Before the
National Telecommunications and Information Administration
Washington, DC 20230

In the Matter of	)	
	)	
Dual-Use Foundation Artificial Intelligence	)	Docket No. 240216-0052
Models with Widely Available Model	)	
Weights	)	
	)	

COMMENTS OF NEW AMERICA'S OPEN TECHNOLOGY INSTITUTE

March 27, 2024

Prem M. Trivedi
Nat Meysenburg
New America's Open Technology Institute
740 15th Street NW, Suite 900
Washington, D.C. 20005

Executive Summary

New America's Open Technology Institute welcomes the opportunity to submit comments in response to the National Telecommunications and Information Administration (NTIA)'s request for public submissions on the risks, benefits, and policy and regulatory approaches relating to "Dual-Use Foundation Artificial Intelligence Models with Widely Available Weights." Our comments emphasize the importance of policy and regulatory approaches that encourage the development of a strong open-source AI ecosystem in the United States. We take a broader framing of "openness" than is captured by the phrase "models with widely available weights."

Section II briefly discusses terminology and definitions related to openness and risk, noting that there is no decisive definition of open or closed, and that the focus should be on defining the marginal risks posed by more open systems.. Section III briefly situates the current debate in the context of the history of open-source software development and discusses the relationship between open code and open societies. Section IV analyzes the benefits and risks of more open models relative to more closed ones along **five policy objectives**: (a) cybersecurity and related harms to people; (b) foreign policy, through the prism of geopolitics and geoeconomics, (c) transparency and public accountability, (d) economic health and innovation, and (e) community control and benefits. Section V offers four principles and recommendations for U.S. government and related stakeholders developing policy and regulatory approaches to open models: (1) Study marginal risk and articulate harms with specificity, (2) create common policy and regulatory requirements that apply to foundation models no matter where they fall on the spectrum of openness, (3) consider the broad range of relevant national security and foreign policy objectives before recommending sweeping policy or regulatory action aimed either at vaguely defined or very specific defined security risks, and (4) develop a thoughtful approach to cybersecurity software liability that accounts for the need to incentivize innovation in open AI models.

I. Introduction

The Open Technology Institute ("OTI") is a program within New America that works at the intersection of technology and policy to ensure that every community has equitable access to digital technology and its benefits.[1] OTI promotes universal access to digital technologies that are both open and secure, using a multidisciplinary approach that brings together advocates, researchers, organizers, and innovators. OTI welcomes the opportunity to respond to the National Telecommunications and Information Administration's (NTIA) request for comments about the risks and benefits of foundation AI models with widely available weights.[2] Our comments are largely responsive to aspects of questions 1, 2, 3, 5, and 8.

[1] *New America*, Open Technology Institute, https://www.newamerica.org/oti/about/.

[2] *Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights*, Docket No. 240216-0052, 89 FR 14059, National Telecommunications and Information Administration (hereafter "NTIA RFC") (Feb. 26, 2024).

This call for comments and public debate is timely, as there has been a rise of closed AI models in the United States that largely reflect existing concentrations of power in the hands of a few leading companies. This trend is a cause for concern, as it risks extending the dynamics we now experience with large technology companies to a period of rapid and transformative evolution in foundation AI models. It would be a mistake for the United States to adopt a policy approach to foundation AI models that hampers the chance for large and small open source systems to thrive alongside large companies' more closed models. At a minimum, the United States should adopt common guardrails and rules requiring meaningful transparency and accountability from *all* AI foundation models, regardless of where they fall on the complicated spectrum of open to closed.

At OTI, we favor an ecosystem in which open AI models can flourish alongside proprietary ones. Our comments analyze the benefits and risks in the context of five major U.S. policy objectives: cybersecurity, foreign policy, transparency and public accountability, economic health and innovation, and community control and benefits. Although open AI models—just like closed models—present risks, the many benefits of open models play an essential role in furthering these five objectives.

This week, OTI joined 46 other civil society organizations and academics in urging Secretary of Commerce Gina Raimondo to protect AI openness and transparency. Ensuring that open and transparent AI models flourish is critical to developing trustworthy AI applications that can bolster American innovation, global competitiveness, and an equitable AI future for all Americans. In order to counter or at least offset the trend toward dominant closed AI systems and continued concentrations of power in the hands of a few companies,[3] The U.S. government should take a coordinated interagency approach designed to ensure that the vast potential benefits of a flourishing open model ecosystem serve American interests, broadly defined.

