Comments of the Wikimedia Foundation
To USAID's AI and Global Development Playbook
1 March, 2024

The Wikimedia Foundation, the nonprofit that operates Wikipedia and other free knowledge projects, thanks USAID for the opportunity to provide comments to inform the Agency's AI and Global Development Playbook. We offer feedback based primarily on our experience as an organization responsible for supporting a global effort to make the sum of all human knowledge freely accessible. In that role, we have learned much about both developing technology and working in partnership with diverse communities around the world. We hope our experience can be useful in the context of global development.

2. What applications of AI or AI technologies are most promising for advancing the SDGs and supporting humanitarian assistance? How can these applications be advanced responsibly? Are there any sectors that are particularly well suited to applications of AI? Are there potential limitations or trade-offs that should be considered when applying AI in these contexts?

The Wikimedia Foundation believes that AI and machine learning (ML) tools can advance certain Sustainable Development Goals' (SDGs) objectives when these tools are built in the public interest, consistent with human rights standards, and with input from the communities they are meant to assist through processes that are participatory, transparent, and open. Our position is informed from our experience as the host of crowdsourced projects like Wikipedia—the world's largest freely accessible online encyclopedia and a long-standing community-governed open source initiative. We thank the United States Agency for International Development (USAID) for the opportunity to share the lessons we have learned over the past two decades.

Wikipedia and the other Wikimedia projects directly support multiple SDGs through our commitment to free knowledge, consensus-based community-led processes, and contributions to the global public good. Our vision is to ensure that anyone, anywhere, can participate in the sum of all human knowledge. Wikipedia exists across 300 languages and is read some 15 billion times each month.[1] Along with other Wikimedia projects, it is freely available to everyone, everywhere, supported by charitable donations. Encyclopedic knowledge is added by volunteer editors from around the world, who jointly set content moderation policies, and engage in lively public debates to discuss and improve that content.

[1] Wikimedia Foundation, *Statistics*, https://meta.wikimedia.org/wiki/Statistics (last visited 29 February, 2024).

Wikipedia's commitment to knowledge equity has led to multiple initiatives that seek to correct historical systems of power and privilege that have excluded the voices and knowledge of women and other underrepresented groups from canons of information.[2] These initiatives not only foster freedom of expression (SDG 16.10), but also directly contribute to other various SDGs, such as education (SDG 4), gender equity (SDG 5), and responsive, inclusive, participatory, and representative decision-making (SDG 16.7).

Wikimedia project volunteers have long used AI tools and bots to help scale their own activities, like detecting vandalism or translating content. At the same time, the openly available information on Wikipedia is a key source of training data for other AI tools such as large language models (LLMs) and chatbots.[3] This openness also enables the use of Wikipedia as an essential education tool that teachers around the world use to help people of all ages learn Indigenous languages, critical information and media literacy, and other 21st century skills.[4]

In our experience, involving communities in the design and deployment of AI from the onset ensures a two-way learning process: the affected community members can learn more about the tool and its intended application, and the technology developers can learn about the community's unique contextual needs and challenges. This process also helps to build trust. Participatory co-design of AI tools, exemplified by Wikimedia projects (and discussed further in response to Question 11), offers a promising avenue for advancing the SDGs in a manner that is inclusive, accountable, and also aligns with the well-being and rights of those it aims to serve.

Although the Foundation does not work directly on matters of humanitarian aid efforts, our recommendations are informed by experience with protecting and supporting access to reliable knowledge as well as supporting volunteers who are at-risk and/or located in regions and countries undergoing various kinds of crises. While some AI tools may help to enhance the efficiency and effectiveness of humanitarian aid efforts, there is a risk of prioritizing the concerns of funders over the needs of aid recipients. Said differently, optimizing for efficiency may not always produce the most desirable outcome in the context of humanitarian aid. Human decision makers should be mindful of the tradeoffs associated with AI-powered suggestions. Relatedly, the deployment of AI can

[2] Wikimedia Foundation, *Promoting Knowledge Equity*, https://wikimediafoundation.org/our-work/education/promoting-knowledge-equity/.
[3] Jon Gertner, *Wikipedia's Moment of Truth*, The New York Times Magazine (18 July, 2023) https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html.
[4] Wikimedia Foundation, *Wikimedia in Education*, https://wikimediafoundation.org/our-work/education/; *Teaching 21st Century Skills*, https://wikimediafoundation.org/our-work/education/teaching-21st-century-skills/.

heighten tensions and mistrust between affected communities and service providers. Balancing technological advancement and safeguarding beneficiaries' rights requires careful navigation of issues such as privacy, consent, and technical literacy (discussed further in response to Questions 6 and 10).

