Mozilla's Response to the
National Telecommunications and Information
Administration's Request for Comments
on

Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights

March 2024

Contact:

Jenn Taylor Hodges (she/her), jhodges@mozilla.com

Table of Contents

About Mozilla

Mozilla's mission is to ensure the internet is a global public resource, open and accessible to all. An internet that truly puts people first, where individuals can shape their own experience and are empowered, safe, and independent.

Founded as a community open source project in 1998, Mozilla consists of several organizations, most notably the non-profit Mozilla Foundation, which leads our movement-building work, and its wholly owned subsidiary, the Mozilla Corporation, which leads our market-based work, including the development of the Firefox web browser. They work in close concert with each other and a global community of tens of thousands of volunteers under the single banner: Mozilla.

For the past five years, Mozilla has been committed to advancing trustworthy AI. Mozilla recently published a paper, Accelerating Progress Toward Trustworthy AI, that outlines how Mozilla and its allies are advancing openness, competition, and accountability in AI. Mozilla is putting its resources behind these priorities as well: The Mozilla Foundation has been dedicating 100% of its $30M a year budget to philanthropic activities, advocacy, and programmatic work on this topic. Mozilla is also investing another $30M in research and development on trustworthy AI via Mozilla.ai, as well as $35M in responsible tech startups — including startups with a focus on trustworthy AI — through Mozilla Ventures.

As an independent and mission-driven organization, Mozilla is committed to working with regulators to develop effective policies that ensure that innovation and growth in AI serve the public interest. We put people above profit..

Executive Summary

Today, as AI is becoming an increasingly important part of the technology industry, we're seeing early signs of history repeating itself. In recent years, many leading players across the industry have moved towards less openness in AI, gating models behind proprietary APIs and documenting fewer and fewer details of the development process as well as about critical inputs and outcomes. Creating an incentive structure that promotes openness and knowledge-sharing thus becomes a key building block of any strategy aimed at driving progress across the U.S. economy — not just in a handful of well-resourced corporate research and development labs.

It's important we pay close attention to what's happened in the past to make sure we harness the benefits of openness while avoiding past mistakes. Open source software and 'open source' AI aren't the same thing — but they're underpinned by many of the same principles and values.

Against this backdrop, our submission to the NTIA's request for comments seeks to answer a range of important questions that warrant consideration in any conversation about openness in AI, with regard to the consultation's focus on widely available model weights but also beyond. Drawing on Mozilla's own history as part of the open source movement, this submission seeks to help guide difficult conversations about openness in AI. First, we shine a light on the different dimensions of openness in AI, including on different components across the AI stack and development lifecycle. Second, we argue that openness in AI can spur competition and help the diffusion of innovation and its benefits more broadly across the economy and society as a whole; that it can advance open science and progress in the entire field of AI; and that it advances accountability and safety by enabling more research and supporting independent scrutiny as well as regulatory oversight. In the past and with a view to recent progress in AI, openness has been a key tenet of U.S. leadership in technology — but ill-conceived policy interventions could jeopardize U.S. leadership in AI.

Good policymaking on AI, and on openness in AI in particular, therefore requires a careful balancing of benefits and risks as well as analytical rigor in taking into account the various dimensions and actors in the AI ecosystem. Rash decisions and ill-considered solutions may cause irreparable damage to the 'open source' AI ecosystem, and with it to the prosperity and safety of the American people. Against this backdrop, we make the following recommendations:

- Impose proportionate and carefully considered regulatory obligations relating to 'open source' AI.

- Support the ‘open source’ AI community in developing norms and practices around responsibly developing and openly releasing AI models and components.
- Invest in and provide resources for the development and maintenance of ‘open source’ AI.
- Involve federal agencies responsible for protecting civil rights, promoting competition, and advancing scientific research in the development of any policy touching on openness in AI.

We hope these measures can help construct a better AI ecosystem — one that is more innovative, more competitive, and more accountable.