II. The Gradient of Openness and the Question of Marginal Risk

There is no easy binary that opposes "open" and "closed" AI models. Commentary and research that suggest otherwise unhelpfully distort the reality—which AI technical and governance experts have repeatedly explained—that AI models sit somewhere on a spectrum or "gradient" of openness.[4] We commend NTIA for forthrightly acknowledging this gradient and focusing on the

[3] Mark Surman, Ayah Bdeir, Lindsey Dodson, Alexis-Brianna Felix, and Nik Marda, Accelerating Progress Toward Trustworthy AI, Mozilla, Feb. 22, 2024.

[4] See, *e.g.*, The Future Society, Toward Effective Governance of Foundational Models and Generative AI ("Several speakers challenged the notion of a binary between "open" and "closed" models, pointing toward a spectrum of options regarding the level of access to system components such as datasets, code, model cards, and model weights.")

need to discuss marginal risks associated with open models relative to closed models or what is already publicly available online.[5]

To put it plainly, “open” and “open source” have many meanings, with different actors using the term in different, often self-serving, ways. As Widder, West, and Whittaker noted in a paper last year:

> Broadly, the terms ‘open’ and ‘open source’ are used in the context of AI in varying ways to refer to a range of capabilities that can be broadly bucketed as offering attributes of *transparency* - the ability to access and vet source code, documentation and data - *reusability* - the ability and licensing needed to allow third parties to reuse source code and/or data - and *extensibility* - the ability to build on top of extant off-the-shelf models, ‘tuning’ them for one or another specific purpose. While the terms ‘open’ and ‘open source’ are used variably to refer to these attributes, in practice there are gradients of openness that offer vastly differing levels of access. Some systems described as ‘open’ offer little more than an API or the ability to download a hosted model. In these cases, many question whether the term should be applied at all, or if it is ‘openwashing’ systems that should be understood as ‘closed’. Other more maximal versions of ‘open’ AI go further, offering access to the source code, underlying training data and full documentation, as well as licensing the AI system for wide reuse under terms aligned with the mandates of the Open Source Initiative’s definition of ‘open source’.[6]

In the context of foundation AI models, openness can manifest in many ways, including:

- Open code that can be downloaded and used by others for training and application to new contexts
- Open model weights
- Open licenses for model use
- Information about model inputs (data sources, training, methodology)
- Information about and plans to mitigate against undesirable downstream effects (e.g., malicious actors fine-tuning the model to cause clear harms).

[5] Remarks of NTIA Administrator Alan Davidson, Mar. 22, 2024, https://www.ntia.gov/speechtestimony/2024/national-security-and-open-weight-models (“One thing we have already learned is the importance of focusing on the marginal or differential risks and benefits of open weights. For example, we need to measure the risks of open-weight models relative to the risks that already exist today from widely-available information, or from closed models. We have also been encouraged to hear that this is not a binary choice of “open” vs. “closed.” Rather there is a broader “gradient of openness” that we need to consider and that may offer broader options for policy.”)

[6] David Gray Widder, Sarah West, and Meredith Whittaker, Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, at 2-3, August 17, 2023.

To plot whether a model is open or closed along just these five factors is challenging, to say the least. To illustrate the concept of the gradient of openness—and how difficult it is to try to situate models along it—we offer simplified breakdowns, with examples, of four categories. Much more fluidity on the gradient is possible than these four categories represent; we have chosen them as an exercise in illustrative line-drawing.

Table 1: Illustration of the Gradient of Openness

	Attributes	**Foundation Model Example(s)**
Open	Open code + transparency about model inputs (sources, training, methodology etc.) + transparency about downstream effects + published model weights	Mixtral, LLaMA[7]
	Open code + no transparency about sources and training + no published model weights	Grok,
↑	Closed code + no open model weights + public access + some transparency about model inputs and downstream effects	ChatGPT, Gemini
Closed	Closed code + proprietary access + perhaps no transparency about model inputs or downstream effects	Many companies' internal systems (e.g., likely for warehouse management and distribution)

Throughout these comments, we use "open models" to refer generally to the top two rows, but our preference is for models that embrace the many aspects of transparency in the topmost row. But because of definitional issues around "open" and "open source" that are vital but unlikely to soon be decisively resolved, it is more helpful to talk about the benefits and risks of models toward the open end of the spectrum than those at the closed end. In light of these definitional difficulties and debates, we think the most productive use of policy makers and regulators' effort is to develop a set of common rules and technical standards that would apply across all foundation models (see Section V below for OTI's recommendations).

