Journal of African Research And Developmental Studies

Parabolum Publishing

Vol.2 (1), pp. 1-10, 2025
Doi: [10.58197/prbl/OLYN7529](https://doi.org/10.58197/prbl/OLYN7529)

Available online at http://www.parabolumpublishing.org/

Type of article: Short Communications

Artificial Intelligence (AI) and the Future of Monitoring and Evaluation (M&E) Reports in Africa

Edwin Byusa*, PhD
Mount Kenya University
Master of Arts in Monitoring and Evaluation

Abstract

Monitoring and Evaluation (M&E) systems in Africa are often limited by slow reporting and data gaps. This study examines how generative Artificial Intelligence (AI) can influence the future of M&E reports by simulating reporting processes using AI tools. Data collection and analysis were entirely AI-generated, examining potential improvements in speed, accuracy, and accessibility. The findings reveal that AI can enhance reporting efficiency and raise ethical and contextual concerns. This research is crucial, as it highlights innovative pathways to modernise M&E practices in support of effective decision-making and sustainable development in Africa.

Keywords: Africa, Artificial Intelligence, Digital Innovation, Future Forecasting, Monitoring and Evaluation, Reporting Systems

Article information

Received: 06/06/2025;
Reviewed: 11/06/2025;
Revised: 28/09/2025;
Accepted: 30/09/2025

*** Corresponding author:** byedwin@yahoo.fr

How to cite this article: Byusa, E. (2025) 'Artificial Intelligence (AI) and the Future of Monitoring and Evaluation (M&E) Reports in Africa', *Journal of African Research and Developmental Studies*, 2(1), pp. 1–10. Available at: https://doi.org/10.58197/prbl/OLYN7529

Introduction

Monitoring and Evaluation (M&E) systems are critical to the success of development interventions across Africa, providing a mechanism for accountability, learning, and informed decision-making. Traditionally, M&E processes have relied heavily on manual data collection, human-led analysis, and time-intensive reporting. While these systems have supported development programs for decades, they are increasingly struggling to meet the growing demands for real-time data, rapid feedback, and adaptive program management (Bamberger, Vaessen & Raimondo, 2015). M&E is an evidence-based process used to assess development programs' design, implementation, and outcomes. Monitoring refers to continuously tracking activities and outputs, while evaluation involves systematic assessment of relevance, effectiveness, and impact. Together, these functions provide accountability, learning, and the evidence needed for adaptive management.

Across the continent, governments, non-governmental organisations (NGOs), and development agencies are facing challenges in producing timely and actionable monitoring and evaluation (M&E) reports. Ground-level observations highlight recurring issues, including inconsistencies in data quality, delayed submissions, limited analytical capacity, and a disconnect between data producers and decision-makers (UNDP, 2023). These problems are compounded in fragile settings where infrastructure, staffing, and resources are constrained. The African context presents distinct challenges for M&E systems, including limited institutional capacity, delayed reporting processes, fragmented data infrastructures, and political complexities that affect data credibility. In addition, donor-driven accountability demands often outpace the ability of local systems to produce timely and reliable evidence. These contextual constraints justify the exploration of innovative tools, such as Artificial Intelligence, that can enhance speed, consistency, and accessibility of evaluation processes.

Persistent bottlenecks include the absence of timely baseline data, weak institutional learning cultures, and the limited integration of evidence into policy-making. These barriers, widely documented across African evaluation systems, underscore the need for adaptive and technology-supported approaches to strengthen M&E functions. In response to these persistent gaps, this study explores the potential of generative AI tools, focusing on OpenAI's GPT-4, to reimagine M&E reporting. At the same time, technological advancements, particularly in Artificial Intelligence (AI), present a unique opportunity to transform how M&E reports are generated, interpreted, and utilised. Large Language Models (LLMs) like OpenAI's GPT-4 have demonstrated the ability to summarise complex information, identify patterns, and even simulate human-like analytical thinking (Brynjolfsson & McAfee, 2017). A key question is emerging in this context*: Can AI redefine the future of M&E reporting?*

