DATA&
SOCIETY

February 2, 2024

Response to Request for Information (RFI) Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence

Data & Society is an independent, nonprofit research institute studying the social implications of data-centric technologies, automation, and artificial intelligence. Through empirical research, policy, and media engagement, our work illuminates the values and decisions that drive these systems and helps shape futures grounded in equity and human dignity.

The AI ecosystem is at an inflection point: As noted in a recent letter to NIST from the House Committee on Science, Space, and Technology, AI safety tools and methodologies and claims about AI technologies themselves are contested and lack scientific validity.[1] Accordingly, NIST has the important task of building consensus across the AI ecosystem to construct empirically sound, rights-protecting AI safety methods that can be widely adopted.

Our comment argues that the process of achieving consensus to build AI safety and security practices, including transnational and global consensus, must be participatory and include multiple stakeholders, including civil society and historically impacted communities. The process must also be critical and precise about the problems that methods and standards can address.

Specifically, we recommend that NIST:

1. Preserve and expand the sociotechnical focus of the AI Risk Management Framework (AI RMF) for the companion resource for generative AI.
2. Define and address the risks and harms of generative AI in an expansive manner, in recognition of the challenges in mapping, measuring, and mitigating trustworthiness characteristics in generative AI over traditional machine learning approaches.
3. Include a wide swath of professions, skills, and disciplinary expertise to effectively govern generative AI.
4. Take advantage of the benefits of existing assessment practices to identify and govern generative AI risks.
5. Create practical guidance for evaluating and auditing AI capabilities that take into account relationships between stakeholders, with a focus on capabilities and limitations through which AI could cause harm.

[1] US Congress, House of Representatives, Committee on Science, Space, and Technology, "Letter Regarding the Establishment of the Artificial Intelligence Safety Institute (AISI) at NIST", 118th Cong. (2023), https://republicans-science.house.gov/_cache/files/8/a/8a9f893d-858a-419f-9904-52163f22be71/191E586AF744B32E6831A248CD7F4D41.2023-12-14-aisi-scientific-merit-final-signed.pdf

6. Ensure that red-teaming standards are informed by the possibilities and limitations of red-teaming practices.
7. Invest in global alliances for standard-making and actively seek an understanding of similarities and differences in the ways different countries are approaching the challenge of regulating AI, in order to ensure safe, secure, and trustworthy AI across languages, borders, and interests.

1. NIST should preserve and expand the sociotechnical focus of the AI RMF for the companion resource on generative AI.

NIST's AI RMF rightly notes that "AI systems are inherently sociotechnical in nature, meaning they are influenced by societal dynamics and human behavior."[2] To ensure that generative AI systems work as intended and are non-discriminatory and trustworthy, it is essential that all AI systems (including and especially generative AI systems) are governed by risk management and accountability methods that center a sociotechnical approach.

Examples abound of generative AI systems' potential or demonstrated harms that are discriminatory, further social divisions, and/or threaten people's safety.[3] A sociotechnical analysis and approach to AI research, development, deployment, and accountability can mitigate these threats. These methods include:

- **Drawing from observations** gathered through quantitative, qualitative, or mixed methods approaches. Interview-based or ethnographic studies, computational analysis of logged data, sociological audits, case studies, and historical analysis are all employed in

[2] Tabassi, E., Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, [online], (2023), https://doi.org/10.6028/NIST.AI.100-1, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=936225

[3] Electronic Privacy Information Center, "Generating Harms: Generative AI's Impact and Paths Forward", (May 2023), https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf;
Marie Lamensch, "Generative AI Tools Are Perpetuating Harmful Gender Stereotypes," Centre for International Governance Innovation, 14 June 2023, https://www.cigionline.org/articles/generative-ai-tools-are-perpetuatingharmful-gender-stereotypes/;
Center for Countering Digital Hate, "AI Image Tool Midjourney Is Being Used to Generate Conspiratorial and Racist Images," Research + Policy (CCDH, 11 August 2023), https://counterhate.com/research/ai-image-tool-midjourney-generate-racist-and-conspiratorial-images/#about;
Johana Bhuiyan, "WhatsApp's AI Shows GunWielding Children When Prompted with "Palestine,"' *The Guardian*, 3 November 2023, sec. Technology, https://www.theguardian.com/technology/2023/nov/02/whatsapps-ai-palestine-kids-gun-gaza-bias-israel.;
Zachary Small, "Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History," *The New York Times*, 4 July 2023, sec. Arts, https://www.nytimes.com/2023/07/04/arts/design/black-artists-bias-ai.html.;
Alexandre Magueresse, Vincent Carles, and Evan Heetderks, "Low-Resource Languages: A Review of Past Work and Future Challenges" (arXiv, 12 June 2020), https://doi.org/10.48550/arXiv.2006.07264.;
Gabriel Nicholas and Aliya Bhatia, "Lost in Translation: Large Language Models in Non-English Content Analysis," Center for Democracy and Technology (blog), 23 May 2023, https://cdt.org/insights/lost-in-translation-large-language-modelsin-non-english-content-analysis/.

