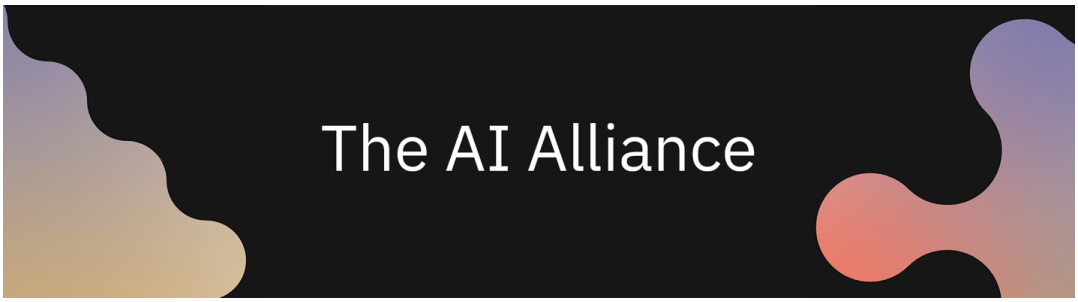

Response to Request for Comment on
"Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights"
Docket No. NTIA–2023–0009
March 27, 2024

Executive Summary

President Biden's "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" (Executive Order 14110) directs the U.S. National Telecommunications and Information Administration (NTIA) to submit a report to the president on the "potential benefits, risks, and implications of dual-use foundation models for which the model weights are widely available, as well as policy and regulatory recommendations pertaining to those models." NTIA seeks public input informing its report recommendations (NTIA–2023–0009).

NTIA's report to the president should find that a thriving ecosystem of widely available foundation model weights is necessary to promote:

A. Innovation, by offering unique privacy and security benefits for AI technologies that will catalyze scientific discovery, free expression, and free enterprise.
B. Economic growth, by democratizing access to productivity-enhancing AI technologies to entrepreneurs and small- and medium-sized enterprises, to underrepresented regions, and by promoting competition; and
C. AI safety, by enabling open, community-driven AI risk management frameworks for building AI safety solutions.

NTIA should implement Executive Order 14110 by partnering with the U.S. National Institute for Standards and Technology (NIST) AI Safety Institute to ensure that any future regulation is shaped by independent, evidence-based research on:

A. Reliable methods of assessing the marginal risks posed by open foundation models;
B. Effective risk management frameworks for the responsible development of open foundation models; and
C. Balancing regulation with the benefits that open foundation models offer for expanding access to the technology and catalyzing economic growth.

Table of Contents

I. Introduction

The AI Alliance is an international community of researchers, developers, and organizational leaders committed to supporting and enhancing open innovation across the artificial intelligence (AI) technology landscape. We are enabling developers and researchers to accelerate responsible, open innovation in AI while ensuring scientific rigor, trust, safety, security, diversity, and economic competitiveness. By bringing together leading developers, scientists, academic institutions, companies, and other innovators, the AI Alliance pools resources and knowledge to address safety concerns and provides a platform for sharing and developing solutions that fit researchers, developers, and adopters around the world.

The content in this filing is provided by the AI Alliance and is not intended to reflect the views of any particular member organization.

We value the opportunity to provide feedback to NTIA about the benefits of widely available foundation model weights. Our comment begins in Section Two by showing how open foundation model weights will play a critical role in advancing scientific discovery, free expression, and free enterprise. Section Three shows how the economic benefits of foundation models will likely accrue most where open foundation model weights are most widely adopted. Both of these sections are most responsive to RFC question #3 (“What are the benefits of foundation models with model weights that are widely available

as compared to fully closed models?") Section Four describes how community-driven risk management frameworks are addressing foundation model risks. This section is most responsive to RFC question #5 ("What are the safety related or broader technical issues involved in managing risks and amplifying benefits of dual-use foundation models with widely available weights?").