3. Relatedly, what are the most risky or harmful applications of AI in Global Majority countries? Why? Can their risks or harms be mitigated, and if so, how?

The Foundation sees the use of AI applications to enhance the efficiency, scale, and effectiveness of orchestrated disinformation and propaganda campaigns as particularly problematic in Global Majority countries. The problems are exacerbated by how data voids and historical information asymmetries intersect.[5] As a reflection of our world, Wikipedia is impacted by systemic inequalities along the lines of gender, race and ethnicity, geography, and wealth in terms of whose knowledge is represented and who participates on the platform. This privileges certain viewpoints not just because of who is actively able to share and police content on Wikipedia, but also because of who is not on the platform to defend and correct the narratives spread. For this reason, Wikipedia encounters the issue of data voids, which are common vectors through which disinformation is spread and content is manipulated. On most platforms, these voids occur when obscure search queries return few results associated, making them ripe for exploitation. On Wikipedia, the vulnerability is similar, but the voids also relate to people (including volunteer community moderators), and are expressed as knowledge gaps that result from the underrepresentation of certain audiences, groups, and cultures as volunteer community moderators and editors.[6] We believe the risks of data voids and knowledge gaps can be mitigated by supporting communities with education and technical tools, and by empowering communities to directly engage in contributing to, as well as curating, the online information ecosystem.

Due to the key role of the Wikimedia projects in the online information ecosystem, their content is a high-value target for powerful actors seeking to promote narratives that support their own economic, political, or geopolitical goals. Wikimedia's community-led model embodies successes as well as pitfalls for countering disinformation. The Foundation's 2020 Human Rights Impact Assessment found that Wikipedia's nonprofit, community-led model acts as a bulwark against harmful content, since it does not

[5] Michael Golebiewski & danah boyd, *Data Voids: Where Missing Data Can Be Easily Exploited*, Data & Society (29 October, 2019) https://wikimediafoundation.org/our-work/education/teaching-21st-century-skills/; Miriam Redi et al., *A Taxonomy of Knowledge Gaps for Wikimedia Projects* (29 January, 2021), https://arxiv.org/abs/2008.12314.
[6] Wikimedia Research, *Programs: Address Knowledge Gaps*, https://research.wikimedia.org/knowledge-gaps.html.

incentivize “viral” content and prioritizes accurate, unbiased information.[7] The same report found that the global community of volunteer editors on Wikipedia effectively controls and improves content systematically via initiatives and tools that they have developed over the last two decades of the platform’s existence.[8] The community-governed approach to curating and contributing verifiable information to the Wikimedia projects works well to combat most forms of disinformation campaigns. However, this model is more vulnerable to manipulation when volunteer editor communities are small or lack diversity.

Supporting people’s ability to engage in community self-governance empowers and makes them more diverse and resilient communities. The Foundation supports volunteers and partner institutions through grantmaking programs to address structural and historical inequalities, and dedicates staff and resources to support global community governance mechanisms.[9] Some of these efforts aim to improve volunteer diversity, others seek to improve the reliability and integrity of the knowledge documented in Wikimedia projects, while others ensure that community hierarchies created to set and enforce rules for behavior and content are well governed and equitable.[10]

Technical tools can also support community efforts to combat disinformation. As AI tools become more sophisticated and ubiquitous, Wikipedia volunteers and others will benefit from having more powerful tools of their own to assist and amplify their efforts to mitigate and counter disinformation attacks. AI and ML tools co-designed by the Foundation and Wikimedia communities help editors to support content integrity, counter vandalism, and scale the ability to monitor content quality in real time.

The lessons we have learned through decades of supporting communities working to advance the mission of freely and openly sharing knowledge with the world can be applied in the context of AI and global development. We urge USAID to consider where