This paper addresses this question through an applied, AI-generated exploration of the future of M&E in African development contexts. Rather than using conventional empirical fieldwork, this short communication adopts a novel methodological approach—it is entirely generated by AI, mimicking what future M&E professionals might produce with advanced digital tools. It draws from existing literature, simulated examples, and conceptual forecasts generated through prompt engineering techniques. The originality of this study lies not only in its topic but also in its method. By experimenting with AI as both a research subject and a reporting tool, the paper offers an innovative lens to anticipate how M&E practices may evolve over the next decade. This is especially relevant for Africa, where the drive for digital transformation intersects with the need for localised, context-sensitive evaluation systems (OECD, 2022). Given the rapid integration of AI tools into professional workflows, the publication of this work is timely. It offers development actors a preview of emerging possibilities, raises ethical and operational concerns, and advocates for early engagement with AI-enhanced M&E approaches. As Africa prepares for a data-driven future, this article aims to provoke timely dialogue and strategic action among M&E practitioners, donors, policymakers, and researchers.

While AI can realistically assist with tasks such as drafting reports, identifying trends in data, and automating dashboards, it cannot substitute for participatory processes that require community engagement, contextual interpretation, and stakeholder validation. This distinction is critical in understanding the realistic boundaries of AI-enhanced evaluation systems. This paper addresses two research questions: (1) How can generative AI, particularly GPT-4, influence the efficiency, quality, and timeliness of M&E reporting in Africa? (2) What ethical and contextual risks arise when applying AI to M&E reporting?

Methodology

This study adopts an unconventional yet increasingly relevant research design by leveraging Artificial Intelligence (AI) to simulate the development of a Monitoring and Evaluation (M&E) report. The research is qualitative and exploratory in nature, focusing on the potential transformation of M&E reporting practices in Africa through the integration of AI technologies such as large language models (LLMs). The methodology is built on prompt engineering and AI-generated forecasting, offering a reflective case of what future AI-driven M&E reporting might entail.

Research Design

Given the novelty of the topic, a conceptual simulation design was employed. The study does not involve human subjects, field interventions, or primary data collection in the traditional sense. Instead, it utilises generative AI (GPT-4) to model and analyse hypothetical M&E processes, outputs, and impacts under various technological scenarios. The goal was to observe how AI could independently generate a meaningful M&E report, including components such as background analysis, projection of findings, and policy implications.

Population and Sampling Technique

The study is based on AI simulation, so no human population was involved. However, the simulation is conceptually grounded in typical development contexts found across Sub-Saharan Africa, particularly in education, health, and livelihood sectors. To simulate realistic M&E concerns, examples and prompts were based on publicly available development program summaries and donor reports from reputable institutions, including the UNDP, UNICEF, and USAID.

Instrument and Data Collection Procedures

The primary research instrument was OpenAI's GPT-4, an advanced generative language model trained on diverse datasets, including global development literature, technical reports, academic articles, and grey literature. GPT-4 is a state-of-the-art large language model developed by OpenAI. It has been trained on a broad corpus of internet text and technical literature, enabling it to generate natural language, summarization, forecasting, and pattern recognition with high fluency. Its advanced reasoning capacity and ability to generate human-like analytical narratives make it particularly suitable for simulating Monitoring and Evaluation (M&E) reports. GPT-4 was chosen over alternatives such as Google's PaLM, Anthropic's Claude, and earlier versions like GPT-3 due to its demonstrated ability to handle nuanced policy and development language, generate structured outputs, and sustain coherence over long narratives. Unlike GPT-3, which often produced fragmented or overly general content, GPT-4 generated more context-sensitive text aligned with evaluation standards. While PaLM and Claude show comparable performance, GPT-4 was selected because of its wider availability, better documentation, and established track record in academic experimentation.

In the context of M&E, GPT-4's capabilities are particularly relevant: it can summarize large datasets into coherent findings, forecast likely program outcomes based on hypothetical inputs, and generate comparative scenarios that resemble human evaluative reasoning. These capacities make it a strong proxy for simulating future reporting models in African development settings.