A. Defining "Open" AI.

The open-source community broadly agrees that software-focused definitions of "open-source" licensing need modification to be effectively applied to AI models or systems.[1] Because dialogue about defining "open-source AI" remains ongoing, we believe it is premature to adopt a formal definition of "open-source AI" here.[2]

In lieu of a formal definition, we use the words "open" and "openness" to refer to conditions in which foundation model weights are publicly available under a permissive license that allows for research and commercial use. This definition encompasses the AI models most relevant to NTIA as it drafts its recommendations to the president.

II. Many of the foundation models that will serve as the next medium for scientific discovery, free expression, and free enterprise will be open.

A. Open foundation models are advancing scientific discovery and free enterprise.

Across multiple scientific disciplines, open foundation models are accelerating fundamental research and showing promise for enabling future breakthroughs. One open foundation model, GHP-MOFassemble, is accelerating research into improving the effectiveness of carbon capture technology.[3] Another model, the AI Foundation Model for Earth Observations, promises to speed up the analysis of satellite images and boost climate-related discoveries.[4] Another model, ESM-Fold, has been used to accelerate research on increasing the effectiveness of COVID-19 antibodies.[5] A breakthrough open foundation model capable of universal representation of living cells, Universal Cell Embeddings, has

[1] *E.g.*, Open Source Initiative, *Why do we need a new Definition of Open Source just for AI?*, https://opensource.org/deepdive ("The traditional view of Open Source code and licenses when applied to AI components are not sufficient to guarantee the freedoms to use, study, share and modify the systems.").
[2] *See* Megan Morrone, *"Open" software needs an AI rethink*, Axios, Feb. 15, 2024 https://www.axios.com/2024/02/15/open-source-ai-definition-openai-meta ("the people and companies creating today's most advanced AI models don't even agree on what 'open' AI means.")
[3] *See* Rob Mitchum, *Researchers generate a carbon capture breakthrough using AI, physics and supercomputers*, UIC Today, Feb. 14, 2024, https://today.uic.edu/researchers-generate-a-carbon-capture-breakthrough-using-ai-physics-and-supercomputers/; Park, H., Yan, X., Zhu, R. et al. A generative artificial intelligence framework based on a molecular diffusion model for the design of metal-organic frameworks for carbon capture. Commun Chem 7, 21 (2024). https://doi.org/10.1038/s42004-023-01090-2. The GHP-MOFassemble framework relies on an open foundation model called DiffLinker available at https://github.com/igashov/DiffLinker.
[4] *See IBM and NASA open source the largest geospatial AI foundation model on Hugging Face*, IBM, Aug. 3, 2023, https://research.ibm.com/blog/nasa-hugging-face-ibm ("IBM is now making its foundation model public … It's the largest geospatial model to be hosted on Hugging Face and the first open-source AI foundation model NASA has collaborated to build … it can analyze geospatial data up to four times faster than state-of-the-art deep-learning models, with half as much labeled data, IBM has estimated.").
[5] *See* Hie, B.L., Shanker, V.R., Xu, D. et al. Efficient evolution of human antibodies from general protein language models. Nat Biotechnol 42, 275–283 (2024). https://doi.org/10.1038/s41587-023-01763-2

discovered new cell types and functions.[6] Dr. Eric Topol, the director of the Scripps Research Translational Institute, told the *New York Times* in reaction to Universal Cell Embeddings that a "vital discovery about biology that otherwise would not have been made by the biologists — I think we're going to see that at some point."[7] More than 90 leading scientists recently signed a statement affirming the belief that "the benefits of current AI technologies for protein design far outweigh the potential for harm," noting that "many researchers in our community benefit from open-source scientific software, which has enabled rapid innovation and broad collaboration."[8]

Open foundation models are beginning to drive the next wave of innovation for business. Furniture company Wayfair is using an open foundation model, Stable Diffusion, to help customers generate new images and visualize redecorated rooms.[9] Cloud computing company VMWare is using an open foundation model, StarCoder, for code generation assistance.[10] Shopify is using Llama 2 to create product descriptions and marketing content.[11] Forbes recently surveyed 600 business owners using or planning to use AI and found that many were planning on using AI for customer relationship management, digital personal assistants, inventory management, content production, product recommendations, accounting, supply chain operations, recruitment and talent sourcing, and audience segmentation.[12]