---

[7] Article One & Wikimedia Foundation, *Assessing the Human Rights Impacts of Wikimedia Free Knowledge Projects* (July 2020), https://meta.wikimedia.org/wiki/File:Article_One_Wikimedia_Foundation_July_2020_HRIA_(English).pdf.
[8] Costanza Sciubba Caniglia, *Wikimedia is an antidote to disinformation: Introducing a repository of anti-disinformation projects* (19 October, 2023), https://diff.wikimedia.org/2023/10/19/wikimedia-is-an-antidote-to-disinformation-introducing-a-repository-of-anti-disinformation-projects/.
[9] Lisa Seitz Gruwell, *The power of knowledge for good: New Wikimedia Equity Fund grantees drive change at the intersection of free knowledge and racial justice*, Medium (8 September, 2021), https://medium.com/freely-sharing-the-sum-of-all-knowledge/power-of-knowledge-for-good-ee94e5681a3b.
[10] Wikimedia Foundation, *Open the Knowledge: Promoting radical knowledge equity*, https://wikimediafoundation.org/our-work/open-the-knowledge/; Wikimedia Research, *Research 2030: Knowledge Integrity*, https://meta.wikimedia.org/wiki/File:Knowledge_Integrity_-_Wikimedia_Research_2030.pdf.

opportunities to support and empower communities could produce more equitable outcomes and mitigate some of the risks of global AI deployment.

6. How should data for AI systems be collected, used, stored, managed, and owned to further the SDGs and support humanitarian goals? Which aspects of data management are unique or particularly salient for AI? How should the objective of ensuring sufficient data accessibility for AI training be reconciled with other objectives, such as ensuring privacy protections, in different contexts?

To further the SDGs and support humanitarian goals, data for AI systems should be collected, used, stored, managed, and owned in a manner that prioritizes human rights, ethical principles, and community empowerment.

Examples of the complexity introduced by AI in work supporting the SDGs are the ongoing debates and emerging practices around the need to promote open data sources while safeguarding against the extraction and exploitation of sensitive or vulnerable data. Safeguarding and promoting open data in relation to Indigenous knowledge and traditions aligns with SDG goals to further diversity and education, and are relevant to Wikimedia's goal to address knowledge gaps globally (see our response to Question 3). However,a recent report from the United Nations International Children's Emergency Fund (UNICEF) and Digital Public Goods Alliance (DPGA) has also emphasized the tension between the benefits of open Indigenous language data sets and the risk that the imbalance of available resources for AI development and use will entrench exploitative relationships and further distort historical and cultural understandings.[11]

Pioneering safe data collection, storage, and use frameworks that support open access to data can prevent its abuse and potential harms. These frameworks should: 1) support community-led data governance; 2) require transparency in how data is collected, stored, used, and deleted; and, 3) promote data minimization and other security and rights-enhancing measures.

Consider the example of Wikimedia projects: Their openness supports their status as a global public good, one that is fundamental to the development of the SDGs. Managed correctly, data sources that are publicly available via open data licenses can maximize transparency and reusability. This openness and transparency means that anyone can participate in identifying and addressing knowledge gaps related to gender, ethnicity and/or background that exist on the platforms. Similarly, anyone can leverage this

[11] Digital Public Goods Alliance, *AI Community of Practice Discussion Paper*, (10 August, 2023), https://digitalpublicgoods.net/blog/ai-community-of-practice-discussion-paper/.

repository of culture and heritage for educational purposes like accessing oral histories or teaching endangered languages so that these might live on.[12]

In contrast to the vast stores of public data made available by Wikimedia projects, the Foundation collects very little personal information about people who access or contribute to the projects, and retains it for only a short time. The Foundation advocates data minimization as well as pseudonymization and aggregated or de-identified data, which also improve cybersecurity—we do so in accordance with the United Nations Human Rights Council Resolution on the promotion, protection and enjoyment of human rights on the Internet, which states that technical solutions to secure and protect the confidentiality of digital communications are important to ensure the enjoyment of all human rights offline and online.[13]

The transparency inherent to open source projects also means that the shortcomings of a data set—i.e., biases, incompleteness, and errors—are known and can be addressed. This transparency also improves understanding about how that data has been tagged and modified to develop AI tools.

The Foundation is creating model cards for every ML model hosted on our servers as part of our work to make open source, transparent, human-centered ML tools.[14] Model cards are primary documentation about a model, reporting on the reasons why it was made, proper and improper use cases, and model evaluation scores, especially for marginalized groups.[15] A similar concept exists for datasets.[16]

This kind of awareness is crucial to hold AI developers accountable for the effectiveness and potential harms of their AI tools, especially when they are borrowed from private partners to be used in development contexts. We believe industry actors should lead by example and make their models and data available under open licenses so that international development organizations and partners can iterate and adapt these tools to be better suited for development and humanitarian aims. Incorporating these

