UNEG Reference Document

UNEG
United Nations Evaluation Group

UNEG Ethical Principles for Harnessing AI in United Nations Evaluations

This document was prepared by the Data and AI Working Group 2025 and was approved as a UNEG Reference document at the Annual General Meeting 2025.

Table of contents

Introduction

1. The Global Digital Compact, adopted as part of the Pact for the Future at the September 2024 Summit for the Future, marks a significant step towards establishing global governance of digital technologies, including artificial intelligence (AI). The compact underscores the importance of grounding AI development and deployment in human rights, echoing the United Nations (UN) vision of integrating ethical principles in AI use, that resonate with its mission of promoting global peace, security, human rights and, sustainable development[1]. This includes UN evaluations where AI applications must uphold United Nations Evaluation Group's (UNEG) Ethical Guidelines for Evaluations and Norms and Standards for Evaluations in the UN System, as well as enhance accuracy and efficiencies.

2. Central to this ethical approach of utilizing AI in evaluation are the principles of fairness, transparency, accountability and privacy, particularly when handling sensitive data from evaluated programmes and initiatives. The principle of beneficence directs the use of AI towards human well-being and social good. These principles ensure that the UNEG's use of AI in evaluation aligns with its goals of setting standards for UN evaluation that reflects good and responsible practices.

3. This paper discusses the increasing interest and practice of applying AI in UN evaluations and the rationale for developing a more comprehensive guidance protocol for establishing responsible use of AI for evaluation. It also presents five key ethical principles that underscore the responsible integration of AI in UN evaluation processes.

The Emergence of AI in UN Evaluation Practice

4. Developments in AI have marked a significant advancement in the capacity to process and analyze vast datasets. Over the past decade, AI methodologies have found extensive applications in development practices, with notable use-cases in complex evaluations which necessitate further exploitation of secondary data. These AI analytics offer the capability to systematically analyze both quantitative and qualitative data in a timely and cost-effective manner, as highlighted in a recent study by Bravo et al., 2023[2]. Recent breakthroughs in Generative AI, particularly in areas like natural language processing and content creation, have significantly expanded the potential applications of AI in evaluation. Furthermore, an increasing number of UN agencies have taken proactive steps to explore the responsible use of AI across the evaluation lifecycle[3], including the application of AI analytics in analyzing various forms of internal monitoring data and reports[4], and use of AI for various evidence syntheses exercises. As the integration of AI in UN evaluation grows, it is increasingly crucial to establish and adhere to ethical principles in AI use that ensure fairness, transparency and accountability, safeguarding against potential biases and promoting responsible innovation in this area.

Harnessing AI in UN Evaluations

5. By harnessing responsible AI use, UN evaluation functions could strengthen the delivery of timely and robust, high-quality evidence to address complex global issues and decision-making. A key way of

[1] Vota, W. (2022, October 26). 10 Ethical Use Principles for Artificial Intelligence in the UN System. ICT works. https://www.ictworks.org/artificial-intelligence-principles-united-nations/
[2] Bravo, L., Hagh, A., Joseph, R., Kambe, H., Xiang, Y., & Vaessen, J. (2023). Machine Learning in Evaluative Synthesis: Lessons from Private Sector Evaluation in the World Bank Group.
[3] UNFPA Independent Evaluation Office (2024). GenAI-powered evaluation function at UNFPA.
[4] UNICEF (2022). Artificial Intelligence Strategy Working paper.

advancing the integration of AI in UN evaluations is through UNEG interagency knowledge exchange. UNEG's Data/AI Working Group currently organizes a series of regular meetings and webinars centered on AI's emerging role in evaluations. Such platforms serve as a hub for sharing expertise and best practices, fostering a dynamic and cooperative environment among evaluators and stakeholders. Pilot studies, as presented by various UN agencies, are also crucial part of this learning and adoption. These initial trials are aimed to test AI's applicability and effectiveness in real-world evaluation scenarios, identifying challenges and shaping the ethical guidelines for AI application in evaluation.

6. To enhance its position at adopting AI technology, UNEG member agencies are contemplating a strategic initiative to form partnerships with leading figures in the field of AI, ranging from academic experts to prominent players in the tech industry. Such partnerships are poised to provide UNEG with access to state-of-the-art AI solutions and innovative methodologies, while ensuring alignment with UN values and evaluation standards. In parallel, a significant emphasis is being placed on the training and development of UN evaluation staff to equip them with the necessary skills to apply AI techniques effectively and responsibly in their work.