B. Openness has unique privacy and security benefits.

Community-driven development of open foundation models can drive state-of-the-art privacy and safety improvements.[13] One research paper, *MasterKey: Automated Jailbreak Across Multiple Large Language Model Chatbots*, shows how widely available model weights enable scrutiny of model vulnerabilities and drive security improvements for closed and open models alike.[14] Another paper shows similar benefits for data privacy.[15] This is because, as one Brookings Institution report notes, open models "can be cross-examined and interrogated for bugs or possible improvements," resulting in "collaborative development and an engaged community" that can create "accessible, robust, and high-quality code."[16]

---

[6] *See* Universal Cell Embeddings: A Foundation Model for Cell Biology Yanay Rosen, Yusuf Roohani, Ayush Agarwal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R. Quake, Jure Leskovec bioRxiv 2023.11.28.568918; doi: https://doi.org/10.1101/2023.11.28.568918 (preprint). Model weights available at https://github.com/snap-stanford/UCE.
[7] Carl Zimmer, *A.I. Is Learning What It Means to Be Alive*, The New York Times, Mar. 10, 2024, https://www.nytimes.com/2024/03/10/science/ai-learning-biology.htm
[8] Responsible AI x Biodesign, *Community Values, Guiding Principles, and Commitments for the Responsible Development of AI for Protein Design*, Mar. 8, 2024, https://responsiblebiodesign.ai/.
[9] *See* Belle Lin, *How Did Companies Use Generative AI in 2023? Here's a Look at Five Early Adopters*, The Wall Street Journal, Dec. 29, 2023, https://www.wsj.com/articles/how-did-companies-use-generative-ai-in-2023-heres-a-look-at-five-early-adopters-6e09c6b3.
[10] *See* Matt Marshall, *How enterprises are using open source LLMs: 16 examples*, VentureBeat, Jan 29, 2024, https://venturebeat.com/ai/how-enterprises-are-using-open-source-llms-16-examples/
[11] *Id*.
[12] Katherine Haan, *How Businesses Are Using Artificial Intelligence In 2024*, Forbes, Apr. 24, 2023, https://www.forbes.com/advisor/business/software/ai-in-business/.
[13] Evidence about the brittleness of closed model security casts some doubt on the argument that closed models are necessarily more resistant to malicious misuse than open models. *See* Arvind Narayanan and Says Kapoor, *AI safety is not a model property,* AI Snake Oil, Mar. 12, 2024, https://www.aisnakeoil.com/p/ai-safety-is-not-a-model-property.
[14] *See* https://arxiv.org/abs/2307.08715.
[15] *See* https://arxiv.org/abs/2403.04801.
[16] *See* Alex Engler, *How open-source software shapes AI policy*, Brookings Institution, Aug. 10, 2021, https://www.brookings.edu/articles/how-open-source-software-shapes-ai-policy/ ("OSS enables and increases AI adoption by

Twenty-five leading AI experts across industry, academia, and civil society agree that "model weights are essential for several forms of research across AI interpretability, security, and safety" and that "model weights enable external researchers, auditors, and journalists to investigate and scrutinize foundation models more deeply."[17]

To put it succinctly, as more than 1,200 people across the field affirmed in a *Joint Statement on AI Safety and Openness*, if "our objectives are safety, security and accountability, then openness and transparency are essential ingredients to get us there."[18]

C. Openness protects freedom of expression.

Access to foundation model weights decentralizes control over core model features that can affect access to accurate information and freedom of expression. Freedom House found in its recent report, *The Repressive Power of Artificial Intelligence*, that authoritarian governments are building centralized foundation models that limit access to accurate information and embed censorship.[19] Open model weights, by contrast, decentralize control over a foundation model's knowledge and better guarantees that each developer can moderate content for the developer's customers or community.[20] This approach is comparable to the decentralized, open approach taken for the technical infrastructure behind the internet.[21] And as the U.S. Department of State, Bureau of Democracy, Human Rights and Labor recently highlighted, the internet's open technical infrastructure plays a crucial role in thwarting authoritarian governments seeking to advance centralized control over the Internet for censorship and surveillance.[22]