[12] Amrit Sufi, *How to help save endangered languages in India – a project on oral culture digitization*, Wikimedia Diff (8 August, 2022), https://diff.wikimedia.org/2022/08/08/how-to-help-save-endangered-languages-in-india-a-project-on-oral-culture-digitization/; Gareth Morlais, *Using technology to promote Welsh Language: Wikipedia*, (7 August, 2017), https://digitalanddata.blog.gov.wales/2017/08/07/using-technology-to-promote-welsh-language-wikipedia/.
[13] United Nations Human Rights Council, *The promotion, protection and enjoyment of human rights on the Internet : draft resolution*, (27 June, 2016), https://digitallibrary.un.org/record/845728?ln=en.
[14] Chris Albon, *apply() Conference 2022 | More ethical machine learning using model cards at Wikimedia*, (31 May, 2022), https://www.youtube.com/watch?v=t4GMq7MC7Js.
[15] Margaret Mitchell et al., *Model cards for model reporting*, (14 January, 2019), https://arxiv.org/abs/1810.03993.
[16] Pushkarna, Zaldivar, and Kjartansson, *Data Cards: Purposeful and Transparent Data Set Documentation for Responsible AI*, (3 April, 2022) https://arxiv.org/abs/2204.01075.

principles into data practices allows AI systems to contribute positively in advancing the SDGs as well as supporting humanitarian efforts that respect the rights and autonomy of affected communities.

10. What other AI-enabling infrastructure or resources are needed to advance responsible AI development and use?

We call USAID's attention to the social resources, not merely to the technical or physical resources, necessary to support AI development. All stakeholders—e.g., elected officials, regulators, policymakers, staff at international development organizations, and citizens around the world—need educational materials and experts to help them understand emerging technologies like AI. Building and maintaining an understanding of the functions and potential impacts of AI presents a challenge even for well-resourced regulatory bodies, since these are a rapidly developing set of technologies, often developed behind closed doors by small groups of elite experts. To be prepared to engage in discussions about the development, use, and regulation of AI tools, people need to understand: 1) how these tools are built; 2) what benefits and risks may come with their deployment and use; and, 3) the tradeoffs associated with various approaches to regulation.

We encourage USAID to support and invest in efforts to help stakeholders at all levels to better understand the technical workings and the policy implications of AI tools. In the international development context, this may be especially important for staff at development organizations and the local communities with whom they engage, so that they are empowered to make informed decisions and better hold private industry partners to account.

We acknowledge that many communities around the world have basic needs that go unmet, lacking fundamentals like access to clean water, shelter, food, and healthcare. We do not suggest that they should be excluded from discussion about the development and use of AI tools, but rather that these communities should be empowered to assess their needs and make collective decisions about development priorities.

15. How should the public be informed about AI risks and harms in your context, and engaged on AI governance issues? What efforts around community engagement seem promising? What communities should be engaged who are not part of existing discussions?

The Foundation employs a combination of risk and impact assessments as well as transparency reports to keep Wikimedia volunteer communities and the public informed about efforts to develop, govern, and use technical tools. For example, the Foundation is finalizing a human rights impact assessment to better understand the opportunities and risks associated with AI and ML technologies within the free knowledge ecosystem; we intend to publish it later this year. We believe it is important to consider the risks and potential impacts of our projects through the lens of human rights, and to communicate to the public our assessment of them. These actions are part of our due diligence to respect human rights, and we suggest that similar steps should be part of AI development and governance efforts globally.

Our commitment to this approach is not new: The Foundation has a history of considering the impacts of technology on free and open knowledge and the humans who work to share it.[17] Our research as well as technology and product development teams publish information about proposed and ongoing projects (including the model cards mentioned in our response to Question 6), information about the development process, and updates as projects progress.[18] Transparency and engagement with stakeholders are central aspects of how the Foundation functions.

In developing AI and ML tools, the Foundation adheres to principles of human-centric design and processes that are open, participatory, and respectful of fundamental rights like privacy.[19] When the Foundation builds AI tools, we do so in consultation with the very people who will be using them. They are our strongest partners in considering the context-specific effects of AI—such as on small language communities. Any AI tools we develop are also open source and transparent, which further empowers our editing communities to adapt these tools to their evolving needs. Working with communities that might be affected by these tools by means of transparent processes helps to safeguard against the potentially negative impact that such technology can have when deployed in a sensitive setting. This is especially important to building trust when many communities and individuals curating and verifying knowledge are indeed part of a setting in which the negative legacies of power, colonialism, extraction, and silencing of specific communities are long-standing.