1. Introduction

Mozilla has been on the frontline of defending the open internet for 25 years. Our history is deeply intertwined with that of the open source movement. When Microsoft was cornering the web browser market in the 1990s, Mozilla’s open source browser Firefox provided an alternative and broke Microsoft’s chokehold on the market — with privacy, security, and openness at its heart. Still, we’ve seen similar dynamics play out again and again over the past decades, with increased concentration and centralization in the market, favoring the construction of walled gardens and a push to capitalize on open source innovation without giving back to the open source community.

Today, as AI is becoming an increasingly important part of the technology industry, we’re seeing early signs of history repeating itself. In recent years, many leading players across the industry have moved towards less openness in AI, gating models behind proprietary APIs and documenting fewer and fewer details of the development process as well as about critical inputs and outcomes. Creating an incentive structure that promotes openness and knowledge-sharing thus becomes a key building block of any strategy aimed at driving progress across the AI industry and the U.S. economy more broadly — not just in a handful of well-resourced corporate research and development labs.

It’s important we pay close attention to what’s happened in the past to make sure we harness the benefits of openness while avoiding past mistakes. Open source software and ‘open source’ AI aren’t the same thing — but they’re underpinned by many of the same principles and values.

This is why Mozilla is deeply involved in strengthening the open ecosystem in AI and rallying the community. In addition to our long-standing work in this area, ahead of the United Kingdom’s AI Safety Summit late last year, Mozilla organized a joint statement with over 1,800 signatories that emphasized how openness is a boon to AI safety and security. Further, in February 2024, Mozilla and the Columbia University Institute of Global Politics brought together over 40 leading scholars and practitioners working on openness and AI to explore what 'open' should mean in the AI era. And this March, Mozilla, the Center for Democracy and Technology, and a wide-ranging group of civil society organizations and academic experts wrote to Secretary Raimondo to emphasize the importance of openness in AI.

Against this backdrop, our submission to the NTIA's request for comments seeks to answer a range of important questions that warrant consideration in any conversation about openness in AI, with regard to the consultation's focus on widely available model weights but also beyond. Drawing on Mozilla's own history as part of the open source movement, this submission seeks to help guide difficult definitional conversations about openness in AI; to shine a light on the many benefits openness offers to innovation and research, competition, and accountability and safety in AI; and to outline key considerations and potential next steps for policymakers deciding on what public policy around openness in AI and AI more broadly should look like.

2. Lessons from 25 years in the open source movement

Responding to questions 6, 6a, and 9

In 2021, Mozilla completed its "Reimagine Open" project, charting the history and future of the intricate relationship of openness and the internet. At the outset of the internet and the web, their open architecture and design created the conditions for the internet to thrive, enabling broad access to technological building blocks and the opportunity to shape and participate in online experiences.

However, this open architecture of the internet and the values underpinning it gradually eroded, with open source software being increasingly enveloped in proprietary systems and platforms. As we wrote back in 2021:

> *"Today's internet has moved away from these values. The term "open" itself has been watered down, with open standards and open source software now supplanted by closed platforms and proprietary systems. Companies pursuing centralization and walled gardens claim to support "openness." And tools for online accountability have failed to scale with the incredible diversity of online life. The result is an internet that we know can be better."*

These dynamics are already visible in today’s AI ecosystem. Advanced AI models and other critical components, such as training datasets, are closed off and turned into platforms, not widely released. Foundational research and innovation is confined to corporate labs, not shared. And terms like ‘community’ and ‘openness‘ are too often projected to evoke a sense of transparency and accountability, but not truly lived. This is in stark contrast to the time before the rush to commercialize AI disincentivized many tech companies from sharing progress in AI more widely, for example when Google open-sourced the transformer architecture — the architecture underpinning today’s large language models — in 2017 as well as early large language models like BERT, laying the foundation for the current boom in generative AI.

This has broad implications for the AI ecosystem as a whole: the AI market is consolidating before it has even fully taken shape, with companies again converging on corrosive business models that center a “move fast and break things” ethos over building trust and ensuring that the technology works for everyone. With the prospect of control over AI being concentrated in the hands of a few companies — many of whom already control most other corners of the internet — ensuring true accountability is likely to become yet another uphill battle.