III. The economic benefits of AI will likely accrue most where open foundation models are most widely available.

A. Foundation models will catalyze economic growth.

Foundation models will increase worker productivity and grow economies. According to Goldman Sachs Research, the widespread adoption of foundation models could increase U.S. productivity by 1.5% on an annual basis.[23] Another report, from the International Monetary Fund, *The Macroeconomics of Artificial*

---

reducing the level of mathematical and technical knowledge necessary to use AI. … Since OSS code is all public, it can be cross-examined and interrogated for bugs or possible improvements. With collaborative development and an engaged community, as often arises around popular OSS, this collaborative-competitive environment can frequently result in accessible, robust, and high-quality code.")

[17] *See* https://arxiv.org/abs/2403.07918.

[18] *See Joint Statement on AI Safety and Openness*, Mozilla, Oct. 31, 2023, https://open.mozilla.org/letter/

[19] *See* Allie Funk, Adrian Shahbaz, and Kian Vesteinsson, *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence*, Freedom House, Oct. 3, 2023, https://freedomhouse.org/sites/default/files/2023-11/FOTN2023Final.pdf.

[20] *See* Mark Gimein, *AI's Spicy-Mayo Problem*, The Atlantic, Nov. 24, 2023, https://www.theatlantic.com/ideas/archive/2023/11/ai-safety-regulations-uncensored-models/676076/.

[21] *See* Steven Vaughan-Nichols, *Open source is actually the cradle of artificial intelligence. Here's why*, ZDNet, https://www.zdnet.com/article/why-open-source-is-the-cradle-of-artificial-intelligence/.

[22] *See* Bureau of Democracy, Human Rights, and Labor, *Funding Opportunity Announcement: Supporting Critical Open Source Technologies That Enable a Free and Open Internet*, U.S. Department of State, Feb. 21, 2023, https://www.state.gov/supporting-critical-open-source-technologies-that-enable-a-free-and-open-internet/.

[23] Goldman Sachs, Research, *AI may start to boost US GDP in 2027*, Nov. 7, 2023, https://www.goldmansachs.com/intelligence/pages/ai-may-start-to-boost-us-gdp-in-2027.html. ("Generative artificial

Intelligence, acknowledges that AI may "end up being less promising or less ready to bring to market than initially hoped," but also forecasts that "AI might be applied to a substantial share of the tasks done by most workers … and massively boost productivity in those tasks."[24] That is no small outcome. As the IMF notes, productivity is the single largest determinant of "the wealth of nations and the living standards of their people."[25] Overall, according to an estimate by the consulting firm McKinsey, foundation models may generate between $2.6 trillion to $4.4 trillion in economic growth across the global economy.[26]

B. Openness brings foundation models to entrepreneurs and small- and medium-businesses.

Open foundation models are often the most affordable, cost-effective option for entrepreneurs and small- and medium-businesses. Building a new, enterprise-specific foundation model often requires prohibitively expensive investments in model training.[27] Access to open foundation models, which are typically free to procure and more affordable to customize than starting from scratch, substantially lowers the barrier to entry.[28] These models also enable a thriving ecosystem of foundation model development support and cloud service providers serving open foundation models to enterprise customers. This makes it possible for more businesses to build foundation models at lower cost, ensuring that corporate resources are not the sole determinant of whether a company can realize the benefits of a bespoke foundation model.[29] This spreads the productivity benefits of foundation models to more sectors of the economy.

C. Open foundation models promote competition and choice.

Openness creates increased competition in the foundation model marketplace by enabling downstream developers to build innovative, custom products.[30] Growing the number of foundation model-based products reduces overall market concentration and increases options for enterprise customers and end

---

intelligence has the potential to automate many work tasks and eventually boost global economic growth … In the baseline scenario, the Goldman Sachs Research economists estimate AI could increase US productivity growth by 1.5 percentage points annually assuming widespread adoption over a 10-year period.")

[24] Erik Brynjolfsson and Gabriel Unger, The Macroeconomics of Artificial Intelligence, International Monetary Fund, Dec. 2023, https://www.imf.org/-/media/Files/Publications/Fandd/Article/2023/December/20-25-Brynjolfsson-final.ashx.