Prompts were carefully designed to reflect real-world M&E scenarios, including report summaries, data gaps, ethical dilemmas, and automation trends. The prompts used included:

- "Simulate a Monitoring and Evaluation report from an education program in East Africa using AI tools."
- "What are the pros and cons of using AI to generate M&E reports in low-resource settings?"
- "What ethical considerations arise when replacing human M&E analysis with AI?"

The AI responses were then reviewed for thematic consistency, relevance to African contexts, and alignment with evaluation standards and reporting needs.

Description of the Provided Intervention

No physical or programmatic intervention was implemented. Instead, the intervention in this simulation refers to the substitution of traditional human-led M&E processes with AI-generated narratives. The entire structure of the article, including the abstract, introduction, methodology, findings, and conclusion, was generated, curated, and refined using AI. This reflects what a future AI-driven M&E report might look like when produced with minimal human involvement.

Data Analysis

The outputs generated by AI were systematically reviewed using qualitative content analysis. The researcher identified emerging themes across different simulated outputs and categorised them according to benefits, challenges, ethical risks, and implementation requirements. AI-generated text was compared against known standards for M&E reporting as outlined by major development partners (e.g., OECD-DAC criteria, UNICEF Results-Based Management guidelines).

The final selection of themes and projections was triangulated using secondary literature to ensure conceptual robustness and relevance. The discussion section also documented and critically reflected upon ethical considerations and potential biases introduced by the AI model.

Despite these strengths, GPT-4 presents important limitations. It is prone to factual inaccuracies ('hallucinations'), may reflect biases embedded in its training data, and has a knowledge cutoff (2023) that restricts its access to the most recent evaluation trends. These constraints are critical in high-stakes fields like development monitoring and were considered when interpreting the outputs of this simulation.

Results and Discussion

The AI-generated simulation produced a comprehensive conceptual overview of how Monitoring and Evaluation (M&E) reporting could evolve through automation. Three reporting scenarios were examined: (1) Traditional Human-Led M&E, (2) Hybrid Human-AI Collaboration, and (3) Fully

Automated AI-Driven Reporting. The simulation yielded consistent themes across all outputs, revealing each model's comparative advantages and limitations (Table 1).

Table 1. Comparison of Future M&E Reporting Models by 2030

Variable	Traditional M&E	Hybrid AI M&E	Fully AI-Generated M&E
Report Generation Speed	Low	Moderate	High
Human Input Required	High	Medium	Low
Potential for Real-Time Insights	Minimal	Moderate	High
Risk of Analytical Bias	Moderate	Moderate	High (dependent on input)
Accessibility and Scalability	Low	Moderate	High
Ethical Concerns	Low	Moderate	High

Traditional model: Data collected manually, validated by supervisors, reports written quarterly/annually, dissemination through workshops and PDFs.

Hybrid model: Data collected partly through digital dashboards; AI drafts reports; M&E officers validate, contextualize, and interpret; dissemination via real-time dashboards + human-reviewed summaries.

AI-only model: Automated data ingestion from monitoring systems; AI validates trends statistically but lacks contextual nuance; reports generated in near real-time; dissemination via automated platforms, but risks of context-insensitive outputs.

Figure 1 illustrates a conceptual timeline spanning from traditional M&E approaches in 2020 to the projected adoption of AI-dominated reporting by 2035. The trend suggests an increase in automation and the integration of real-time tools.