---

[17] Jonathan T. Morgan, *Designing ethically with AI: How Wikimedia can harness machine learning in a responsible and human-centered way*, Wikimedia Foundation, (18 July, 2019), https://wikimediafoundation.org/news/2019/07/18/designing-ethically-with-ai-how-wikimedia-can-harness-machine-learning-in-a-responsible-and-human-centered-way/.
[18] Wikimedia Apps, *Wikimedia Apps/Team/Android/Machine Assisted Article Descriptions*, https://www.mediawiki.org/wiki/Wikimedia_Apps/Team/Android/Machine_Assisted_Article_Descriptions; *Machine learning models*, https://meta.wikimedia.org/wiki/Machine_learning_models.
[19] Wikimedia Product, *Wikimedia Product/Inclusive Product Development/ Draft Playbook*, https://www.mediawiki.org/wiki/Wikimedia_Product/Inclusive_Product_Development/Draft_Playbook.

We encourage USAID to integrate these values into its playbook for AI and global development, and to ensure that international development projects involving AI tools are transparent and accountable to impacted communities.

16. What are the best ways to improve inclusivity and stakeholder representation in AI design, deployment, governance, or policymaking in the context of global development (at the global, regional, and local levels)?

At the highest level, stakeholder inclusion and participation depend on four elements: 1) capacity, 2) relevance, 3) accessibility, and 4) respect. First, stakeholders must have the time and resources necessary to allow them to take part in AI development processes. Second, the potential impacts of an AI development project must be clear to stakeholders so that they can assess the value of participation. Third, these opportunities for participation must be visible, open to, and understandable for all stakeholders. Fourth and finally, decision-makers must acknowledge, respect, and be responsive to stakeholder contributions and concerns.

We suggest that the following courses of action help to establish a more comprehensive approach to fostering greater representation and participation in dialogues about AI:

Be specific about what kinds of AI applications will be considered. "Artificial Intelligence" is a general term that gives potential stakeholders little information to assess whether their interests may be impacted. Providing more context and information about the intended applications and uses of an AI technology helps people better anticipate and understand these potential impacts. For example, characterizing a consultation as relating to the impacts of "generative AI tools for translation and summarization of text for educational purposes" gives stakeholders far more clarity than does "AI and education."

Provide appropriate venues to discuss a variety of applications. Policymakers may wish to solicit stakeholder perspectives on a broader range of AI tools. However, rather than convening a large cross section of stakeholders with different interest in various AI tools, dividing those consultations into more focused venues will help decision-makers and stakeholders discuss the development, uses, and impacts of AI tools with more precision and nuance.

Ensure that accurate information about opportunities for participation, and participation itself, is available in all relevant languages. This is perhaps especially important at the global level, but remains relevant even at the local level. Language

inclusivity is a key enabler in ensuring equitable accessibility and participation, especially in what concerns Global Majority countries.

17. What are best practices for ensuring human rights are respected and protected in the development, deployment, and use of AI in the context of a risk-based approach to AI governance? Are there mechanisms, processes, and capacity in place to hold actors accountable for harms resulting from AI systems in your context? What should be done to create and operationalize those accountability mechanisms, and ensure their sustainability?

The Foundation identifies, follows, and recommends three sets of best practices. The first involves conducting thorough human rights impact assessments (HRIAs) to identify and mitigate potential risks associated with AI technologies, ensuring that they do not infringe upon fundamental rights such as privacy, freedom of expression, and non-discrimination.[20]

The second is implementing robust accountability mechanisms, transparency measures, and stakeholder engagement processes, which can help foster trust and accountability in AI systems. Twice a year, the Wikimedia Foundation publishes a transparency report on requests received to alter or remove content from the projects, and to provide nonpublic information about users.[21]

Third and last, we follow best practices for involving affected communities throughout the entire design process. Promoting the participation and inclusion of historically marginalized groups across our projects can enhance the comprehensiveness and effectiveness of human rights safeguards in the development and deployment of AI technologies, as well as mitigate the propagation of harmful biases by these tools.

18. Please list any other organizations you think should be consulted as the AI in Global Development Playbook is developed (please note it may not be possible to consult with every organization).

AI Now Institute, Creative Commons, Data & Society, Digital Public Goods Alliance (DGPA), Global Network Initiative (GNI), Open Future Foundation

[20] Article One & Wikimedia Foundation, *Assessing the Human Rights Impacts of Wikimedia Free Knowledge Projects* (July 2020), https://meta.wikimedia.org/wiki/File:Article_One_Wikimedia_Foundation_July_2020_HRIA_(English).pdf.
[21] Wikimedia Foundation, *Transparency Reports*, https://wikimediafoundation.org/about/transparency/.