WFP EVALUATION

Info Brief

Ethical Use of Artificial Intelligence in WFP Evaluations

1. Purpose

1. This brief provides information to guide the use of Artificial Intelligence (AI) in WFP evaluations, as applicable by evaluation managers, firms and Evaluation Teams to produce any evaluation deliverables. While the brief does not prescribe or elaborate on technical aspects (types, methods, or tools), it recognizes that a basic understanding of technical characteristics is essential for addressing ethical, data-related, and evaluative considerations in AI-supported work. The brief highlights risks of AI in evaluations and how to mitigate them.

2. Background

2. The humanitarian community, including WFP and its stakeholders, is navigating a rapidly changing landscape marked by data proliferation and the emergence of AI and Machine Learning (ML). WFP considers AI as an opportunity for use to develop a data-centric workforce and culture. [1]

3. AI and ML have the potential to significantly enhance WFP's activities and operations by identifying patterns, trends and insights and enabling more informed strategic and operational decision-making. Harnessing AI responsibly offers WFP new capabilities to deliver assistance faster, more efficiently, and more equitably to those we serve. WFP's Interim AI Guidance[2] helps WFP employees to ensure responsible uptake of AI.

4. The emergence of AI, building on the exponentially rapid developments of generative AI over the past decade, has created significant openings in the capacity to process and analyze vast datasets, which are relevant for use in evaluation. AI-powered tools increasingly offer appealing use-cases for complex evaluations that necessitate exploitation of secondary data, offering expanded capability to systematically analyze quantitative and qualitative data with less time and costs. Hence, most evaluators have increasingly sought to harness these new capabilities.

5. Applying AI technology to evaluation however presents significant risks that must be identified and managed to ensure effective and ethical use; to ensure alignment with the UN's core values of fairness, equality, and respect for human rights. Ethical oversight is therefore a moral obligation to uphold these values.

6. Considering the attractiveness of AI tools in relation to many aspects of evaluation work, including processing large datasets, or interpreting these into analytical findings, the appeal of these technologies to evaluation teams will continue to become increasingly strong and therefore needs to be framed by appropriate norms.

---

[1] WFP Global Data Strategy, 2025
[2] WFP Interim Artificial Intelligence Guidance, 2025 (internal only)

September 2025

3. Risks of AI in evaluations

7. The United Nations System White Paper on AI Governance[3] identifies current and evolving AI risks to include regulatory design challenges; opacity, explainability and interpretability of AI data; challenge of tracking and monitoring decentralized AI systems; inter-related risks from data and cybersecurity; interrelated risks from copyrights and intellectual property rights; growing AI divide; lack of effective enforcement; and inclusivity among others. Evaluators should also familiarize themselves with these general AI risks before deciding on the use of AI for evaluations.

8. AI systems used in evaluations **may perpetuate or exacerbate existing societal inequalities because AI may be relying/built on data based on structural biases and prejudices, which may further embed/entrench such biases**, thereby leading to discriminatory outcomes and undermining the UN's commitment to global equality. This risks stem from many AI models having been trained on big data (mainly the Internet) that was formed from and therefore reflected mainly Western perspectives. Biases generated by the algorithms underlying AI are often unknown by the user. Inclusivity and attention to diversity of perspectives are imperative to ensuring accessibility and consideration of diverse needs and socio-cultural references.

9. Use of AI presents a risk of **violation of UN ethical principles in evaluations, especially on privacy and data protection, and accuracy and reliability**. There are concerns relating to privacy, consent, data ownership, and the potential exacerbation of existing power imbalances, also underlining the need for careful consideration and use of AI to generate evaluative evidence. Handling sensitive data, especially concerning vulnerable groups, is a key aspect of UN evaluations, necessitating strong data privacy and security measures, vital to maintain trust among data providers and those affected by evaluations. Ethical risks of exposing internal UN documents to open-source Generative AI models include exposing sensitive data, violation of international data privacy laws, error propagation, misinterpretation of context, and systemic bias.

10. AI often struggles with **understanding complex, real-world contexts, particularly those involving cultural nuances or unique situations**, when humans can intuitively navigate such complexities. This raises the risk of AI extracting data and drawing conclusions out of context.

11. The often complex and opaque nature of AI systems and solutions to evaluators, given the complexity of their technological basis, may not allow users to fully understand the implications of using open-access tools, which compound the above-mentioned risks.

12. Building on the principles for the Ethical Use of Artificial Intelligence in the United Nations System[4], and ensuring that adoption of AI in UN evaluations aligns with UN's core values, UNEG has developed five ethical principles for use by UN agencies in navigating the complexities of integrating AI into their evaluation processes in an ethical way[5]. These ethical principles include transparency and accountability; fairness and inclusivity; data protection and privacy; validity and reliability; and human-rights-based AI implementation. WFP adopted these high-level principles alongside other UN evaluation system agencies, in February 2025.