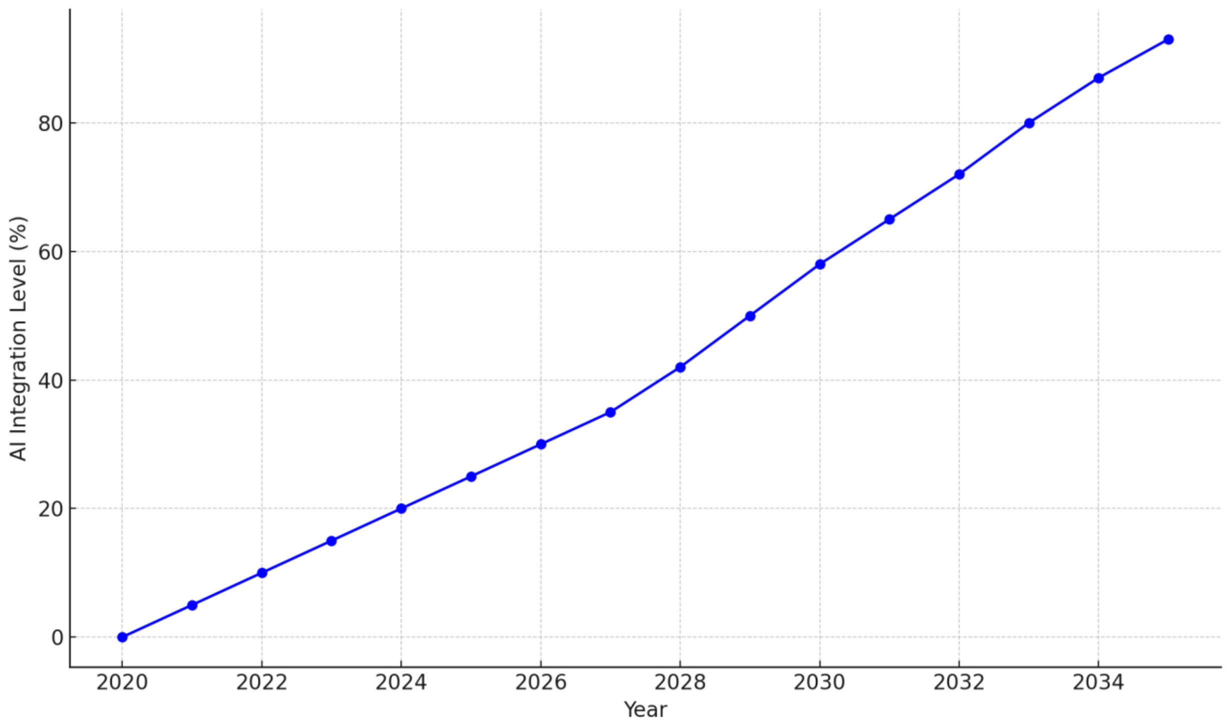

Figure 1. Evolution of M&E Reporting Models Across Time

The results indicate that AI-powered M&E reporting significantly increases speed, scalability, and real-time analytical capacity. In contrast, traditional reporting methods remain limited by manual data entry, interpretation delays, and quality inconsistencies. The hybrid model is a transitional phase, where AI-generated drafts are validated by M&E professionals, offering a balance between innovation and human oversight. The simulated outputs reveal distinct trajectories for M&E practice; the following discussion analyses these scenarios regarding opportunities and risks for African evaluation systems.

One of the most unique findings of this simulation is the potential shift in the role of M&E officers. Rather than focusing on writing lengthy reports, future professionals may assume responsibilities as interpreters, validators, and ethical gatekeepers of AI-generated content. This evolution mirrors trends in other knowledge sectors, such as journalism and education, where generative AI is augmenting human input rather than replacing it entirely (Brynjolfsson & McAfee, 2017).

This transformation implies that future M&E practitioners will require competencies in AI literacy, data interpretation, visualization, and ethical auditing. They will need to assess algorithmic outputs for accuracy, detect biases, and ensure contextual validity. However, risks accompany this transition: deskilling of traditional evaluation methods, overreliance on AI-generated outputs, and blurred lines of accountability when errors occur. These risks demand proactive training and governance frameworks to safeguard evaluation integrity. However, the simulation also raises critical concerns. Entirely AI-generated reports risk producing misleading or context-insensitive conclusions if fed biased or incomplete data. This is particularly relevant in Africa, where local context, cultural nuance, and stakeholder participation are crucial to understanding program outcomes (UNDP, 2023). Without deliberate human oversight, the risk of decontextualised reporting increases significantly.