III. Historical context: Open societies and the ideal of open-source software

While we do not argue for a reductive one-to-one equation of open-source software (OSS) development or open AI models with attributes of open societies, the relationship between the design and governance of open software/models and the ethos of open societies is worth

[7] The difficulties of trying to plot just a few foundation models on a simplified chart (for instance, Mixtral vs. LLaMA) perfectly makes the point that regulatory policy should not be based on parsing these differences.

examining. While open and closed models have specific, if contested, meanings in the context of software, they also have some broader resonance in the context of politically open and closed societies. In a seminal 1999 article Open Code and Open Societies, Lawrence Lessig describes the "Open Source Software Moment" taking shape at the time.[8] Lessig lays out the ideal of open source:

> The idea that through this collective, essentially volunteer, effort, one of the most powerful operating systems on the planet [Linux] could be developed is, to put it mildly, surprising. … This idea—or ideal—of open source software is not limited to this OS … It extends to many of the core technologies that make the Net run. And this idea, or ideal, of putting into the commons one's work product—of giving away what one makes, with no guarantee of compensation—might all sound wildly 60s-ish, wildly idealistic: Marx applied to code. It sounds alien to our tradition, foreign to what has made our national flourish. … Until one thinks again … about the way science works. For basic science functions much the same—progress made and given to the next generation.[9]

Lessig goes on to explain that the Internet, and perhaps most dramatically the World Wide Web, is built on public standards. These open source choices created "the fastest growing network in our history." Open source, Lessig concludes by quoting an unnamed software engineer, "is the Internet."[10]

The last thirty years of the Internet's evolution have exposed the extent to which large proprietary models have contributed to what Jonathan Zittrain calls the "appliancized" nature of the Internet.[11] While they have punctured the image of Lessig's ideal, his core points remain more than aspirational values for internet governance: "The values of Internet governance are about the values of governance; and the balance the Internet must strike is between that part that leaves itself open to these values, and that part that doesn't." For Lessig, the concept of the commons is key. "[W]e can't privatize every feature of cyberspace," he writes. "The opposite of private is not the government. The opposite of the private is the commons. … There is a value in preserving that space, regardless of efficiency."[12]

IV. Assessing Benefits and Risks of Open and Closed Models

The preservation of the commons and the parallels between how open source software and scientific inquiry operate help to explain the vast potential of more open AI models. Open-source

[8] Lawrence Lessig, Open Code and Open Societies: Values of Internet Governance at 104, *Chicago-Kent Law Review,* 1999.
[9] Lessig, *id.* at 107-108.
[10] *Id.* at 109.
[11] Jonathan Zittrain, *The Future of the Internet and How to Stop It.*
[12] Lawrence Lessig, Open Code and Open Societies: Values of Internet Governance at 116, *Chicago-Kent Law Review,* 1999.

software development, including the development of open foundation models, is critical to responsible AI development and critical to ensuring that benefits of the "AI revolution" maximize their potential to serve people equitably.[13] The reasons are manifold. As OTI and nearly 50 academics and civil society organizations wrote in an open letter earlier this week:

> For decades, open source software has provided building blocks for everything from creating art to designing vaccines. According to recent estimates, open source software is worth more than $8 trillion in value[14] and is a part of 96% of commercial software.[15] The U.S. government is one of the biggest users of open source software in the world,[16] and funds open source approaches ranging from boosting cybersecurity to protecting human rights and fighting cancer.[17]

NTIA also rightly highlights in its RFC some of the ways in which open AI models present tremendous positive potential:

> Dual use foundation models with widely available weights (referred to here as open foundation models) could play a key role in fostering growth among less resourced actors, helping to widely share access to AI's benefits…. Open foundation models can be readily adapted and fine-tuned to specific tasks and possibly make it easier for system developers to scrutinize the role foundation models play in larger AI systems, which is important for rights- and safety-impacting AI systems (e.g. healthcare, education, housing, criminal justice, online platforms etc.).
>
> … Historically, widely available programming libraries have given researchers the ability to simultaneously run and understand algorithms created by other programmers. Researchers and journals have supported the movement towards open science, which includes sharing research artifacts like the data and code required to reproduce results.[18]

[13] See, *e.g.*, World Economic Forum, Why Open Source is Crucial for Responsible AI Development, Dec. 27, 2023, https://europeansting.com/2023/12/27/why-open-source-is-crucial-for-responsible-ai-development/#; Kapoor & Bommasani et al., On the Societal Impact of Open Foundation Models, Feb. 27, 2024.

[14] Manuel Hoffman et al., "The Value of Open Source Software," Harvard Business School, January 2024.

[15] Synopsys, "2024 Open Source Security and Risk Analysis Report," February 2024. (Analyzed 1,067 commercial codebases across 17 industries in 2023, and found that 96% of those codebases contained open source.) See also, Chinmayi Sharma, "Tragedy of the Digital Commons," North Carolina Law Review, October 2022. ("Google, iPhones, the national power grid, surgical operating rooms, baby monitors, surveillance technology, and wastewater management systems all run on open-source software… Without it, our critical infrastructure would crumble.")

