



| |
|--|
| SESSION NUMBER & TITLE |
| Parallel session 15: AI-powered evaluation: Maximizing efficiency while minimizing risks |
| SESSION TAKE AWAY |
| The use of artificial intelligence (AI) can fast track the delivery of the SDGs by providing innovative and efficient solutions to address global challenges. While AI has numerous advantages, it also has significant risks that need to be addressed. Evaluation can play a crucial role in mitigating AI-related risks by providing a framework to assess and validate AI systems. Furthermore, incorporating AI in various stages of an evaluation can revolutionize the evaluation process by enhancing efficiency gains, improve evidence extraction and synthesis, enable real time evaluation by supporting data collection and timely feedback and analysis, and enhance evaluation communication for greater utilization of evaluation. The opportunities in powering evaluation with AI, also come with several risks and pitfalls which should be identified and mitigated. Keeping in mind the various opportunities AI presents for evaluation, as well as the risks, the participants will learn how to leverage AI in evaluation in a balanced way. |
| SUBTHEME |
| Innovations in evaluation |
| ORGANIZATION/S |
| United Nations Population Fund (UNFPA) |
| LEAD PRESENTER: [Marco Segone, UNFPA] |
| KEY MESSAGES |
| <ol style="list-style-type: none"> 1. The following approaches need to be considered to reduce harms and mitigate potential risks of AI: (i) robust testing and validation; (ii) data quality and bias; (iii) explainability and transparency; (iv) regular monitoring and maintenance; (v) ethical guidelines and regulations; (vi) human must have a complete oversight and control; (vii) public awareness and education; (viii) collaboration and knowledge sharing; (ix) international conventions and regulations; (x) robust cybersecurity measures; (xi) enhance digital and AI literacy to combat digital divide. 2. Human rights, transparency and accountability must light the way, including for generative AI (GenAI)-powered evaluation. We must collectively build essential guardrails for the deployment of AI in evaluation. In order to do so, there should be (i) optimistic outlook for AI-empowered evaluation; (ii) responsible and ethical GenAI; and (iii) partnering and sharing experiences for optimization. |
| PANELIST/DISCUSSANT #1: [Uyen Kim Huynh, Innovation Specialist, Evaluation Office, UNICEF] |
| KEY MESSAGES |
| <ol style="list-style-type: none"> 1. GenAI has the capacity that is proliferating quickly. Wisdom from human would discern which knowledge is important. 2. UNICEF Policy on AI guidance for children in 2020 and was revised in 2021. Having low resources for traditional evaluation tools is where UNICEF explore AI. Insights could be drawn from big data. UNICEF has an Internal database of more or less 2,000 reports. UNICEF has started to identify which interventions to end child marriage per country. They have piloted |

machine learning techniques to synthesize reports; construct baseline and evidence from specific interventions on violence against children; and scrape twitter data. GenAI can process and automate data collection and have it ready for analysis. It also helps fast track a lot of data needs for baselines and counterfactual values, which would improve planning, analytics and help evaluators write better.

PANELIST/DISCUSSANT #2: [Martin Prowse, Evaluation Specialist, Independent Evaluation Office, GCF]

1. There is a growing consensus on how to use GenAI in evaluation. Output from GenAI is just one source of data. We, as evaluators, have developed skills for checks and balances, which could also be applied on GenAI.
2. Climate projects expands in various regions and sectors. These interventions could be categorized in two: mitigation and adaptation. Adaptation projects are more complicated than mitigation projects. There is a need to move to the future tense, where interconnectiveness of climate issues and sustainable development have to factored in. The question of “how adaptation is measured?” is where GenAI can come in as there is no metrics for adaptation yet, and no scales of indicators. GenAI would help us synthesize data.

MODERATOR: [Maya Vijayaraghavan, Principal Evaluation Specialist, IED, ADB]

KEY MESSAGES

1. Everyone, including evaluators, should use GenAI wisely and build the capacity to prompt.
2. It is important to manage risks and pitfalls.
3. We should maximize the tool to efficiently perform evaluation work.

Q&A SUMMARY

How does UNICEF plan to address data privacy and security while using generative ai?

There should be paramount emphasis on security and privacy in data management. It has to adhere to ethical guidance and practices that protect children’s data. Use ethical consents when collecting data from stakeholders... international standards... ai strategy – 1) creation of data technologies that would eliminate biases; 2) highly dencentralized organization – ict and ethics departments are consolidating good practices that would help regulate the use of ai; 3) continue advocating open-ai solutions; 4) staff upskilling for digital and AI literacy. Governance – regulatory framework is important, much better digital literacy, emphasizing risks and limitations.

Leveraging AI benefits, how do you navigate limitations with stakeholders?

Estimating cost in adapting to climate change is difficult. Evidence and systematic reviews use score cards to assess confidence level. It can be used on generative AIs.

What is the difference between GenAI and responsible GenAI

Closed system of database/AI tools protects data which google cannot access ensures control. cost is considered. However, trade-off includes quality assurance of the results.

Marco: Evaluation is about building relationship and trust with stakeholders instead of just providing figures. AI will potentially free us up on more fundamental aspects of evaluation.

Uyen: AI will do whatever you prompt to do. AI should be a thought partner. We can work smarter and see AI as netscape of the internet.

Martin: Evaluators should keep respondents' information confidential.

SESSION COORDINATOR

Bryan Noel Lazaro