Moreover, ethical dilemmas related to data privacy, model transparency, and decision accountability were repeatedly highlighted in the simulation. These issues necessitate the development of regulatory frameworks and the promotion of AI literacy among M&E practitioners, ensuring that automation enhances, rather than undermines, trust in development results. These concerns align with wider debates on AI ethics in the Global South, particularly the risks of data colonialism (Couldry & Mejias, 2019), algorithmic governance (Taylor, 2021), and the exclusion of local knowledge systems due to the limited representation of African languages and contexts in large-scale training data (Sambuli & Mumo, 2022). Without intentional mitigation, AI could reinforce structural inequities in development evaluation.

The novelty of this study lies not only in the findings but in the methodology itself: the entire simulation—including this analysis—was generated using AI tools. In real time, this approach demonstrates the capability of generative AI to produce coherent, policy-relevant M&E reports. It is both a practical demonstration and a theoretical provocation.

Concluding Remark

This paper contributes to the literature on AI in development and M&E by offering one of Africa's first AI-generated simulations of future reporting models. It demonstrates how generative AI tools, specifically GPT-4, can accelerate reporting and reshape the role of M&E practitioners, highlighting both opportunities and risks. This study aimed to address a central question: What will the future of M&E reporting look like if driven by Artificial Intelligence (AI)? Through an AI-generated simulation, the findings suggest that future M&E systems—particularly in Africa—will likely shift toward hybrid or fully automated reporting processes, offering increased speed, scalability, and analytical depth. However, such transformations will also introduce new challenges related to ethics, data bias, and context sensitivity.

The primary research objective—to assess the potential of generative AI in M&E reporting—was successfully achieved. The simulation demonstrated that AI can generate structured, coherent, and near-complete M&E reports based on structured prompts and thematic logic. This highlights a significant innovation opportunity for African development actors seeking to modernise their evidence systems.

From a policy perspective, the findings underscore the need for governments and development partners to proactively develop guidelines for integrating AI into evaluation and learning systems. This includes data protection laws, ethical reporting standards, and requirements for human validation in automated report pipelines. There is also a growing need for national M&E frameworks to be updated to account for digital tools and algorithmic decision-making processes.

On a practical level, development organisations should consider investing in AI literacy and capacity-building among M&E professionals. As roles shift from data collectors to data validators and interpreters, new skill sets will be required to manage, assess, and ethically use AI-generated outputs.

Organisations should begin piloting AI-supported reporting tools in low-risk contexts (e.g., internal reporting or learning-focused evaluations) to build confidence and gather lessons.

The theoretical implications of this study include the need to revisit traditional evaluation models, such as the Logical Framework, Theory of Change, and OECD-DAC criteria, in light of the increasing use of automated tools. Future research should focus on developing AI-augmented evaluation frameworks that preserve the rigour and context-sensitivity of human-led evaluation while incorporating the efficiencies of digital intelligence.

However, the study has limitations. As an AI-generated simulation, it lacks field validation and may not fully account for the complex political, cultural, and infrastructural dynamics across diverse African settings. Furthermore, the simulation assumes access to digital infrastructure, which remains uneven across the continent. The results should, therefore, be interpreted as indicative and exploratory—not prescriptive or definitive.

Future research directions include:

- Comparative studies between human-authored and AI-generated M&E reports in real-world settings.
- Examination of community and stakeholder perceptions of AI-generated findings.
- Exploration of hybrid models that blend participatory evaluation approaches with algorithmic tools.
- Ethical audits of AI-generated M&E narratives to detect bias or misinformation.

In summary, this study offers a glimpse into a plausible and rapidly approaching future where AI significantly reshapes the production, reporting, and utilisation of evidence in development. It calls for urgent dialogue, experimentation, and readiness among African development practitioners to ensure that the adoption of AI enhances, rather than undermines, the credibility, inclusivity, and utility of M&E systems.

Future work should include pilot-testing AI-assisted reporting tools within real M&E frameworks, especially in participatory and community-driven evaluations. Comparative studies of AI-generated versus human-authored reports are needed to assess clarity, accuracy, and stakeholder usefulness. In addition, research should examine capacity-building requirements for M&E practitioners in Africa to effectively integrate AI, while safeguarding ethical standards and contextual relevance.