That is why we need to be wary of self-serving narratives propagated by those dominating the industry. Instead, we need to ask ourselves how openness in AI can be preserved, promoting open innovation, open science, and accountability, while simultaneously ensuring that making AI more open contributes to a trustworthy and safe ecosystem.

3. What is ‘open’ in AI?

Responding to questions 2, 3e, 4, 5f and 6

Discussions of openness in AI must go beyond a focus on model weights and a binary distinction between ‘open’ and ‘closed’. While there are many parallels between open source software and ‘open source’ AI, there are also differences. Namely, openness in AI has many different dimensions and goes far beyond just releasing source code.

First, there are a variety of different components in AI. While most of the public debate focuses on the release of openly available models (i.e., model weights), this is only part of the picture. While the open release of a model allows its use and adaptation, other components are necessary, for example, to enhance transparency and enable deeper scrutiny and reproducibility. For instance, this includes several technical artifacts:

- Model weights, e.g. pre-trained weights or training checkpoint weights

- Source code, e.g. for training and finetuning a model, for inference (i.e., querying a model), or for user interfaces through which users can interact with a model
- Data, e.g. pre-training data, finetuning data, or evaluation data

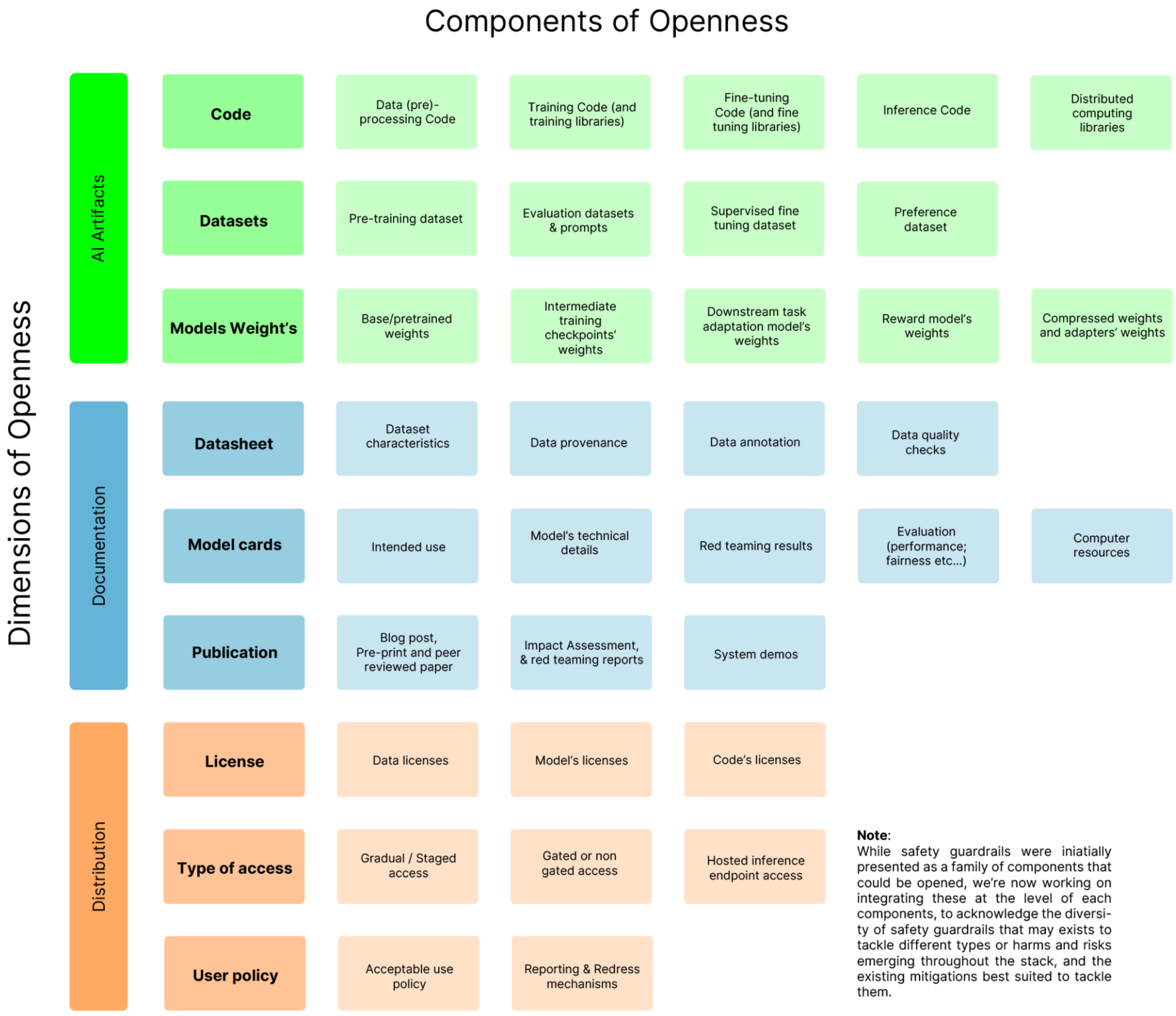
However, there are also several non-technical artifacts, particularly technical documentation that can, amongst other things, provide insights into a model's design, data used throughout its development, risks and (potential) impacts stemming from a model, and information about evaluation processes and results. Further, openness goes beyond the release of technical and non-technical artifacts — collaboration within and across communities and the documentation of development and decisions have long been a common feature of the open source ecosystem, and can aid reproducibility and understanding in AI research and development.