The Rationale for Ethical Principles in AI Adoption in Evaluation

7. Ethical principles are crucial in the adoption of AI for UN evaluations to ensure alignment with the UN's core values of fairness, equality and respect for human rights. AI, if not ethically managed, risks embedding biases leading to discriminatory outcomes, thereby undermining the UN's commitment to global equality[5]. There is also a growing body of AI ethics research that highlights the harms of irresponsible and unethical use of AI towards marginalized groups, including the potential for misuse of this technology and unintended consequences[6]. Ethical oversight is essential not only as a technical requirement but as a moral obligation to uphold human rights and UN values.

8. The handling of sensitive data, especially concerning vulnerable populations in crisis or conflict situations, including personal data protection, necessitates robust data privacy and security measures aligned with UN standards. This is vital for maintaining trust among data providers and those affected by evaluations. The complex nature of AI demands a transparent and accountable framework that enables stakeholders to understand and evaluate AI-driven decisions, crucial for maintaining the credibility of UN evaluation processes.

9. Inclusivity and diversity are also imperative in ethical AI adoption, ensuring accessibility and consideration of diverse needs, in line with the UN's mission for global peace, security and well-being. The Global Digital Compact highlights the need to sets standards in AI use; promote responsible and ethical AI use globally; and foster international cooperation in AI governance in close partnership with the Global South countries.

10. In the context of evaluations, which often involve critical assessments impacting various initiatives at the local, national, regional and international levels, the ethical use of AI is paramount. Fairness is a

[5] Office of Information and Communications Technology of the United Nations Secretariat (OICT) (2022). "Principles for the Ethical Use of Artificial Intelligence in the United Nations System," High-level Committee on Programmes (HLCP), endorsed by the United Nations System Chief Executives Board for Coordination, Sept 2022.
[6] Adapted from UNFPA Independent Evaluation Office (2024). GenAI-powered evaluation function at UNFPA.

primary ethical concern, ensuring that AI systems used in evaluations do not perpetuate or exacerbate existing societal biases and inequalities[7].

Challenges and Limitations to Applying AI to UN Evaluations

11. Applying AI technologies to UN evaluations presents several challenges and risks that must be addressed to ensure effective and ethical use. A significant hurdle is the lack of standardized templates and formats of evaluation reports which are essential for properly training AI models. Another major concern is the risk of biases and discrimination in AI algorithms, particularly when evaluating programmes affecting vulnerable populations or crisis situations. If the training data is not representative of the diverse evaluation sample, it can lead to exclusionary and biased AI generated outcomes. Generative AI requires careful attention, as its outputs may contain misinformation necessitating thorough human review and validation.

12. Ethical implications also pose a challenge, including issues related to privacy, consent, data ownership, and the potential exacerbation of existing power imbalances. These concerns highlight the necessity of careful consideration and management when using AI for evaluative evidence.

13. While AI has advanced, especially in natural language processing (NLP) and machine learning, it still faces limitations in adaptability and insight compared to human capabilities. AI often struggles to understand complex, real-world contexts which humans intuitively navigate, particularly those involving cultural nuances or unique situations not represented in the training data. Additionally, AI can find it challenging to generalize knowledge across different domains or unfamiliar situations, whereas humans are typically more adept at applying their knowledge to new contexts.

14. AI's limitations extend to creativity and problem-solving as well. It relies on data patterns and correlations, lacking the human capacity for novel idea generation, lateral thinking and abstract problem-solving. Emotional intelligence is another area where AI falls short. It cannot match human skills in empathy, social understanding and adapting to emotional contexts. In other words, AI systems do not possess the ability to make ethical judgments or understand the moral implications of the data they analyze. This is particularly relevant in UN evaluations, where ethical considerations are paramount. As stated earlier, human oversight is necessary to interpret AI-generated data within an ethical framework, ensuring that evaluations align with the UN's core values while also removing any biases and misinformation.

15. Another significant limitation of AI is its inability to verify the reliability and validity of data and information[8]. Its filtering may not always be accurate in terms of timeliness, facts, or logic, requiring human oversight. For much of the work being conducted using NLP, these approaches are based on machine learning, relying heavily on the data they are trained on. If this data is biased, incomplete or not representative of the real-world scenario, the AI's conclusions can be flawed and unreliable. Therefore, prioritizing accuracy in AI-generated outputs is crucial for ensuring the credibility and utilization of evaluations.

[7] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization.
[8] Punia, Sanjeev Kumar, et al. "Performance Analysis of Machine Learning Algorithms for Big Data Classification: ML and AI-Based Algorithms for Big Data Analysis." IJEHMC vol.12, no.4 2021: pp.60-75. http://doi.org/10.4018/IJEHMC.20210701.oa4.

16. Given AI's relatively nascent applications throughout UNEG member agencies, these technologies should still require continuous monitoring and assessment along each step of the process. This will help ensure that agencies meet intended objectives and address the inherent challenges in AI use in evaluation, maintaining the crucial role of human expertise and oversight in the process.