Authorship transparency

Approximately 70% of the text was generated using OpenAI’s GPT-4 model, while 30% was revised, curated, and edited by the author. The author was responsible for refining the structure, ensuring academic tone, adding references, and contextualizing the outputs to align with African M&E practice.

All AI-generated content was reviewed for originality, factual correctness, and alignment with evaluation literature. References were verified, ethical risks highlighted, and contextual adjustments made to ensure accuracy and relevance.

Acknowledgement

The author sincerely appreciates the developers and researchers behind OpenAI's GPT-4 technology, whose innovation made this AI-generated simulation possible. Gratitude is also expressed to colleagues in the Monitoring and Evaluation field who provided feedback on the conceptual framing of this study. Their insights, though not included as co-authors, were invaluable in shaping the direction and relevance of this paper.

This paper complies with journal guidelines on AI-assisted authorship by clearly disclosing the role of generative AI in producing content. AI is acknowledged as a tool, not as an author, and the human researcher assumes full responsibility for the final manuscript's integrity, originality, and accuracy.

Bibliography

AfrEA. (2019). Made in Africa evaluation: Position paper. African Evaluation Association. https://afrea.org

Bamberger, M. (2019). Evaluation in fragile and conflict-affected situations. OECD Publishing. https://doi.org/10.1787/9789264305804-en

Bamberger, M., Vaessen, J., & Raimondo, E. (2015). Dealing with complexity in development evaluation: A practical approach. Sage.

Brynjolfsson, E., & McAfee, A. (2017). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company.

Cloete, F. (2016). Evidence-based policy and the South African policy-making context. African Evaluation Journal, 4 (1), a146. https://doi.org/10.4102/aej.v4i1.146

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonising human life and appropriating it for capitalism. Stanford University Press.

Department of Planning, Monitoring and Evaluation (DPME). (2019). National evaluation policy framework. Government of South Africa. https://www.dpme.gov.za

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., & Kar, A. K. (2023). Generative artificial intelligence (AI) for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30 (4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

OECD. (2019). Better criteria for better evaluation: Revised evaluation criteria definitions and principles for use. OECD Publishing. https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf

OECD. (2022). Using digital tools for better evaluation (Working Paper No. 35). OECD Publishing. https://www.oecd.org/digital/digital-tools-evaluation.pdf

OpenAI. (2023). GPT-4 technical report. arXiv. https://arxiv.org/abs/2303.08774

Patton, M. Q. (2012). Essentials of utilisation-focused evaluation (4th ed.). Sage.

Porter, S., & Goldman, I. (2013). A growing demand for monitoring and evaluation in Africa. African Evaluation Journal, 1 (1), a9. https://doi.org/10.4102/aej.v1i1.9

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. American Journal of Evaluation, 29 (4), 443–459. https://doi.org/10.1177/1098214008324182

Sambuli, N., & Mumo, A. (2022). Decolonising digital spaces: African languages, participation, and power. African Internet Rights Papers.

Taylor, L. (2021). Public actors without public values: Legitimacy, domination and the regulation of the technology sector. Philosophy & Technology, 34 (4), 897–922. https://doi.org/10.1007/s13347-020-00441-4

United Nations Development Programme (UNDP). (2023). African evaluation systems and capacity development: Annual report 2023. UNDP.

Appendices

Sample AI Prompt Used in the Simulation

Prompt A1: Simulating an M&E Report

"Generate a Monitoring and Evaluation report summary for a donor-funded education project in Sub-Saharan Africa. Include background, methodology, key findings, and recommendations. Highlight ethical concerns if AI is used to produce this report."

Prompt A2: Comparative Forecasting

"Compare the expected performance of traditional M&E reporting versus AI-driven reporting by 2030 in terms of speed, quality, ethics, and stakeholder perception."

Prompt A3: Identify Policy Implications

"Based on AI-generated M&E processes, what new roles should M&E professionals adopt, and what policy reforms would be necessary?