Second, we have to distinguish between different modalities of releasing AI components, from the type of license under which they are released to where they are hosted — here, there are many shades between fully 'open' and 'fully' closed. For example, a component may be openly released, but under a non-commercial license or with restrictions on how or for what purposes it may be used. This would be a deviation from how software is released under traditional open source licenses, but would still fall on the broader spectrum of openness. There also is a "gradient" of how AI components — particularly model weights — can be accessed: from providing no outside access to only providing access via interfaces and hosted API access to making the component available for download.

In a recent workshop co-organized by Mozilla and Columbia University's Institute for Global Politics — the "Columbia Convening on Openness and AI" — we convened AI experts and representatives of the open community to further map out these dimensions. Mozilla recently published the technical and policy readouts from the workshop to serve as a resource to the community. In this context, Figure 1 below provides a preliminary mapping of the different dimensions of openness across the AI stack, and will be revised based on discussions at the workshop.[1]

Additionally, the table in the Annex attempts to provide a more extensive overview of the (non-exhaustive) components of the AI stack that were discussed at the workshop, and the benefits and drawbacks that bringing openness to each of these components would bring to the AI ecosystem.

[1] An updated version of Figure 1 will be published in the coming weeks.



AI Artifacts	Code	Data (pre)-processing Code	Training Code (and training libraries)	Fine-tuning Code (and fine tuning libraries)	Inference Code	Distributed computing libraries
	Datasets	Pre-training dataset	Evaluation datasets & prompts	Supervised fine tuning dataset	Preference dataset	
	Models Weight's	Base/pretrained weights	Intermediate training checkpoints' weights	Downstream task adaptation model's weights	Reward model's weights	Compressed weights and adapters' weights
Documentation	Datasheet	Dataset characteristics	Data provenance	Data annotation	Data quality checks	
	Model cards	Intended use	Model's technical details	Red teaming results	Evaluation (performance; fairness etc...)	Computer resources
	Publication	Blog post, Pre-print and peer reviewed paper	Impact Assessment, & red teaming reports	System demos		
Distribution	License	Data licenses	Model's licenses	Code's licenses		
	Type of access	Gradual / Staged access	Gated or non gated access	Hosted inference endpoint access		
	User policy	Acceptable use policy	Reporting & Redress mechanisms			

Note:
While safety guardrails were iniatially presented as a family of components that could be opened, we're now working on integrating these at the level of each components, to acknowledge the diversity of safety guardrails that may exists to tackle different types or harms and risks emerging throughout the stack, and the existing mitigations best suited to tackle them.

