

Session Summary

PARALLEL SESSION 9: AI & ML IN ACTION: REVOLUTIONIZING EFFICIENCY AND INNOVATION IN EVALUATION 3 September 2024, 3:30 – 5:00 PM

SESSION TAKE AWAY

How can artificial intelligence (AI) and machine learning (ML) transform the field of evaluation? Find out in this session at the 2024 Asian Evaluation Week, where we will explore cutting-edge initiatives from the Asian Development Bank (ADB), the European Bank for Reconstruction and Development (EBRD), and the World Bank (WB). You will learn how these institutions are using advanced language models and machine learning to increase the efficiency of processing evaluation reports and improve access to valuable evaluation knowledge for accountability and learning.

SUBTHEME

Tech-Driven Evaluation: Navigating Development Impact in the Digital Era

PANELIST/DISCUSSANT #1: CHRISTOPHER DAVID NELSON, MANAGER, INFRASTRUCTURE AND SUSTAINABLE DEVELOPMENT UNIT, INDEPENDENT EVALUATION GROUP, WB

1. The World Bank's use of AI and machine learning in evaluation has evolved into operational tools that extract valuable lessons for project delivery, but there is a need to rethink how technology is incorporated into evaluation to maximize its potential.
2. Utilizing machine learning to analyze completion reports has revealed that food security interventions lasting beyond three years significantly improve performance and project ratings, highlighting the importance of long-term planning in development projects.

PANELIST/DISCUSSANT #2: GABRIELE FATTORELLI, DIRECTOR, INDEPENDENT EVALUATION DEPARTMENT, EBRD

1. The EBRD is leveraging existing AI technology to enhance efficiency and reduce costs in its operations, which are predominantly focused on the private sector. The approach emphasizes democratic use, safety, and a phased implementation
2. While EBRD is in the early stages of its AI journey, it aims to expand its use of AI tools. There is a strong interest in collaborating with other MDBs to share evaluation knowledge and make AI tools more accessible to all staff and stakeholders.

PANELIST/DISCUSSANT #3: MAYA VIJAYARAGHAVAN, PRINCIPAL EVALUATION SPECIALIST, INDEPENDENT EVALUATION DEPARTMENT, ADB

1. ADB's IED and IT Department developed EVA, an AI tool, to efficiently extract valuable lessons from a vast collection of evaluation documents. The development process took into consideration three sets of users: (i) operations departments; (ii) evaluation departments; and (iii) high-level users. This significantly improved efficiency and effectiveness of IED's work.
2. Currently, ADB has developed an organization-wide workbench called ADB Genie. This AI-powered research assistant uses three distinct language models to offer multiple

Session Summary

perspectives on the same set of data. The tool is being continuously refined based on user feedback.
MODERATOR: SONA SHRESTHA, DEPUTY DIRECTOR GENERAL, INDEPENDENT EVALUATION DEPARTMENT, ADB
1. The panelists are optimistic on the potential of AI and machine learning tools in evaluation, particularly in extracting evaluation knowledge and findings, and possibly across the entire evaluation cycle. 2. Despite the potential, there was an emphasis on the importance of approaching AI and machine learning with a healthy degree of skepticism and being aware of the associated risks. 3. The journey of integrating AI and machine learning is gradual, and continuous feedback and improvement are crucial to meet the needs of different users.
Q&A SUMMARY
Lesson Extraction Based on Quality Data. The AI tool uses a cognitive full-text search to identify lessons throughout the entire document, not just from designated sections, and attempts to rewrite poorly worded lessons, but the quality of input data remains crucial. Coordination to Improve Usage. There is a need for enhanced coordination and the potential use of deep learning to improve AI tools, reduce resource-intensive tagging, and ensure sustainable and unbiased data usage. Using AI Tools for Triangulation. AI tools should be used to enhance triangulation by raising new questions and ensuring rigorous processes. While still in the early stages, they are continuously refined based on user feedback and quality testing. Use in Evaluation Beyond Summarization. While AI tools have potential in various stages of evaluation, their use has been limited so far, and there is a need for staff sensitization and careful consideration of the interaction and deep learning aspects to maximize their effectiveness.
SESSION COORDINATOR
CHERIE ANNE R. QUIRANTE