[16] Eric Goldstein and Camille Stewart Gloster, "We Want Your Input to Help Secure Open Source Software," Cybersecurity and Infrastructure Security Agency, August 2023. See also, federal policy supporting open source and open innovation, e.g., Tony Scott and Anne Rung, "M-16-21 Federal Source Code Policy: Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software," August 2016.

[17] See, e.g., Rachel Berkowitz, "How Berkeley Lab Helped Develop One of the World's Most Popular Open-Source Security Monitoring Platforms," Lawrence Berkeley National Laboratory, February 2023; "Supporting Critical Open-Source Technologies That Enable a Free and Open Internet," State Department, November 2023; and "CANcer Distributed Learning Environment," National Cancer Institute, February 2023.

[18] NTIA RFC at 14060.

On balance, the benefits of open AI models, whether foundation models or compound models built on top of them, outweigh the risks. While we at OTI are deeply mindful of risks, we urge policymakers, regulators, and researchers to articulate not just risks but measurable and foreseeable harms as precisely and specifically as possible. Without this kind of rigor, "safety" and "risk"-focused discourse can morph into a bogeyman that intimidates consumers, legislators, and regulators into overstating the risks of openness and significantly underweighting its benefits. The subsections that follow analyze some of the relative risks and benefits of open and closed models along five major policy objectives: cybersecurity; U.S. foreign policy; transparency, public accountability, and human rights; innovation and competition; and community control and benefits.

A. Cybersecurity and related harms to people

A short piece from the Wilson Center entitled "Open Source Software and Cybersecurity: How unique is this problem?" is worth quoting in full:

> Both open source and proprietary models of producing software will inevitably contain vulnerabilities. Vulnerabilities can come from dependency management (what, how, and which software packages are pulled into a new software project) to bad-faith actors (people that intentionally break into systems, or contributors intentionally changing the software to be exploitable) whether the software is developed internally or in the open.
>
> Best practices for enhancing security in software already exist, and these apply to both open and proprietary code, packages, and systems. No matter the model of development, these best practices can guide developers during every stage of the life cycle, from development of software to architecting a system.[19]

The Wilson Center goes on to note that OSS's widespread success has driven adoption and use. As a result of the massive uptake of OSS, and not because of anything inherent in OSS itself, there are attendant cybersecurity vulnerabilities.[20]

[19] Ashley Schuett, Alison Parker, and Alex Long, Open Source Software and Cybersecurity: How Unique is this Problem?, Wilson Center, Nov. 10, 2022.

[20] *Id.* ("OSS is widely adopted in both open and proprietary systems, resulting in decentralized usage of code that can contain vulnerabilities. Because the use of OSS is more widespread than proprietary code, it is difficult to track these vulnerabilities. In turn, it can be harder to identify and remediate them. This is not due to the nature of OSS itself, but to its widespread success and use.")

1. *Benefits*

The cybersecurity benefits of open-source software have long been understood, including by U.S. cybersecurity and national security practitioners and experts. Among open source software's greatest security strengths is the ability of security researchers to independently examine code for vulnerabilities and then feed those findings back into an established ecosystem where software developers can be alerted and patches written. Open source also provides the ability for other developers to patch older code if the original authors are no longer willing or able.

Microsoft's 2023 Digital Defense Report emphatically makes the case for open-source software's benefits to cybersecurity: "Open-source collaboration also drives innovation and enhances skills through shared tools and techniques, leveraging the inclusivity and diversity of a community. This is vital for understanding the current supply chain threat landscape and scaling mitigation efforts against emerging threats."[21]

It is worth noting that the Department of Defense, arguably the most security-conscious federal agency in the United States, relies heavily on open-source components of its cyber and weapons systems[22], not simply because open source integration drives innovation but *also* because it produces more secure products and systems when accompanied by a coordinated approach to rigorous vetting.

2. *Risks*

The overarching question, which NTIA itself has recognized, is how the emergence of open foundation models present *marginal* risk relative to more closed models or information publicly available on the internet. Broad claims about open models causing mis-and disinformation seem to conflate the large zones in which risks to information integrity challenges are common among open and proprietary systems. In fact, we have clear evidence that proprietary models are being frequently weaponized by bad actors.

Much has been made of the impact of generative AI on disinformation, and some commentators have casually equated open generative AI models with significantly higher risks of disinformation.As nearly fifty academics and civil society organizations emphasized in an open letter to Secretary of Commerce Raimondo, "the claim that open models make it easier to operate disinformation campaigns needs to be compared against the ease of conducting disinformation campaigns using closed models like DALL-E 3 and existing tools like Photoshop.[23]" Indeed,

[21] Microsoft Digital Defense Report 2023 at 116.

