the event, the TrustedTech Africa team identified several strategic next steps to sustain momentum and deepen impact. A key priority is to institutionalize the roundtable as a recurring engagement either on a quarterly or annual basis bringing together a broader and more diverse group of stakeholders to build dialogue and advance practical solutions around AI trust and safety.

In addition, the team is exploring collaboration opportunities with AI Safety Nigeria, particularly by integrating AI Trust and Safety discussions into their “Launch and Learn” series. This partnership presents an opportunity to reach wider technical and policy-focused audiences; however, its successful implementation will require dedicated funding and resource support.

The TrustedTech team also plans to work closely with developer training organizations that participated in the event to deliver targeted capacity-building sessions. These trainings will focus on equipping developers with the knowledge and tools to integrate AI safety principles into the design and development of their products. A key component of this effort will be the introduction of the SafetyMeter as a practical framework to support ideation, risk assessment, and responsible product development from the early stages.

Furthermore, the team intends to initiate collaborative technical research with select partner organizations represented at the event. This research will aim to generate context-specific insights and contribute to the growing body of knowledge on AI safety and governance within African markets.

06 Conclusion

AI presents a powerful opportunity for Africa's growth and innovation. However, without intentional governance, education, and ethical safeguards, these technologies may deepen existing inequalities and introduce new risks. To move forward effectively, Africa must:

- Build strong AI literacy systems
- Develop robust governance frameworks
- Prioritize trust, safety, and inclusivity

Only then can AI truly serve as a tool for sustainable and equitable development across the continent.