[25] *Id*.

[26] *See* McKinsey & Company, *The Economic Potential of Generative AI: The Next Productivity Frontier*, Jun. 2023, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.

[27] *See* https://arxiv.org/abs/2403.07918.

[28] One exception is that certain use cases, like customer service chatbots, are amenable to the fine-tuning options made available by closed model providers. But realizing many of the business benefits of foundation models depends on access to pretrained model weights. E.g., MosaicML, *Stardog: Customer Spotlight*, Dec. 19, 2023 https://www.mosaicml.com/blog/stardog-customer-spotlight ("we quickly learned that if you have a smaller model that you fine-tune for a specific task, you can match or exceed the quality of [a closed model] and you have more control over data security and privacy. … Fine-tuning is not really a nice-to-have for us. It's a necessity.").

[29] *See* Oguz A. Acar and Andrés Gvirtz, GenAI Can Help Small Companies Level the Playing Field, Harvard Business Review, Feb. 1, 2024, https://hbr.org/2024/02/genai-can-help-small-companies-level-the-playing-field.

[30] *See* Will Douglas Heaven, *The open-source AI boom is built on Big Tech's handouts. How long will it last?*, MIT Technology Review, May 12, 2023, https://www.technologyreview.com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/ ("if the trend toward closing down access continues, then not only will the open-source crowd be cut adrift—but the next generation of AI breakthroughs will be entirely back in the hands of the biggest, richest AI labs in the world.").

users.[31] The availability of open foundation models is also likely applying market pressure on closed developers to lower prices and "compete against free."[32] This has wide benefits; in general, as the White House has noted, "when firms have to compete for customers, it leads to lower prices, higher quality goods and services, greater variety, and more innovation."[33]

D. Openness brings foundation models to underrepresented regions.

Openness can bring foundation model development to underrepresented regions, including the Global South. By publication index (h-index), a widely-used measure of research impact, Africa and Latin America trail the Global North on AI research and development.[34] According to one analysis, none of the top 100 most-cited companies or universities for AI research were based in Africa or Latin America.[35] Open models can help close the gap. Unlike closed models, which support select languages and are tailored to select audiences, open models empower members of the Global South to build and localize models that speak their own language and understand their own culture.[36] As one study found, Africa-based "technologists note that their solutions often perform better than tools from large multinational companies, simply because they, the technologists, can speak the language."[37]

IV. Community-driven risk management frameworks are addressing foundation model risks.

A. Community-driven, multi-stakeholder efforts are advancing foundation model safety.

Openness is not a one-size-fits-all solution and risk management is an important aspect of the responsible development and release of a foundation model.[38] That's why stakeholders across industry, academia, civil society, and government are working together to build AI risk assessment and mitigation