[22] Of course, weapons systems integrate both open and closed hardware and software components.

[23] See, e.g., Sayash Kapoor and Arvind Narayanan, "How to Prepare for the Deluge of Generative AI on Social Media," Knight First Amendment Institute at Columbia University, June 2023. ("[T]he bottleneck for successful disinformation operations is not the cost of creating it.")

Microsoft's own research team notes the weaponization of LLMs like ChatGPT, which runs on a closed foundation model, by adversary governments including China.[24]

There is at least one important area where analysis of marginal risk shows the rise of open foundational models causing measurable harms: the spread of child sexual abuse material (CSAM) and, in particular, computer generated CSAM (CG-CSAM). A 2023 paper establishes how open AI models using Stable Diffusion are contributing to the proliferation of CG-CSAM.[25] These known harms are grave, and we do not have decisive solutions in hand. The authors propose a number of potential mitigations that need further research.[26]

Public debate and NTIA's RFC has also focused on the relationship between model computing power and risk. In short, computing power is an imperfect proxy for risk, and using it as a proxy can create perverse incentives. Defining risk thresholds with computing power benchmarks cannot account for improved models that require less computing power. In addition, choosing static computing thresholds (instead of articulating risk through impacts/effects), also creates incentives for malicious model developers to stay just under a compute threshold that triggers heightened regulatory scrutiny. Similarly, malicious downstream model users can simply shop around for powerful models that fall below an established threshold and might therefore be likely to have less in-built safeguards against misuse.

Overall, NTIA and the U.S. government more broadly should identify, evaluate, and target the specific risks from openness in AI, including developing better proxies for risk that are not solely based on the amount of computing power used to train a model.[27]

One other risk inherent in the broad open-source software project is also worth considering. How can maintainers of code be incentivized to perform what is largely volunteer work? This long-standing problem in open source software security is likely to evolve further in the context of maintaining smaller, more bespoke models (not necessarily foundation models).[28]

[24] Staying ahead of threat actors in the age of AI, Microsoft, Feb. 14, 2024.
[25] David Thiel et al., Generative ML and CSAM: Implications and Mitigations, at 7-8 ("Several decidedly negative outcomes have been observed and pose a high risk to child safety as the availability of CG-CSAM grows. One likely scenario is that the advent of realistic CG-CSAM generates hundreds of thousands of reports to technology platforms, NGOs handling CSAM cases, and law enforcement, thus overloading the ability of companies and organizations to effectively handle reporting and investigations. The investigators will have the added challenge of determining whether the victim in the scenario is in fact a real person.")
[26] *Id* at 9-14.
[27] Rishi Bommasani, "Drawing Lines: Tiers for Foundation Models," Stanford University Center for Research on Foundation Models, November 2023. ("the relationship between compute and impact is quite tenuous and not evidentiated… there is no demonstration that compute robustly predicts results on risk evaluations, let alone demonstrations that compute predicts the impact foundation models have in society… compute is a measure of upstream resource expenditure, naturally divorced from downstream societal impact.")
[28] Luis Villa, The largest problems require government collaboration: Tidelift's response to the ONCD RFI

Software cybersecurity vulnerabilities have long bedeviled the U.S. software ecosystem. Some have argued that the problem is posed less by the absence of liability-based disincentives and more a failure to identify incentives for maintainers of software to continually identify vulnerabilities and update code.[29] But the Biden White House's 2023 National Cybersecurity Strategy directs the Office of the National Cyber Director to consult with academics and civil society to grapple with the hard question of how to assign liability within the value chain from developers to downstream users. The White House and other federal agencies should engage with the broader open-source software/open AI model community in the development of a new cybersecurity software liability framework that preserves the innovation potential of open AI models.

B. U.S. Foreign Policy: Geopolitics & Geoeconomics

Although debates about a U.S. policy or regulatory approach to open AI models might not obviously seem to implicate U.S. foreign policy interests, U.S. domestic policy on open models is highly relevant to foreign policy objectives, particularly when it comes to China.

1. Benefits

First, a thriving open model ecosystem could produce broad regulatory harmonization with key economic and international security partners like the European Union, India, Brazil, and Japan. Each of these countries are seen broadly as valuable trading partners as well as vital security partners in the context of a U.S.-China competitive dynamic. A U.S. regulatory regime, therefore, should take a broadly similar approach to these partner countries by avoiding differential treatment of open-source or otherwise more open AI models. A fragmented approach could engender unnecessary diffusion in the global economy. U.S. departure from what is likely to be a broadly common approach among democracies also would miss an opportunity to plant a decisive global flag in favor of openness in AI model development, further isolating countries like China and Russia whose governance models necessitate tight government control over foundation models and whose governing ethos is largely inimical to openness and public transparency.