4. Mitigating risks of AI in WFP evaluations

Terms of Reference

13. The evaluation manager should ensure that evaluation terms of reference have incorporated the following paragraphs:

   a. Artificial Intelligence (AI) technologies may only be used in the framework of this evaluation with prior written approval from the WFP Office of Evaluation (OEV). Upon receiving this approval, the evaluation team shall clearly and comprehensively disclose in the inception report, the intended utilization of AI tools in evaluation, including the purpose, scope and nature of the proposed AI usage.

   b. The evaluation **team** shall uphold ethical standards and accuracy in the application of AI tools. This includes diligently checking the accuracy and reliability of AI-generated outputs and assuming full responsibility for its reliability and validity.

   c. The **evaluation** team shall ensure that their use of AI technologies comply with relevant normative and ethical frameworks applicable to the use of AI in the United Nations system / WFP. These include but not limited to: the Digital & Technology Network Guidance on the Use of Generative AI Tools in the United Nations System, Principles

---

[3] United Nations System White Paper on AI Governance.pdf
[4] CEB Principles for ethical use of Ai in the UN system
[5] UNEG Ethical Principles for Harnessing AI in United Nations Evaluations

for the Ethical Use of Artificial Intelligence in the United Nations System[6], WFP Global Data Strategy[7], UNEG Ethical Principles for Harnessing AI in United Nations Evaluations, WFP AI strategy (2025)[8], WFP's Personal Data Protection and Privacy Framework[9], WFP's Interim AI Guidance[10], and the Principles for Ethical Use of AI in the United Nations System[11]. The evaluation team shall employ AI tools in a manner that upholds the principles of non-discrimination, fairness, transparency, and accountability and take appropriate measures to avoid the exclusion, disadvantage or harm of any group in connection with the use of AI technologies.

d.   The evaluation team shall ensure that any data used in connection with AI tools is handled in accordance with WFP data protection standards and confidentiality obligations. AI tools shall not be used in a manner that compromises the privacy or security of evaluation data.

Transparency on the use of AI

14. The evaluation manager and evaluation teams shall ensure that evaluation reports include a disclaimer regarding the use of AI, as appropriate.

Ethical use of AI in evaluation services

15. To ensure ethical use of AI in evaluation services, the evaluation team shall conduct its evaluation activities in conformity with (a) the Principles for the Ethical Use of Artificial Intelligence in the United Nations System[12]; (b) UNEG Ethical Principles for Harnessing AI in United Nations Evaluations[13]; c) Data principles in the WFP Global Strategy[14] for safe, secure, and trustworthy AI; (d) any other relevant WFP ethical or guidance documents on the use of AI, issued by WFP from time to time and formally communicated to the evaluation team during the duration of the evaluation.

16. The evaluation team shall integrate the applicable principles and requirements from such documents into in all aspects of AI utilization in WFP evaluation work and analysis, to the extent relevant and applicable.

17. The evaluation team shall respect human rights, protect human dignity, and avoid any use of AI that could reinforce discrimination, inequity, or bias.

Verification of results

18. The evaluation team shall implement appropriate verification processes to validate the results generated by AI analyses. The evaluation team shall be responsible for ensuring the reliability and validity of the AI analyses performed, including taking the necessary measures to detect and address any errors, bias, discrimination or unintended adverse outcomes.

19. The evaluation team shall ensure that where AI has been used for data synthesis and analysis, there is additional data review and validation by humans before using it as evidence to inform evaluation findings, conclusions and recommendations. This will ensure that meaningful human oversight is maintained over the interpretation and use of AI-generated outputs in evaluation deliverables.

20. Along with the AI analyses, the evaluation team shall provide a clear and comprehensive description of the methodology employed for the AI analyses. The evaluation team shall maintain documentation of the methodology and verification processes and make such documentation available to WFP upon request.

---

[6] Principles for the Ethical Use of AI in the UN System_1.pdf
[7] WFP Global Data Strategy, 2025
[8] WFP Global Artificial Intelligence Strategy.
[9] WFP Personal Data Protection and Privacy Framework
[10] WFP Interim Artificial Intelligence (AI) Guidance (2025)
[11] Principles for the ethical use of artificial intelligence in the United Nations system
[12] Principles for the ethical use of artificial intelligence in the United Nations system
[13] UNEG Ethical Principles for Harnessing AI in United Nations Evaluations
[14] WFP Global Data Strategy, 2025

Monitoring of AI outputs and processes

21. The evaluation manager shall use WFP AI Guidance to review the use of AI outputs and processes related to evaluation products, to ensure adherence to ethical AI principles. Following a review by the evaluation manager, the evaluation team shall address and rectify any biases, errors, or shortcomings in the AI analyses.

Data protection principles

22. The evaluation team shall ensure that any AI tools that could be processing personal data incorporate the data protection principles outlined in WFP's Personal Data Protection and Privacy Framework[15] by design and by default and uphold these principles throughout the whole data lifecycle.

23. The evaluation team shall provide WFP with documentation specifying the categories of data ingested and what data security measures were applied in relation to the use of AI in evaluations. It shall also provide evidence of review by a human or proof of any human intervention, before using the outcomes as evidence to inform evaluation findings.

Generative AI and data security

24. The evaluation team shall not use proprietary Generative AI tools that store or process data externally (external to WFP or contractor). Using WFP non-public data with proprietary AI tools may contravene WFP policies, jeopardize data privacy and protection, and can cause information security incidents.

25. The evaluation team shall have proof that AI systems/models that are used in the evaluation have passed international cybersecurity standards, WFP and UN data and data security standards[16], and that the AI models being used have been scrutinized for best practices.

---

[15] WFP Personal Data Protection and Privacy Framework
[16] UN data privacy, ethics, and protection guidance note on big data for achievement of the 2030