Ethical Principles for AI in UN Evaluations

17. When adopting AI for use in conducting and managing UN evaluations, several ethical principles need to be carefully considered to ensure that the technology aligns with the UNs' values and standards. This section presents five key principles that underpin the ethical use of AI in conducting and managing evaluations.

1. Transparency and Accountability

18. AI systems used in UN evaluations must demonstrate full methodological transparency through detailed documentation aligned with UNEG Norms and Standards. This includes clear documentation of training data sources, model parameters, decision-making algorithms, explainability and established oversight frameworks enabling regular audits by relevant UN bodies. Moreover, AI-enabled evaluation methodology must be fully explainable and transparent to all stakeholders, ensuring affected populations, evaluation users and other parties can understand both the process and the rationale behind AI-driven findings.

19. Transparency and accountability in AI-enabled evaluations are essential to maintain the credibility and trust in UN evaluation findings that inform global policy decisions. The "black box" nature of AI systems demands rigorous documentation and oversight to ensure methodological integrity and maintain public trust. This principle is critical because UN evaluations influence resource allocation and policy decisions affecting vulnerable populations worldwide. By mandating transparent AI processes and clear accountability mechanisms, the UN ensures that evaluation findings remain trustworthy and aligned with international standards while fostering institutional learning and continuous improvement in AI-enhanced evaluation practices.

2. Fairness and inclusivity

20. Ensuring fairness and actively mitigating bias in AI-enabled evaluations is fundamental to upholding the UN's mandate to promote equality and protect vulnerable populations worldwide. AI systems have the potential to perpetuate societal inequalities through skewed data representation and flawed algorithmic design that misrepresent marginalized communities. Further, skewed findings can influence resource allocation and policy decisions affecting millions of lives. Given the UN's commitment to leaving no one behind (LNOB), mandating rigorous bias testing and implementing safeguards aligned with the UNEG Ethical Guidelines ensures AI-enhanced evaluations promote equity rather than reinforce discrimination. This systematic approach strengthens the UN's role in advancing social justice through evidence-based evaluation practices that accurately reflect diverse population needs.

3. Data Protection and Privacy

21. Robust data protection and informed consent protocols are fundamental ethical requirements when deploying AI in UN evaluations, particularly given the increasing sophistication of AI systems to process sensitive personal information across diverse cultural contexts. The collection and processing of personal data through AI systems presents unique vulnerabilities that can disproportionately affect marginalized populations and those in conflict zones[9]. This principle is necessary because UN evaluations often handle sensitive data from vulnerable populations who may face severe consequences if their information is compromised or misused. Beyond basic data security, informed consent in AI-enhanced evaluations must address the complexities of algorithmic processing and potential future uses of data that may not be immediately apparent to participants. The principle's importance is amplified in conflict-sensitive environments where data breaches could endanger lives or exacerbate existing tensions. By mandating stringent data protection measures and culturally appropriate consent processes, the UN upholds its duty of care while ensuring that technological advancement in evaluation practices does not come at the cost of individual privacy and security.

4. Validity and reliability

22. The integration of AI in UN evaluations requires commitment to methodological rigor and systematic validation to ensure credible findings that inform global policy decisions. This principle is vital because AI systems, despite their analytical sophistication, can produce biased or unreliable results without proper validation through established evaluation frameworks. As van den Berg et al. (2021)[10] emphasize, the complexity of UN programmes demands multiple validation layers to ensure AI-generated insights accurately capture diverse population needs and intervention contexts. The stakes are particularly high given the UN's influential role in international development, where UN evaluation findings impact millions of lives. By combining AI analysis with traditional evaluation methods through rigorous validation protocols, the UN both safeguards against algorithmic bias and enhances evaluation quality. This approach ensures AI complements rather than replaces proven methodologies while maintaining the highest standards of evidence-based decision-making in humanitarian and development contexts. AI-generated insights must meet UNEG quality criteria for credibility and reliability which necessitates humans being involved in the validation and interpretation process of the outputs. Results should be triangulated with traditional evaluation methods and validated through established UN external quality assurance processes, that include human oversight before informing decisions.

5. Human Rights-Based AI Implementation

23. The integration of AI in UN evaluations must fundamentally operate within a human rights-based framework that places human dignity and rights at the center of both technological implementation and evaluation processes. This principle is paramount because it directly connects to the UN's foundational mandate and the universal values enshrined in international human rights law. By prioritizing human rights in AI-enabled evaluations, the UN ensures that technological advancement serves its primary purpose of enhancing human welfare rather than becoming an end in itself. This approach requires careful consideration of how AI tools impact fundamental rights such as privacy, non-discrimination and freedom

[9] Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. Information and Organization, 33(3), 100478.
[10] van den Berg, Rob D. "Ethical guidance: Enabling evaluation's role in tackling bad and doing good". Ethics for Evaluation. Routledge, 2021. 17-44.

of expression, while simultaneously leveraging technology to strengthen the UN's ability to assess and promote human rights compliance in its programmes.