Figure 1: *Preliminary mapping the dimensions of openness across the AI stack (developed for the Columbia Convening on Openness and AI).*

Finally, 'open source' development can take myriad forms. It can take place entirely in-house or entirely in the open, driven by multinational corporations or small communities. At Mozilla, for example, staff developers are working side by side with community members, who can help develop new features or identify and fix bugs in Firefox. While there are notable differences between open source software and 'open source' AI, as discussed throughout this submission, we see some similar patterns emerge across the AI ecosystem. It can thus be helpful to look to the open source software ecosystem — a helpful overview of which can be found in Mozilla's 2018 report on "Open Source Archetypes: A Framework For Purposeful Open Source" — for lessons on how a similar ecosystem may be built up in AI.

4. Why openness matters

Responding to questions 1d, 2, 2a, 2d, 3, 3a, 3b, 3c, 3d, 3e, 4, 5, 5b, 5c, 5f, 5g, 6, 6b, 7a, 7b, 8, and 8a

Recent debates around openness in AI have disproportionately focused on the risks of open technology rather than its benefits. In this respect, it also matters *how* we discuss such questions — and with which terminology. For instance, discussing AI and openness under the banner of 'dual-use' technology creates a highly securitized framing of the debate. While there certainly are national security implications, this also underlines the necessity of approaching these questions from a variety of different perspectives, including economic and scientific impacts as well as impacts on civil rights and liberties. Only by considering multiple perspectives on the issue, policymakers will be in a position to more effectively weigh the risks and benefits of openness in AI in particular.

4.1. Openness and innovation

Openness is the bedrock of AI innovation

Openly available AI components have long served as a bedrock of progress and innovation in AI, from incremental tweaks to paradigm-setting breakthroughs. In fact, much of the current boom in generative AI can be traced back to Google's decision to openly document and share its newly developed transformer architecture — underpinning all state-of-the-art large language models — as well as open-sourcing early large language models like BERT. Similarly, the provision of many open source components — including transformer libraries — has turned the platform Hugging Face into a key hub and resource for AI developers.

At the same time, the sharing of countless smaller innovations has propelled the field of AI forward again and again. For example, open-sourcing libraries for techniques like "Low-Rank Adaptation" (LoRA) and other parameter-efficient finetuning (PEFT) techniques has significantly reduced the cost (and with that, the carbon footprint) of finetuning models and adapting them to specific uses where it may otherwise have been prohibitively expensive for many. With regard to data, initiatives like Mozilla's Common Voice project — compiling the world's largest crowdsourced multilingual voice dataset — can not only lower barriers to entry in fields like voice technology but also expand access to the technology for communities previously (and still!) underrepresented in machine learning training data.

Openness is vital for better science and a robust national R&D ecosystem

Openness helps accelerate scientific research because openly available models can be less expensive and easier to fine-tune and query, and supportive of reproducible, open research.

For example, finetuning a model to adapt it for specific (research) purposes or to modify certain properties of the trained model can be performed using far fewer computing resources and less data — and therefore, at much lower cost — given sufficient access to model weights and other information about a model's architecture (see above). This particularly benefits researchers and developers with limited resources, such as academic research groups or start-ups. Further, openness is a critical enabler of reproducibility in research, both in AI and in research enabled by AI in other disciplines. Without access to the research underpinning new technical developments and innovations, and without access to the technical artifacts developed in the process, researchers will not be able to reproduce research or validate specific claims or findings. This would fundamentally hold back both basic and applied research and obstruct knowledge-sharing as well as the practice of open science. Restrictions on the sharing of model weights or other AI components could threaten the integrity of open science and scientific dialogue, and hinder the diffusion of progress in AI.[2]

Finally, openness helps strengthen the broader research and development ecosystem. A culture of collaboration and scientific inquiry can make it easier to translate AI research and development into products and services. It also enables international collaboration on AI research, helping to identify and attract the best AI talent from around the world to the U.S.. Openness allows the U.S. to lead by example, incorporating safeguards and best practices into AI models that will be used around the world, rather than allowing another country's AI companies to provide the building blocks of the global AI infrastructure.