Second, if the United States were to adopt a policy and regulatory approach to AI models broadly consistent with its storied history in helping to catalyze the FOSS movement, the benefits to its geopolitical and geoeconomic strategy would be clear. There would be broad policy and regulatory harmonization with the EU

Tidelift blog, November 9, 2023. https://blog.tidelift.com/tidelift-response-to-oncd-rfi ("If security problems don't align with the interests and time resources of the volunteer supply chain, adding more "requirements" is not likely to solve the problem—they will at best not respond, and at worst quit doing other maintenance activities!"

[29] *Id.*

Third, the claim that a permissive regulatory environment in the United States for open source model development would hand dangerous weapons to foreign adversaries fails to consider some powerful counterarguments. The Chinese government, for example, does not want to use open U.S. models because they fear that code might contain security exploits targeted at China, and because the governance of those models (through code, transparency to researchers and civil society, and broader public scrutiny) would not afford them the level of control they desire. It would be unwise to base legislative or regulatory policy on an overstatement of risks and — in the process — discourage the development of open AI models in the United States in ways that run counter to broader American security and economic interests.

Fourth, stifling innovation in open models may have the effect of granting a competitive economic advantage to countries that permit open AI models to flourish. And to the extent that the evolution of open AI models become essential to cybersecurity and broader national security offensive and defensive capabilities, a U.S. decision to hamper open model development could be directly deleterious to our own national security capabilities.

2. Risks

Detractors of open-source systems often note that adversary nation-states can use powerful open foundation models, too. While we do not dismiss this possibility, several known factors suggest that this threat may be overstated. Hostile or potentially hostile state actors like North Korea, Iran, and certainly China have the capacity to develop and train their own foundation models. Indeed, many U.S. adversaries will want to train their own models to do precisely what they want, rather than rely on code originating from the United States. In other words, many nations in hostile or competitive relationships with the United States will be *wary* of, not eager to, rely on open-source models that have been developed wholly or predominantly in the United States.[30] develop their own foundation models (and already are); they don't need open U.S. models. They also can and are working to weaponize LLMs like ChatGPT.[31]

C. Transparency and Public Accountability

Models further along the gradient of openness models are more transparent than those that tend toward the closed end of the spectrum. As noted above, model openness has many implications for transparency and therefore for public accountability and democratic health. These benefits need study, policy, and regulatory approaches to be realized. And openness alone cannot "solve the problem of oversight and scrutiny."[32]

[30] Microsoft, Staying ahead of threat actors in the age of AI, Feb. 14, 2024.

[31] *Id.*

[32] David Gray Widder, Sarah West, and Meredith Whittaker, Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, at 2-3, August 16, 2023.

1. Benefits

As Kapoor and Bommasani et al. note, openness in the broader open-source software context and the more specific context of AI foundation models furthers important objectives in transparency and therefore accountability.[33] Relatedly, more open models may also further the crucial ability of people to be able to contest decisions made by AI models. As Jim Dempsey, Susan Landau, Steve Bellovin, and others explain in a recent paper, true "contestability" can depend upon some meaningful transparency into some or all of the following: model methods, criteria, code, and data.[34] More open systems, even those that do not provide access to source code itself, are generally better to further the important objective of contestability.

The presence of more open models may also exert helpful competitive pressure that drives improvements in transparency and explainability among more closed models. Specifically, all AI developers and companies are likely to feel pressure to respond with more information about model data sources, methodology, training, and risk mitigations based on a study of downstream effects.

The better the incentives for meaningful transparency across the board, the more we drive a race toward the top, rather than a slide to the bottom. More open models also increase pressure on companies building closed models to provide meaningful transparency into and to open up key design and governance questions to a broader set of stakeholders. In this sense, the presence of successful open models can theoretically have a salutary effect on the entire ecosystem of foundation models.

2. Risks

At the same time, the charge is that open models represent ungoverned space. While this is an overstatement and a misunderstanding of how open-source software has been governed by research communities for decades, the colloquial equation of open with risks may be rooted in some genuine concerns. In particular, as Bommasani, Kapoor, et al. detail clearly, publishing model weights alone is an insufficient step for transparency. Similarly, in "Open (for Business): Big Tech, Concentrated Power, and the Political Economy of Open AI," Widder, West, and

[33] Kapoor and Bommasani et al. at 4-5.

