Q&A: AI in Evaluation – From Interviews to Insight

gLOCAL Evaluation Week 2026

Sessions:

AI in qualitative data collection: Field experiment *by Ali Safarnejad & Daniel Alonso Valckx*

AI for Social Listening in Evaluation *by Juan Pablo Arguello Yepez*

AI-Enabled Equity Analysis *by Atishay Mathur*

Moderated *by Jane Mwangi*

Q: Will the presentation slides be made available after the session?

A: Yes — the slides are available on the gLOCAL Evaluation Week page.

Q: How did participants react to being interviewed by an AI? I sometimes find it hard enough to get people to sit for an interview with a person!

A: Reaction was genuinely positive across participant types — the term that came up repeatedly was that it felt natural, like a WhatsApp conversation. Two factors seem to drive this. First, the asynchronous format: participants can start the conversation, step away, and return at their own pace, which reduces the coordination burden that makes scheduling human interviews so difficult. Second, the total time participants spent was typically less than with a human interview. There is an important caveat though: because AI-assisted interviewing is still novel, some participants may be spending more time with it out of curiosity — so we'd want to watch whether engagement sustains once the novelty wears off.

Q: How was the AI-interviewer tool trained to handle reticent or difficult-to-engage participants? Are there rules to follow up with a respondent later or replace them with another?

A: The tool uses adaptive probing logic — rather than moving on after a brief or non-committal response, the agent is configured to use follow-up prompts that encourage

elaboration ("could you tell me more about that?" or "can you give me a specific example?"). On replacement of respondents: the tool does not have an automated substitution rule, but the link to the interview remains active even if a participant drops off, with the conversation history remaining historized, and reload in the chat history box, allowing them to resume where they left off. The configuration of follow-up protocols is ultimately determined by the evaluation design and ethical governance framework in place for each deployment.

Q: What AI platform was used to develop the AI-Interviewer tool?

A: The platform is Google Cloud Platform (GCP), built with Python. The underlying models are Google Gemini 2.5 — both fine-tuned and prompted versions. A link to a gLOCAL talk from last year with a fuller architectural walkthrough is available here: https://www.youtube.com/watch?v=1SUMiHIpof8 (starting at 00h20m37s)

Q: Is the AI-Interviewer open source? Can anyone outside of UNICEF use it?

A: The project is currently open in the sense that we are actively seeking to improve and adapt it across different contexts. It is not a downloadable open-source package, but if you have a specific use case in mind, please reach out directly. Where there is mutual interest, we can set up a dedicated instance — similar to what was done for the Maldives TechPath case study — so you can experiment with it in your own setting. Contact details were shared with all registered participants; for anyone viewing this Q&A, you can reach Ali Safarnejad on Linkedin: https://www.linkedin.com/in/ali-safarnejad-phd-a5097b11/

Q: Were participants reading and typing their responses, rather than speaking and having audio recorded during the AI-Interviews?

A: Yes, for this pilot phase participants interacted through a text-based chat interface — reading questions and typing responses. There is a plan to develop a digital avatar interface with audio/video interaction and automated transcription, which would make the experience closer to a spoken interview.

Q: How does the Social and Community Listening (SCL) tool intersect with consent for data processing? What ethical governance is in place?

A: Any evaluation or evidence-generation activity involving AI data collection — whether surveys, social listening, or otherwise — needs to operate within a coherent governance framework. For UNICEF's work, this means compliance with UNICEF's Data Policy and the ethical standards for evidence generation. For activities involving primary data collection from individuals, an independent Institutional Review Board (IRB) review is required. Consent design is part of that review — ensuring that participants understand what data is being collected, how it will be used, and by whom.

Q: Regarding the plan to develop child-friendly AI interfaces — how will UNICEF manage the risk of anthropomorphism? Children can easily develop emotional bonds with friendly digital avatars. What guardrails are in place to prevent relational deception or over-sharing during data collection?

A: That's an important consideration. Any evaluation or evidence-generation activity involving children—whether AI is used or not—must comply with UNICEF's policy on ethics in evidence generation and undergo independent review by an accredited Institutional Review Board (IRB). The IRB process specifically examines potential risks to children, including issues such as undue influence, over-disclosure, etc. So rather than treating anthropomorphism as a purely technical issue, it would be addressed through UNICEF's established ethical review framework which applies the governing principle of "do no harm". In this way, the data collection tools are independently scrutinized and refined before deployment with children.

Q: Have you thought not only about what AI *could* be used for, but what it *should* and *should not* be used for?

A: This question sits at the heart of responsible AI integration in evaluation. Some principles that apply include using AI where it adds genuine value that would otherwise be inaccessible, for example to strengthen scale, consistency, and speed of analysis, rather than as a default replacement for evaluator judgment. There are roles where AI will be unsuitable, for instance, in highly sensitive contexts requiring active trust-building, or situations where the probabilistic nature of LLMs outputs is an unacceptable risk or contexts where participants have limited agency, low digital literacy, limited access to technology, or discomfort engaging with digital tools, which could compromise inclusiveness, informed participation, or data quality. In this regard, as the session presentations demonstrated, the "human in the loop" is not just a technical safeguard, but rather it reflects the epistemological reality that AI output without human interpretation is

data, not intelligence. So the answer to the question is not a fixed list of dos and don'ts — it is an ongoing, context-specific ethical deliberation, supported by governance structures like UNICEF's policies on AI and ethics in evidence generation, and the independent ethical review process.

Q: Is the SCL tool available for use by organizations outside of UNICEF?

A: We have not yet worked with external organizations, but it is possible. If you are interested in exploring how the tool could be applied in your context, please reach out using the contact information was shared with registered participants.

Q: Did any of your tests try to estimate the environmental impact of using these tools?

A: The piloting phases presented in this session have primarily focused on operational feasibility, user experience, and the quality of equity-integration. The environmental impact of running LLMs, while an important consideration, was not assessed. As we move from the small scale experimentation toward wider integration and scale, measuring and mitigating the carbon footprint of these tools will need to be incorporated into our broader AI governance and ethics frameworks.

Q: Given that multilateral agencies require rigorous procurement and ethical review processes, there is often an inherent time lag. How does UNICEF ensure that AI tools don't become obsolete by the time they are formally approved — given that AI evolves on 6–12 month cycles?

A: This tension is real and reflects a broader challenge facing many organizations working with emerging technologies. However, the question should be asked not in how quickly one can adopt the latest AI tools, but rather how can we do so responsibly and effectively. Our primary obligation should be to ensure that any use of AI is safe, ethical and aligned with the principle of ‘do no harm’. In this context, taking the time to conduct appropriate ethical reviews, risk assessments and governance processes is not a limitation, it is safeguard.