sociotechnical research. Sociotechnical research may also propose theoretical framings that synthesize insights from observational studies or shape future studies.

- **Inductive methods** help to discover the unexpected when technology is deployed "in the real world." These unintended consequences are not necessarily good or bad, but reflect that what a technology becomes in practice is dependent on other actors, including users, who enter the picture only after the tech is deployed.

- **Capturing the viewpoint** of those who are impacted by a technology allows others to have a say in how that technology is used and designed, a critical element in laying the groundwork for meaningful participation in AI governance.

Sociotechnical research methods can expose discrimination,[4] improve system design,[5] and reveal paths to accountability in ways that a purely quantitative approach to evaluating AI systems cannot.[6] When building the companion resource for generative AI, NIST should look to the wealth of sociotechnical research that has been conducted about AI and apply those methodologies and learnings to generative AI.

NIST's AI RMF rightly includes a diverse range of representative actors throughout the AI lifecycle stages, including impact assessors, auditors, governance experts, impacted individuals/communities, socio-cultural analysts, advocacy groups, civil society organizations, and researchers. This diverse group should also be included in the companion resource across the AI lifecycle stages of generative AI. Due to generative AI's many use cases, it is important for the companion resource to encourage that historically marginalized communities and sociotechnical researchers be actively involved in informing the development of generative AI systems before they are deployed, in order to prevent as much harm as possible.

2. **NIST should define and address the risks and harms of generative AI in an expansive manner, in recognition of the challenges in mapping, measuring, and mitigating trustworthiness characteristics in generative AI over traditional machine learning approaches.**

---

[4] Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
[5] Elish, Madeleine Clare, and Elizabeth Anne Watkins. 2020. "Repairing Innovation: A Study of Integrating AI in Clinical Care." Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2020/09/Repairing-Innovation-DataSociety-20200930-1.pdf.
[6] Moss, Emanuel and Watkins, Elizabeth and Singh, Ranjit and Elish, Madeleine Clare and Metcalf, Jacob, *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest* (June 29, 2021). Available at SSRN: https://ssrn.com/abstract=3877437 or http://dx.doi.org/10.2139/ssrn.3877437

Historically, methods meant to identify and mitigate instances of algorithmic harm have been predicated on case-specific paradigms. Generative AI upsets this paradigm, with one multi-purpose model often deployed across a wide array of use cases. Where traditional machine learning approaches can systematically catalog the impacts of a system for a particular setting and set of impacted stakeholders, generative AI — especially "foundation models" — are not tethered to a discrete set of uses and users.[7]

This inherent expansiveness of generative AI necessitates a methodological shift to ensure responsible evaluation. To effectively address this issue, NIST should encourage comprehensive approaches that go beyond technical challenges to include ethical, social, and legal implications — particularly by embracing methods that enable participation by the broader community.

In doing so, NIST should consider the following factors when navigating the responsible deployment of generative AI:

- **Risk identification:** A variety of existing approaches enable broader participation in order to comprehensively identify risks — including, for example, convening representative panels with stakeholders such as ethicists, experts, and representatives from civil society and historically impacted communities.
- **Design and measurement challenges:** Participatory efforts across the stages of mapping, measurement, and design can help experts, end-users, civil society, and impacted communities discuss and identify potential harms across applications of generative AI. Many of these methods already exist — for example, participatory design, questionnaires/surveys, feedback sessions, and workshops. These enable collaborative consideration of ongoing challenges and are especially beneficial when considering the wide array of application areas that generative AI is expected to address.
- **Trustworthiness methods:** Participatory efforts help build consensus with respect to technical and ethical related concerns. Methods such as bug bounties, red-teaming, and crowdsourced audits offer promising ways to engage both experts and the broader community in scrutinizing a system's performance in order to identify potential harms. This external scrutiny, facilitated by multiple stakeholders, allows for a more comprehensive assessment of potential harms by revealing test cases that might have not have been noticed by other groups.
- **Continuous evaluation:** The dynamic nature of generative AI necessitates continuous evaluation that establishes robust feedback loops with end-users, impacted communities, and experts. Methods such as surveys, iterative testing, and convening focus groups can help track evolving system performance and its impacts across diverse perspectives.

[7] Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the Opportunities and Risks of Foundation Models." *arXiv preprint arXiv:2108.07258* (2021).

3. **NIST should include a wide swath of professions, skills, and disciplinary expertise to effectively govern generative AI.**

Generative AI's governance will require methods that can address its technical and social dimensions simultaneously. An absence of diversity in expertise will fail to capture the gamut of the impact of generative AI and its potential harms. Overly technical assessments with no accounting for human experience are not useful; similarly, solely relying on human experience of such systems will make it difficult to engage with the underlying technical choices that produce undesirable outcomes. Thus, many forms of disciplinary expertise must be assembled to work towards a deeper understanding of assessing the impact of generative AI and articulating approaches to govern its use. Relevant experts will range widely and include system developers, data scientists, AI/ML and prompt engineers, cybersecurity professionals, responsible AI practitioners, red-teaming experts, auditors, subject-matter experts (of the domain in which generative AI is deployed), social scientists, economists, ethnographers, lawyers, and labor advocates, among others. Such diverse expertise is necessary to assess the range of potential impacts that the deployment of a generative AI system may have on, for example, existing work practices.

A multidisciplinary approach to assessment and governance must be combined with new modes of public participation.[8] Generative AI has broken barriers around the skill set needed to critically engage with this kind of technology. Cybersecurity culture has always allowed anyone from the public to identify bugs, but technical know-how has been a persistent barrier to participation. With generative AI, because prompts are written in natural language, there is potential for anyone in the broader public to participate in identifying flaws within a system. Hence, there is a need to invest in streamlining reporting, and to develop methods to measure which undesirable outputs merit further attention.

Streamlining mitigation avenues would be useful not only for accountability, but as a resource for researchers interested in the nature of public concerns with generative AI.[9] Given the diversity of generative AI models currently being developed, there is a need for a platform to assess how different models respond to the same prompt, and enable the comparison of potentially problematic responses across models. No such platforms are currently available, but

[8] Gilman, Michele. (2023). *Democratizing AI: Principles for Meaningful Public Participation*. Data & Society Research Institute. https://datasociety.net/library/democratizing-ai-principles-for-meaningful-public-participation/ ; https://medium.com/datasociety-points/shaping-ai-systems-by-shifting-power-ee95f7c3edf9

[9] Friedler, Sorelle and Singh, Ranjit and Blili-Hamelin, Borhane and Metcalf, Jacob and Chen, Brian J. (2023). *AI Red-Teaming Is Not a One-Stop Solution to AI Harms: Recommendations for Using Red-Teaming for AI Accountability*. Data & Society Research Institute. https://datasociety.net/library/ai-red-teaming-is-not-a-one-stop-solution-to-ai-harms-recommendations-for-using-red-teaming-for-ai-accountability/

there are ongoing experiments in developing one, such as the AI Democracy Projects Elections Forum. The synthesis, collation, and curation of reports gathered from the public would produce a crowdsourced dataset of prompts that can be used to evaluate future model behaviors. Efforts to put together such crowdsourced datasets are already underway; a good example is the Adversarial Nibbler project.[10]

Finally, the skills needed to identify patterns in undesirable model outputs have only begun to get industry attention. These skills need to be developed further over time and should also be considered crucial for regulators paying attention to public concerns around generative AI. Such skills will only become more nuanced and specialized with more exposure to the ways generative AI is integrated into industry practices across domains, and there is a need to invest in building training materials that offer a foundation in anticipatory thinking in generative AI governance. We believe that the skills to parse through diverse model outputs to identify the locus of mitigation efforts will come to be increasingly professionalized over the coming years. NIST can play a crucial role in setting up the standards for this work.