[34] Susan Landau, James X. Dempsey, Ece Kamar, Steven Bellovin, Recommendations for Government Development and Use of Advanced Automated Systems to Make Decisions about Individuals, March 1, 2024 ("Nonetheless, although designing advanced automated systems to support a meaningful right of contestability is difficult, it is not impossible—and it is often required by law, … The degree of transparency necessary to support contestability will vary by context. Especially where the government fails to justify an outcome in a way that is understandable by the affected individual and his or her representative, it may be necessary for advocates and litigators to delve into how the system was constructed. Datasheets or model cards as documentation for how a system was built could enable contestability (Mitchell et al., Ehsan et al.), but in some cases a deeper examination of methods, criteria, code and data may be necessary along with expert analysis by those asserting challenges.")

Whittaker remind us that meaningful openness is not enough to achieve the lofty aims of democratizing AI or equitably distributing its benefits throughout society. They argue that this is in large part because of the well-entrenched "concentration of power in the tech industry."[35]

D. Economic Health, Innovation, and Competition

In late March 2024, the startup Databricks announced that it had succeeded in building what might be the world's most powerful open source large language model. "Similar in design to the [model] behind OpenAI's ChatGPT," it outperformed every other open source model available, including Meta's LLaMA 2 and Mistral's Mixtral, which are two leading open source (or not quite open source, depending on one's perspective on LLaMA) models. It also performed nearly as well as ChatGPT. Databricks says it plans to release DBRZ under an open-source license.[36] There is much yet to unpack about Databricks' emergence onto the scene, but it is a powerful example of companies' will to innovate in the development of open AI models.

1. Benefits

By one Harvard Business School study's estimation, the economic value of open-source software is $8 trillion.[37] That staggering number may be an over- or under-estimation,[38] but even arbitrarily quartering that estimate results in a remarkable figure estimating the value of open source. Open-source software has served as a key driver of U.S. technological innovation and economic growth, and it would be prudent for U.S. policymakers to create the conditions for innovative open AI models to continue that trend.

2. Risks

[35] David Gray Widder, Sarah West, and Meredith Whittaker, Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, August ("We find that even though there are a handful of meaningfully transparent, reusable, and extensible AI systems, these and all other 'open' AI exists within a deeply concentrated tech company landscape. With scant exceptions that prove the rule, only a few large tech corporations can create and deploy large AI systems at scale, from start to finish - a far cry from the decentralized and modifiable infrastructure that once animated the dream of the free/open source software movement.[16] Given the immense importance of scale to the current trajectory of artificial intelligence, this means 'open' AI cannot, alone, meaningfully 'democratize' AI, nor does it pose a significant challenge to the concentration of power in the tech industry.")

[36] Will Knight, Inside the Creation of the World's Most Powerful Open Source AI Model, Wired, Mar. 27, 2024.

[37] Manuel Hoffman et al., The Value of Open Source Software, Harvard Business School Working Paper 24-038, Jan. 2024.

[38] Luis Villa, Eight triiiiiilllion dollars: the "new" valuation of open source, Tidelift, Feb. 1, 2024. ("[T]he number is almost certainly too small. The authors explicitly exclude operating systems from their measures, and yet we know that open source operating systems are (1) extremely complex to create … and (2) extremely widely deployed, both on servers and on consumer devices (Android phones and in new home embedded devices like TVs). In addition, the survey does not appear to capture the value of web browsers, which are second only to operating system kernels in complexity, possibly more widely deployed than open source operating systems, and central to modern e-commerce. A total value number that captured those missing components would likely be even larger.")

Again, the risks to new entrants in the market stem from the uphill battle against entrenched tech companies with massive training data sets at their disposal. Widder et al. argue that while open foundation models may increase competition in some parts of the AI supply chain, they will struggle to reduce market concentration in the highly concentrated upstream markets of computing and specialized hardware providers.[39]

E. Community Control and Benefits

This section focuses more on impact on and use by people and communities than on macro-economic health and innovation. But these issues aren't neatly segregable and community control over the training and application of foundation models reinforces competitive benefits and economic vitality, and is also furthered by them.

1. Benefits and Challenges

One of the key benefits of a healthy ecosystem characterized by a prevalence of open models is that many people can learn how the technology works. This enables technologists and community leaders to partner in ways that are tailored to address specific community needs and implement community-driven solutions.[40]

Relatedly, open-source projects can also be used to fill technological gaps that aren't being met in the private sector. For example, the OpenCellular project aimed to make it possible for communities not currently served by mobile network operators (MNO) to start their own. This is achieved through open sharing not only of software, but also hardware schematics and other plans. This could drastically reduce the cost of becoming an MNO.

