NETHOPE

NetHope’s Center for the Digital Nonprofit

HUMANITARIAN AI CODE OF CONDUCT

By Elizabeth Shaughnessy, NetHope's Director of Digital Programming

Introduction

The rise of Artificial Intelligence (AI) technologies, in particular those powered by Machine Learning (ML), using Large Language Models (LLMs), and Generative AI, presents a new challenge for humanitarian organizations. Though AI as broadly defined has been around for decades, LLMs and uses of Natural Language Processing (NLP) are creating new opportunities for organizations to make better use of the data they collect, automate processes and create efficiencies, and ultimately (and ideally) improve operations and analysis, and support content generation to improve the lives of those we collectively work to support across humanitarian, development, and conservation environments.

Project ideation and private sector partnerships are already underway at a number of organizations, exploring training models to support analysis of program data, reports, and local contexts. Beyond program work, AI technology has the potential to automate and scale communications and content generation, supporting public fundraising and campaign efforts.

The risks of AI have been well documented, and NetHope has produced a number of resources to help organizations mitigate this risk, including the AI Ethics Train-the-Trainer program and the newly launched Gender Equitable AI Toolkit. Risks include inequity, structural bias, and exclusion or lack of representation, both in reference to the data used to train models to the design of the model itself. A lack of fairness, transparency and accountability also factor into questions around whether a particular AI model is ethical and whether its use is responsible and safe.

These risks can and do have harmful outcomes. Biases in AI have serious impacts on people where it is used for decision-making and access to fair work, education, the economy, and healthcare. Generative AI can and will be used for mis- and disinformation campaigns and threatens an already shrinking digital civic space. How LLMs are trained, with what data, and from whom can also be problematic, with privacy and security issues surrounding personal data collection, use and sharing. AI technologies also tend to be high carbon dioxide producers, with training models producing significant carbon emissions and furthering the impact of climate change for already vulnerable parts of the world.

AI technology has the potential to further entrench inequality, deepen existing divides (including the digital and gender divides), and worsen conflict and fragility. There are opportunities and benefits, and these should be utilized; however, to do so without agreement amongst the sector about how to engage ethically and fairly with this technology risks putting ourselves on the back foot, engaging in a 'race to the bottom', and constantly chasing the next new tool while engaging in potentially dangerous activities that put the people we are meant to support at risk of harm.

We have a unique opportunity in the timeline of AI development to thoughtfully and collaboratively work together to address and mitigate risks, demystify AI technology, and set responsible standards before the use of such technology becomes ubiquitous. By drafting an AI Code of Conduct with and for the sector, we can also support internal, individual efforts by organizations in drafting their own policies and procedures as well as align ourselves better with legislation and governance efforts which are rapidly evolving in both the public and private spheres. Together, we can prepare ourselves and support each other to have robust governance, operations, programming, and advocacy in relation to AI technologies.

Humanitarian AI Code of Conduct

Our use of AI technology and related data practices must:

1. Do No Harm and align with existing humanitarian principles and frameworks.
2. Have a net positive impact on our organizations' missions and for the individuals and communities we serve, and not exacerbate the problems we are working to solve[1].
3. Be fair, inclusive, accessible, and feminist[2]; mitigate bias; and ensure transparency and explainability.

Where there are sector-specific high risk concerns, we agree:

4. Not to use AI to generate photo-realistic images or videos[3] of vulnerable groups, including children and program participants, for the purposes of publication, including campaigning and fundraising.
5. A human will review and, where necessary, contextualize and attribute content generated by or with the help of AI for the purposes of publication.
6. A human will review decisions made or facilitated by AI, in particular where there is risk of real or perceived harm to an individual or community[4].
7. To prioritize safeguarding and child protection and establish any additional guardrails needed where AI is used in these contexts[5].

We will work together to:

8. Collaborate and exchange knowledge and best practices, and collectively support digital literacy and skills in AI across the sector.
9. Align on the approach to due diligence, assessment, and procurement of AI vendors and tools, as well as limits on the commercial use of data.
10. Align where we have different regulatory environments and uphold the highest available standards and practice of data protection, privacy, and security used by model.
11. Understand and address how informed consent manifests in AI contexts and whether it can be meaningfully used as a basis for collecting personal data.
12. Ensure effective governance, accountability, and pathways for redress.
13. Collectively engage with questions and address new risks as the AI landscape evolves[6].

[1] Including inequality, conflict and fragility, and climate change.

[2] An intersectional feminist approach to AI requires an analysis of power and impacts on social justice. As a primer, see this MERL Tech post, the Feminist Principles of the Internet, and Data Feminism.

[3] Or other related forms of content as they emerge, for example voice or audio samples.

[4] Those affected will have a right to human review of AI-made or facilitated decision-making. Further exploration is needed on risk classification as well as feasibility of scale where human review is required.

[5] For example, agreeing to not use facial recognition technology for those under 18 years of age.

[6] For example, on the commodification of data, the ability to opt out, the right to be forgotten, and so on.

Acknowledgements

Gratitude is extended to the support and inspiration behind this document. This includes Elizabeth Shaughnessy, NetHope's Director of Digital Programming, Daniela Weber, Director of The NetHope Center for the Digital Nonprofit, and to NetHope's Member-led AI Working Group. Appreciation is also extended to the founders and partners of NetHope’s Center for the Digital Nonprofit.

Partners of NetHope’s Center for the Digital Nonprofit

avanade box Microsoft okta for good servicenow

NETHOPE

Contact

If you have any questions about this code of conduct, its methods or the data used, contact NetHope's Center for the Digital Nonprofit directly at:
nethopecdn@nethope.org

nethope.org