4. **NIST should take advantage of the benefits of existing assessment practices to both identify and govern generative AI risks.**

Generative AI poses novel risks that have yet to be identified, which may create the impression that we lack relevant governance approaches. In reality, existing governance methods will better help to identify and assess those novel risks *as they arise*. Assessment documentation, ideally paired with some degree of public transparency, is a key component of both risk identification and risk management. Scholars and practitioners alike have proposed and implemented a variety of documentation practices for other types of AI systems that do not need to be overhauled in order to be constructive in understanding the risks of generative AI. These include:

- **Model cards**: Technical/engineering transparency documentation about how a model has been constructed, what data it was trained on, what its intended uses are, what risks have been identified, and what use cases should be excluded for safety and fairness reasons. Model cards are intended to travel with models as they are moved inside of and between organizations, and are repurposed for different products.[11]

[10] Adversarial Nibbler Project. Dataperf. (2023). https://www.dataperf.org/adversarial-nibbler.
[11] Mitchell, M. et al. "Model Cards for Model Reporting. in Proceedings of the Conference on Fairness, Accountability, and Transparency" - FAT* '19 220–229 (ACM Press, Atlanta, GA, USA, 2019). doi:10.1145/3287560.3287596.

- **Risk cards:** Similar in structure and purpose to model cards, risk cards document specific paths to risks from the use of a language model, including examples of potentially harmful input-output pairs.[12]
- **Datasheets:** Information about datasets that can inform their use by multiple parties, these are particularly useful for open-source datasets used widely across scientific contexts and inside of complex engineering organizations that construct many different datasets. They are used to flag specific types of data that may create human and social risks, like PII or demographic data.[13]
- **Human rights impact assessments:** HRIAs have an established practice in other industries and are a promising pathway to understanding some specific types of harms caused by AI.[14] They are also highly aligned with policy priorities in European AI regulation, which tends to be oriented toward international human rights standards.
- **Algorithmic impact assessments:** AIAs use multidisciplinary expertise to assemble a report about likely impacts to people, communities, and society caused by the introduction of, or changes to, a technical system.[15]

These assessments and documentation do double duty: they both give us handholds for governance of a specific system, and provide an evidentiary base for understanding emergent harms from novel systems. NIST should recognize that these documentation practices are especially powerful when accompanied by public transparency. Users, commercial partners, scholars, and regulators should have access to the information necessary to make informed decisions and effectively track trends across the industry.

5. **NIST should create practical guidance for evaluating and auditing AI capabilities that take into account relationships between stakeholders, with a focus on capabilities and limitations through which AI could cause harm.**

---

[12] Derczynski, L. et al. "Assessing Language Model Deployment with Risk Cards." Preprint at http://arxiv.org/abs/2303.18190 (2023).

[13] Gebru, T. et al. "Datasheets for Datasets" in Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (Stockholm, SE, 2018).

[14] Data & Society Research Institute and European Center for Nonprofit Law. "Recommendations for Assessing AI Impacts to Human Rights, Democracy, and the Rule of Law." https://datasociety.net/wp-content/uploads/2021/11/HUDIERA-Full-Paper_FINAL.pdf (2021);
European Center for Nonprofit Law. Framework for Meaningful Engagement: Human Rights Impact Assessments of AI | ECNL. https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai (2023);
Mantelero, A. & Esposito, M. S. "An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems." Computer Law & Security Review 41, 105561 (2021).

[15] Moss, E., Watkins, E. A., Singh, R., Elish, M. C. & Metcalf, J. *Assembling Accountability: Algorithmic Impact Assessment for the Public Interest.* https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/ (2021).

The biggest challenges in evaluating and auditing algorithmic systems are often not *technical*; rather, they are found in negotiating the sociotechnical relationships between stakeholders, such as auditors, vendors, customers, regulators, and users.[16] Given the practical nature of auditing work, the power and economic relationships between stakeholders is significantly determinative of the utility of auditing for protecting the public good.[17] For example, if an independent auditor needs to certify some aspect of a system for a customer, that auditor must also negotiate data and model access from the developer, something that is rarely codified in contracts nor required by regulations. We suspect that given the opportunity to experiment and iterate upon auditing and assessment techniques, current technical limitations around explainability and interpretability will be resolved through a community of practice. Therefore the meaningful roadblocks to establishing a robust audit ecosystem are largely a matter of economic and organizational relationships: who gets to audit whom and under what circumstances?