[31] *See* https://arxiv.org/abs/2403.07918 (access to model weights promotes innovation in downstream markets by "helping to reduce market concentration at the foundation model level from vertical cascading").
[32] *See* Tyler Cowen, *Open-Source Software Is Worth a Lot More Than You Pay for It*, Bloomberg, Feb. 26, 2024 https://www.bloomberg.com/opinion/articles/2024-02-26/open-source-software-is-worth-a-lot-more-than-you-pay-for-it.
[33] *See* Heather Boushey and Helen Knudsen, *The Importance of Competition for the American Economy*, U.S White House Blog, Jul. 9, 2021, https://www.whitehouse.gov/cea/written-materials/2021/07/09/the-importance-of-competition-for-the-american-economy/.
[34] Many factors may explain this divide, including historical inequities, higher costs of internet access, and underrepresentation in available training data for model training. *See* https://arxiv.org/abs/2102.01265.
[35] *See* https://arxiv.org/abs/2102.01265
[36] For example, with adequate training data, the BigScience Large Open-science Open-access Multilingual Language Mode (BLOOM) model can enable developers to more affordably fine-tune for a new language. *See* https://arxiv.org/abs/2212.09535 ("we adapt BLOOM models to support eight new languages (German, Russian, Bulgarian, Thai, Turkish, Greek, Korean, and Guarani) in the resource-constrained settings…"). *See also* MakerereNLP, *Text & Speech for East Africa*, https://www.masakhane.io/ongoing-projects/makererenlp-text-speech-for-east-africa ("The project aims to deliver open, accessible and high quality text and speech datasets for low resourced East African languages from Uganda, Tanzania and Kenya.").
[37] *See* Andrew Paul, *AI programs often exclude African languages. These researchers have a plan to fix that*, Popular Science, Aug. 11, 2023, https://www.popsci.com/technology/african-language-ai-bias/; Kathleen Siminyu et. al.,. *Consultative engagement of stakeholders toward a roadmap for African language technologies*, Patterns, https://doi.org/10.1016/j.patter.2023.100820.
[38] Seger, Dreksler, Moulange, Dardaman, Schuett, Wei, et al, *Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives*, Centre for the Governance of AI, 2023 ("If models are determined to pose significant threats, and those risks are determined to outweigh the potential benefits of open-sourcing, then those models should not be open-sourced. … This is not to say that a given highly capable model should never be open-sourced. Expected model impacts are likely to change with increasing societal resilience and development of new defensive techniques. … The need to conduct risk assessments prior to model release seems to be generally accepted").

frameworks. Our multi-stakeholder organization, the AI Alliance, is already collaborating on establishing improved, community-driven model evaluation testing frameworks and mitigations. Our members are also leading other important safety efforts. For example, ML Commons has established an AI Safety Working Group focused on the development of safety benchmarks for certain foundation models.[39] The Partnership on AI has been growing the list of signatories to its Responsible Practices for Synthetic Media, a framework for how to responsibly develop, create, and share AI-generated media.[40] The Partnership on AI has also been maturing its Guidance for Safe Foundation Model Deployment, a framework for model providers to responsibly develop and deploy AI models.[41] Hugging Face is leading a multi stakeholder collaboration on building a framework for assessing the social impact of foundation models.[42]

B. Community-driven risk management frameworks are keeping pace with AI advancements.

Community-driven risk management frameworks are supplementing government approaches with solutioning that more closely track the pace of foundation model development. For example, the UC Berkeley AI Risk-Management Standards Profile for General-Purpose AI Systems (GPAIS) and Foundation Models proposed a comprehensive adaptation of the U.S. National Institute for Standards and Technology (NIST) AI Risk Management Framework (RMF) to foundation models after the NIST AI RMF did not address it.[43]

Open foundation models inform and validate community-driven risk assessment and risk management solutions. For example, the non-profit Humane Intelligence and the United Kingdom's national academy of sciences, the Royal Society, organized an event with 40 experts in climate science and disease that depended on using the Llama 2 model for a risk assessment exercise.[44]

V. Conclusion

The AI Alliance values this opportunity to highlight the benefits of widely available foundation model weights. We look forward to additional opportunities to show how open science and open innovation is important to realizing many of the benefits of AI advancements.

[39] *See MLCommons Announces the Formation of AI Safety Working Group*, ML Commons, Oct. 26, 2023, https://mlcommons.org/2023/10/mlcommons-announces-the-formation-of-ai-safety-working-group/.
[40] *See* Responsible Practices for Synthetic Media, Partnership on AI, https://syntheticmedia.partnershiponai.org/.
[41] *See* Guidance for Safe Foundation Model Deployment, Partnership on AI, https://partnershiponai.org/modeldeployment/.
[42] *See* https://arxiv.org/abs/2306.05949.
[43] *See* Center for Long-Term Cybersecurity, *AI Risk-Management Standards Profile for General-Purpose AI Systems (GPAIS) and Foundation Models*, University of California, Berkeley, Nov. 8, 2023, https://cltc.berkeley.edu/seeking-input-and-feedback-ai-risk-management-standards-profile-for-increasingly-multi-purpose-or-general-purpose-ai/
[44] *See Billy Perrigo,* The Scientists Breaking AI to Make It Safer, Time, Oct. 26, 2023, https://time.com/6328851/scientists-training-ai-safety/.