

Meeting summary

Quick recap

This meeting was a panel discussion on "Can AI-generated evaluation findings be trusted?" as part of Global Evaluation Week 2026, organized by the Prime Minister's Office and the Tanzania Evaluation Association. The session featured keynote speaker Dr. James Yonas, Permanent Secretary at the Prime Minister's Office, who emphasized that while AI offers significant opportunities for improving evaluation efficiency, trust in AI-generated findings depends on transparency, data quality, human oversight, and ethical standards. The panel included presenters Ms. Akina Munimku, Mr. Pantaleon Shoki, and Mr. Baraka Filingue who discussed the conditional nature of trusting AI findings, the importance of proper governance frameworks, and practical tools for assessing AI-generated evaluation data. Participants actively engaged through chat responses, expressing that AI can improve evaluation quality by identifying patterns, enhancing analytical depth, and supporting faster data analysis, though they emphasized the need for human oversight and professional judgment to ensure trustworthiness.

Next steps

Baraka

- [Share practical guidance and examples of combining free/paid tools \(e.g., Kobo Toolbox, Zotero, ChatGPT\) for data collection, analysis, and reporting in evaluation, especially for teams with limited resources.](#)

Pantaleon

- [Organize further training sessions in collaboration with the Prime Minister's Office on responsible use of AI in evaluation for TANEA members and interested participants.](#)

Sakina

- [Promote and support the integration of national and international AI and evaluation standards \(e.g., OECD DAC, UNESCO, National Digital Economy Strategic Framework\) within the Performance Monitoring and Evaluation Division and broader evaluation community.](#)

Collaboration

- [All presenters \(Pantaleon, Baraka, Sakina\): Develop and disseminate documentation templates and best practices for transparent reporting of AI-assisted evaluation processes, including version control and clear methodology disclosure.](#)

- [TANEA \(Tanzania Evaluation Association\): Continue to update members on new frameworks, standards, and tools for trustworthy AI use in evaluation, and facilitate knowledge sharing among evaluators.](#)
- [All evaluation teams: Regularly review and update their evaluation workflows to include human-in-the-loop validation, source citation, and auditability for all AI-generated findings, as per the recommendations discussed.](#)
- [All participants: Share experiences and lessons learned regarding the use of AI in evaluation, including challenges and solutions, in upcoming TANEA forums and training sessions.](#)

Summary

Technical Meeting Connectivity Issues

The meeting experienced significant technical difficulties with audio and connectivity issues throughout the session. Baraka attempted to manage the technical problems and informed participants that the meeting would be recorded with their consent. Despite multiple attempts to address audio issues with participants like Ritha and Sakina, the meeting remained disrupted with ongoing technical challenges.

AI in Evaluation Findings Discussion

SAKINA opened the session on AI-generated evaluation findings as part of Global Evaluation Week, welcoming participants and introducing the keynote speaker, Dr. Jim Yonas, Permanent Secretary at the Prime Minister's Office in Tanzania. Dr. Yonas emphasized that while AI presents significant opportunities for strengthening evaluation processes, trust in AI-generated findings depends on transparency, data quality, human oversight, and adherence to ethical standards. SAKINA and other speakers discussed the importance of maintaining rigorous evaluation standards and using AI as a tool to enhance rather than replace human judgment and professional expertise.

AI Evaluation Trustworthiness Discussion

Pantaleon presented on the trustworthiness of AI-generated evaluation findings, answering that it is conditional based on various factors including ethical principles, international frameworks, and quality assurance practices. He outlined key ethical principles for AI-based evaluations including fairness, transparency, accountability, and privacy, and discussed international frameworks from organizations like OECD and UNESCO that provide guidance for AI implementation. Pantaleon concluded that while AI-generated findings can be trusted with proper implementation of these frameworks and standards, evaluators must remain cautious of existing gaps and infrastructure challenges in their approach.

AI Governance and Evaluation Standards

Pantaleon emphasized the importance of integrating AI governance principles, stakeholder engagement, and quality assurance in AI-based evaluation practices. Baraka highlighted practical tools for assessing trustworthiness, including continuous observability systems, evaluation frameworks, feedback and attribution tracking, and benchmarking, while stressing that AI outputs should not be accepted without human review and verification. The discussion also covered documentation and transparency standards, the OECD AI principles, and the need to balance AI tool usage with human judgment to ensure responsible and trustworthy evaluation findings.

AI in Evaluation Processes Discussion

The meeting focused on discussing the role and implementation of AI in evaluation processes. Participants shared their views on whether AI can improve evaluation quality, with most agreeing that AI can enhance efficiency and data analysis but should be used as a supportive tool rather than a replacement for human judgment and expertise. Pantaleon emphasized the importance of using AI intelligently with human oversight, following established guidelines, and understanding the context to ensure trust in AI-generated findings. The discussion also highlighted the need for professional skills, ethical standards, and transparency in using AI, with participants suggesting that a multi-layered governance stack combining evaluation standards and AI guidelines could help address these challenges.

AI in Evaluation Tools Discussion

Baraka discussed how evaluation teams can effectively use a combination of free and paid tools like Kobo Toolbox, Google Forms, and Zotero to conduct evaluations despite limited resources. The discussion addressed donor and government perceptions of AI-generated evaluation findings, with Baraka explaining that while these stakeholders appreciate the speed and analytical support AI offers, they remain concerned about evidence quality, transparency, and bias. Pantaleon concluded by emphasizing the importance of digital literacy and AI knowledge in Tanzania's long-term vision, announcing plans for future training sessions through the Tanzania Evaluation Association.

AI in Evaluation Practices

The meeting focused on the role of AI in evaluation practices, with Baraka emphasizing that while AI can support evaluation, it should complement human judgment rather than replace it. Key factors for trustworthy AI in evaluation were identified as data quality, transparency, ethics, human oversight, and accountability. Ritha concluded the session by thanking all participants and presenters, highlighting the importance of using AI responsibly to strengthen evidence-based decision-making.

[View in Zoom](#)