As NIST develops guidelines about how to robustly audit and assess generative AI systems, it should pay keen attention to the procedural and relational elements of these practices. While NIST lacks the power to alter contracts or demand system access, it can template the relationships that support auditing and assessment ecosystems. This would facilitate a change in the expectations of all stakeholders.

We also sound a note of caution about reliance on benchmarking as a core AI accountability tool. As critics have pointed out, AI benchmarking has significant construct validity problems — "general" benchmarks applied to systems trained for specialized purposes at best lack utility, and at worst can be highly misleading.[18] While benchmarking can be used to measure relative performance between systems, it tells you little about the appropriateness of the application of that system to its stated purpose, which is where harm actually happens. Reliance on benchmarking as a mechanism to gatekeep access to commercial markets risks furthering the "AI snake oil" problem, where seemingly precise outputs are irrelevant or inappropriate to the task at hand[19] — this is a variation of Goodhart's Law that "when a measure becomes a target, it ceases to be a good measure." Finally, there are many examples of harmful and dubious data incorporated into core AI benchmarking datasets that have shaped the scientific and industrial

[16] Wright, L. et al. "Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability." (2024) doi:10.17605/OSF.IO/UPFDK.
[17] Birhane, A., Steed, R., Ojewale, V., Vecchione, B. and Raji, I. D. "AI Auditing: The Broken Bus on the Road to AI Accountability." Preprint at https://doi.org/10.48550/arXiv.2401.14462 (2024).
[18] Raji, I. D., Bender, E. M., Paullada, A., Denton, E. and Hanna, A. "AI and the Everything in the Whole Wide World Benchmark." Preprint at http://arxiv.org/abs/2111.15366 (2021);
Khan, M. and Hanna, A. "The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability." SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.4217148 (2022).
[19] Narayanan, A. "How to Recognize AI Snake Oil." Arthur Miller Lecture on Science and Ethics (2019);
Raji, I. D., Kumar, I. E., Horowitz, A. & Selbst, A. "The Fallacy of AI Functionality." in 2022 ACM Conference on Fairness, Accountability, and Transparency 959–972 (ACM, Seoul Republic of Korea, 2022). doi:10.1145/3531146.3533158.

uses of this technology, including bigoted labeling in ImageNet, NIST's own FRVT containing mugshots of minors accused of crimes, and the widely used image dataset LIAON-5B containing child sexual abuse material.[20] Reliance on benchmarking risks drawing attention away from these systems' actual conditions of application.

6. **NIST should ensure that AI red-teaming standards are informed by the possibilities and limitations of red-teaming practices.**

Red-teaming has increasingly drawn attention as a method that addresses the risks of AI systems. Red-teamings' roots come from cybersecurity, in which security engineers within a company try to make a system produce undesired results. By identifying undesired results, red-teaming reveals where a system needs to be fixed.

While red-teaming is a valuable tactic to uncover undesirable results in an AI system, it must be used within a suite of tools, including algorithmic impact assessments, external audits, and public consultation, to truly produce an AI ecosystem that is safe, secure, and trustworthy.[21] Due to the traditionally highly technical nature of red-teaming, this approach alone cannot address nor remediate the more complex sociotechnical concerns raised by AI systems.

As NIST sets standards on AI red-teaming, it should explicitly connect AI red-teaming within the broader ecosystem of accountability practices. As noted earlier, because AI systems are inherently sociotechnical, NIST must resist establishing a singular practice as an answer to the complex problems that AI systems, and particularly generative AI systems, bring. Our research indicates that AI red-teaming works best when the flaws sought out are well-defined, when it accompanies transparency, disclosures, and system access for external groups; and when it is part of a broader assessment process. Critically, stakeholders need to commit adequate plans and resources to mitigate identified harms and, if the AI is already deployed, provide paths to redress for those experiencing harms.[22]

7. **NIST should invest in global alliances for standard-making and actively seek an understanding of similarities and differences in the ways different countries are**