The low (often zero) dollar cost for using open-source also lowers the barrier to entry for those interested in learning skills like coding. A variety of open-source programming languages, and training materials make it possible for someone to start coding on less than one hundred dollars worth of equipment. The benefits for technical literacy, digital equity, and the ability of would-be coders from all over the country to serve their communities are potentially vast.

Open data provides an analogue for the possible community benefit of open models. Governments at every level have spent the last decade or more making a wide array of public data available on the internet in "machine readable" formats. This has allowed researchers and advocates to look into the effectiveness of government programs, or highlight matters of public concern. From mapping urban tree canopies, to looking at healthcare outcomes in rural areas,

[39] *See* David Gray Widder, Sarah West, and Meredith Whittaker, Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI, August 17, 2023.
[40] See, e.g., Bending Generative AI's Trajectory Toward a Responsible Technology Future, Omidyar, Sep. 2023;

opening datasets has provided for new levels of citizen engagement and insight that were only possible because they were available for creative new uses.

V. Recommendations for Policy and Regulatory Approaches

It may be premature to take highly prescriptive regulatory approaches to foundation models, whether those models fall more on the open or closed end of the spectrum. But there are at least four broad recommendations that should guide the NTIA's and the interagency high-level approach to considering policy and regulatory frameworks for foundation models.

A. Study the marginal risk of open models and precisely articulate observable harms.

NTIA and other U.S. government agencies must focus vague discussions about the risks of open AI models on the study and precise articulation of the marginal risk these models pose. Rigorous empirical analysis is necessary to inform thoughtful, targeted interventions. As a coalition of academics and civil society organizations (of which OTI was a part) put it in our open letter: "We urge you to be rigorous in evaluating and targeting the specific risks from openness in AI, including developing better proxies for risk that are not solely based on the amount of computing power used to train a model."

B. Create common requirements for responsible development across the entire gradient of openness.

Precisely because of the definitional indeterminacy we have highlighted about the terms "open," "open source," and "closed," these terms would form an unsound basis for policy or regulation aimed at articulating differential rules or treatment depending on how a model is classified on the gradient of openness. Apart from the lack of consensus and clarity, the mere fact that terms are likely to remain contested for a long time suggests that policymakers who rightly feel the urge to act swiftly should not wait indefinitely. **Instead, policymakers should establish a common baseline of governance requirements for *all* foundation models, rather than considering higher regulatory burdens for certain types of open models.** Taken together with a serious study of marginal risk, the implementation of laws, policies, technical standards, and meaningful transparency norms will help to produce public accountability and a good governance race to the top. Such a policy approach would give upstart companies entering the marketplace and communities across America the best possible chance to ensure that advances in AI produce wide-ranging economic benefits.

C. Consider the broad range of relevant national security and foreign policy objectives before recommending policy or regulatory action aimed at vaguely defined or narrow types of security risks.

As discussed at length in Section IV.B, vague claims of security risks or specific articulations of marginal risk need to be weighed against other factors like regulatory harmonization with partners and participating in economic coalitions competing with China. These geopolitical and geoeconomic considerations are manifestly "security" considerations that need to be balanced on the scale when assessing the benefits and risks of more open foundation models. Returning to the question of marginal risk, U.S. intelligence agencies and U.S. companies should continue to closely study foreign governments' malicious use of models along the gradient of openness. This analysis will be vital to understanding threat vectors and strategies to address them.

D. Develop a thoughtful approach to cybersecurity software liability that accounts for the need to incentivize innovation in open AI models.

We are without easy solutions to updating a framework for developer and/or downstream liability. But the U.S. government should continue the consultative work started pursuant to the 2023 National Security Strategy specifically with the question of open AI models (foundational or otherwise) in mind. Open-source software and open AI model proponents should engage with the discussions around implementing the National Cybersecurity Strategy to ensure that a liability regime and a standard of care are explicitly developed with the implications for innovation in open AI models in mind.

VI. Conclusion

At OTI, we favor an ecosystem in which open AI models can flourish alongside proprietary ones. Our analysis of the benefits and risks of open models in the context of five major U.S. policy objectives (cybersecurity, foreign policy, public accountability, economic health and innovation, and community control and benefits) suggests that on balance the benefits of more open models are manifold and outweigh the risks. This does not mean that models should develop without guardrails, but both guardrails and incentives should be broadly common across the gradient of openness and should push and pull developers toward meaningful transparency about model inputs and downstream effects. Section V (above) contains four key recommendations for policy and regulatory approaches to open AI models.

Respectfully submitted,

/s/ Prem M. Trivedi
/s/ Nat Meysenburg

New America's Open Technology Institute
740 15th Street NW, Suite 900
Washington, D.C. 20005
March 27, 2024