24. The principle's importance is magnified in the context of UN evaluations, where findings and recommendations can affect vulnerable populations worldwide. Furthermore, by maintaining human oversight and embedding rights-based considerations in AI systems, evaluations can more effectively contribute to the SDGs while upholding the UN's commitment to leaving no one behind. This balance between technological innovation and human rights protection ensures that AI serves as a tool for advancing human dignity and well-being, rather than potentially compromising the very values the UN seeks to protect and promote.

Conclusions

25. In conclusion, the ethical integration of AI into UN evaluation processes represents both an opportunity and responsibility. While AI offers powerful analytical capabilities to enhance evaluative practices, its implementation must be guided by robust ethical principles that safeguard human rights and dignity. Central to this approach is meaningful stakeholder engagement, particularly from communities in the Global South, to counterbalance potential biases in AI systems and ensure local knowledge is valued. Through structured consultation processes that engage rights-holders, duty-bearers and local communities, the UN can ensure AI enhances rather than undermines participatory evaluation principles.

26. This approach transcends mere technological advancement; it demonstrates how AI can strengthen evidence-based decision-making. Furthermore, it positions UN evaluations at the forefront of innovative yet responsible practices that enhance the organization's ability to assess progress toward the SDGs and respond effectively to emerging global challenges. As the UN system continues to adapt to an increasingly complex evaluation landscape, this ethical framework ensures that AI integration serves its ultimate purpose: strengthening the UN's capacity to promote human rights, foster sustainable development and advance international cooperation through more rigorous, inclusive and impactful evaluations. This methodological evolution, guided by clear ethical principles and grounded in UN values, will enhance both the quality and utility of evaluations while setting global standards for responsible AI use in development evaluation.

References

Arora, A., Barrett, M., Lee, E., Oborn, E., & Prince, K. (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. Information and Organization, 33(3), 100478.

Bravo, L., Hagh, A., Joseph, R., Kambe, H., Xiang, Y., & Vaessen, J. (2023). Machine Learning in Evaluative Synthesis: Lessons from Private Sector Evaluation in the World Bank Group.

Harlow, H. (2018). Ethical concerns of artificial intelligence, big data and data analytics. Kidmore End: Academic Conferences International Limited. Retrieved from http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/conference-papers-proceedings/ethical-concerns-artificial-intelligence-big-data/docview/2116815150/se-2.

Hogenhout, L. (2021). A framework for ethical AI at the United Nations. arXiv preprint arXiv:2104.12547.

Punia, Sanjeev Kumar, et al. "Performance Analysis of Machine Learning Algorithms for Big Data Classification: ML and AI-Based Algorithms for Big Data Analysis." IJEHMC vol.12, no.4 2021: pp.60-75. http://doi.org/10.4018/IJEHMC.20210701.oa4.

Raso, Filippo and Hilligoss, Hannah and Krishnamurthy, Vivek and Krishnamurthy, Vivek and Bavitz, Christopher and Kim, Levin Yerin, Artificial Intelligence & Human Rights: Opportunities & Risks (September 25, 2018). Berkman Klein Center Research Publication No. 2018-6, Available at SSRN: https://ssrn.com/abstract=3259344 or http://dx.doi.org/10.2139/ssrn.3259344.

Saltz, J.S., Dewar, N. Data science ethical considerations: a systematic literature review and proposed project framework. Ethics Inf Technol 21, 197–208 (2019). https://doi.org/10.1007/s10676-019-09502-5.

UNICEF (2022). Artificial Intelligence Strategy Working paper. UNICEF-Global-Insight-policy-guidance-AI-children

United Nations ICT (2022). Office of Information and Communications Technology of the United Nations Secretariat (OICT). "Principles for the Ethical Use of Artificial Intelligence in the United Nations System," High-level Committee on Programmes (HLCP), endorsed by the United Nations System Chief Executives Board for Coordination, Sept 2022.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization.

UNFPA Independent Evaluation Office (2024). GenAI-powered evaluation function at UNFPA.

van den Berg, Rob D. "Ethical guidance: Enabling evaluation's role in tackling bad and doing good." Ethics for Evaluation. Routledge, 2021. 17-44.

Vota, W. (2022, October 26). 10 Ethical Use Principles for Artificial Intelligence in the UN System. ICT works. https://www.ictworks.org/artificial-intelligence-principles-united-nations/

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2021). From human-computer interaction to human-AI Interaction: new challenges and opportunities for enabling human-centered AI. arXiv preprint arXiv:2105.05424, 5.