[20] Yang, K., Qinami, K., Li, F.-F., Deng and Russakovsky, R. "Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy." https://image-net.org/update-sep-17-2019 (2019); Special Database 18: 3,248 "Mugshots Used for Training Image Recognition Systems." Beautiful Public Data https://www.beautifulpublicdata.com/nist-special-database-18-3-248-mugshots-used-for-training-image-recognition-systems/ (2023).
[21] Friedler, Sorelle and Singh, Ranjit and Blili-Hamelin, Borhane and Metcalf, Jacob and Chen, Brian J. (2023). *AI Red-Teaming Is Not a One-Stop Solution to AI Harms: Recommendations for Using Red-Teaming for AI Accountability*. Data & Society Research Institute. https://datasociety.net/library/ai-red-teaming-is-not-a-one-stop-solution-to-ai-harms-recommendations-for-using-red-teaming-for-ai-accountability/
[22] Id.

approaching the challenge of regulating AI in order to ensure safe, secure, and trustworthy AI across languages, borders, and interests.

Given the preponderance of internet data in training generative AI models, their performance diminishes significantly in languages other than the few that are data-rich, such as English, Spanish, and Mandarin.[23] Along similar lines, practices of evaluating and mitigating potential algorithmic harms, ranging from content moderation[24] to monitoring of generative AI content,[25] has largely focused on English as the primary language of communication. NIST can contribute to advancing responsible global technical standards for AI development by starting with supporting similar standards for model behavior in other languages as they are measured in data-rich languages. This would involve focusing on evaluating whether multilingual models work equally well in languages they support, assessing the accuracy of the translations they offer, and learning from ongoing community-based efforts at NLP research on lower-resourced languages.[26]

The effort to create global standards will require broader ongoing collaboration with stakeholders across the world, and increasing engagements with other national standard-making bodies to build evaluation mechanisms that are robust in a global sense and contextual in a local sense.

This would involve engaging with the methods and interests of the majority world as equally relevant to the process of building standards of AI development within the US. Data & Society's *Primer on AI in/from the Majority World* provides an initial mapping of the conceptual developments in understanding the impact of AI in/from the majority world in its own right, instead of treating it as secondary to the processes of knowledge and technology production in the US.[27] In order to ensure safe, secure, and trustworthy AI across languages, borders, and interests, it is crucial to intentionally learn from and engage with the ways the majority world is approaching the challenges of AI accountability and regulation.

---

[23] Choudhury, Monojit. 2023. "Generative AI Has a Language Problem." *Nature Human Behaviour* 7 (11): 1802–3. https://doi.org/10.1038/s41562-023-01716-4.

[24] Mark Latonero and Aaina Agarwal. 2021. "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar." Carr Center for Human Rights Policy Harvard Kennedy School.

[25] Nicholas, Gabriel, and Aliya Bhatia. 2023. "Lost in Translation: Large Language Models in Non-English Content Analysis." Washington, DC: The Center for Democracy & Technology (CDT): The Center for Democracy & Technology (CDT). https://cdt.org/insights/lost-in-translation-large-language-models-in-non-english-content-analysis/.

[26] Orife, Iroro, Julia Kreutzer, Blessing Sibanda, Daniel Whitenack, Kathleen Siminyu, Laura Martinus, Jamiil Toure Ali, et al. 2020. "Masakhane -- Machine Translation For Africa." *arXiv:2003.11529 [Cs]*, March. http://arxiv.org/abs/2003.11529.

[27] Amrute, Sareeta, Ranjit Singh, and Rigoberto Lara Guzmán. 2022. "A Primer on AI in/from the Majority World: An Empirical Site and a Standpoint." New York: Data & Society Research Institute. https://datasociety.net/library/a-primer-on-ai-in-from-the-majority-world/.

Thank you for the opportunity to inform the direction of NIST's work under the recent AI executive order. NIST is well-positioned to shape AI accountability practices that bring us closer to an AI ecosystem that is safe, secure, and trustworthy. A participatory, multiple-stakeholder approach to consensus-building that employs a suite of AI accountability tools and methodologies is essential to achieve the vision for AI established in EO 14110.

Respectfully submitted,

Serena Oduro, Senior Policy Analyst
Jacob Metcalf, Program Director
Ranjit Singh, Senior Researcher
Briana Vecchione, Technical Researcher
Meg Young, Participatory Methods Researcher