Openness provides 'unknown unknown' benefits for the economy

In addition to substantive debates about potential risks from openly releasing model weights, more recent debates have also invoked 'unknown unknown' risks from openness to national security. Indeed, with new technologies, we are often confronted with unknown risks as we learn more about the technology and its impacts. However,

[2] Further, as U.S. courts have held multiple times, computer source code must be viewed as expressive for First Amendment purposes. For example, in *Junger v. Daley*, the court held that "computer source code is an expressive means for the exchange of information and ideas about computer programming," granting it protections under the First Amendment. A similar argument could be made about the importance of protecting the sharing of information about model weights and other AI components.

this Rumsfeldian claim is now being used to turn vague, unscientific claims of catastrophic AI risk into the animating concern for AI policy, despite a clear lack of empirical evidence — in turn detracting from well-evidenced present-day — i.e., very well-known — harms from AI.

Policymakers must move to prevent and mitigate these harms, in addition to building the capacity to practice effective foresight and quickly counter newly emerging risks — that is, the 'known known' and the 'known unknown'. However, by their very definition, attempts to counter 'unknown unknowns' are bound to be ineffective. In fact, any claim to foresee the things (in Rumsfeld's own words) "we don't know we don't know" borders on mysticism.

If policymakers still seek to address 'unknown unknown' risks from AI, they must also take into account 'unknown unknown' benefits from AI. Indeed, the National Institute for Standards and Technology (NIST) also emphasizes potential benefits in its AI Risk Management Framework, "offer[ing] approaches to minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive impacts" (p. 4). History shows us countless examples of innovation functioning as the exploration of the unknown, creating significant economic growth and other societal benefits in the process. There is so much unknown about the benefits of AI, and policymakers must not ignore this.

History shows the economic value of openness in digital technologies

Openness in digital technologies has come under threat before, and made it harder for open source's economic value to be fully realized. For example, consider encryption in early web browsers of the 1990s — a vital tool for enabling the sharing of payment information across the Internet, and thus a key requirement for e-commerce. This type of encryption was controlled by the U.S. government, making it harder to enable a global ecosystem of commercial transactions on the Internet. The early web browser Netscape successfully fought and won against these export controls on encryption, which enabled the birth of modern international e-commerce — a major boon for both the U.S. and the global economy.

4.2. Openness and competition

More openness in AI can drive competition and the diffusion of innovation

As in open source software development, more openness in the AI ecosystem can come with significant benefits for competition and contribute to a more innovative and contestable AI market. As outlined in a recent paper by Kapoor et al. and as Mozilla reiterated in a submission on competition concerns in generative AI to the European

Commission, there are three key pathways through which more openness in the AI ecosystem can contribute to advancing transparency and innovation in the space:

- Broader access to and enhanced diffusion of technological innovation in AI
- Enhanced customizability to specific purposes and uses of AI components
- Local inference on devices without necessitating third-party data sharing

Each of these can contribute to making AI components more widely available for integration into products and services and potentially more private.

‘Open’ alternatives in AI can help avoid vendor lock-in and lower switching costs

The increased availability of ‘open’ alternatives in the AI market can support competition by reducing switching costs as relying on specific proprietary model APIs or platform ecosystems (like those offered by the leading cloud service providers) can create lock-in effects for customers, both in the private and public sector. At the same time, the benefits of openness do not only accrue around openly available models, but also other components. For example, given the current rush to obtain new training data, developments in the AI ecosystem increasingly favor those who have already built large data collection infrastructures — including in other verticals — and those who can broker exclusive (and expensive) licensing agreements with publishers and other rights-holders. In this context, openly available datasets — ideally curated with privacy and copyright concerns in mind — can lower barriers to training for smaller developers. Even in the hardware stack, open-sourcing software frameworks can be a means to challenging proprietary infrastructure and dominant market actors.

Policymakers should be wary of industry co-optation and concentration of power

As Widder, West, and Whittaker have argued, promoting openness in AI alone is not sufficient for creating a more competitive ecosystem. There are also risks of openness being co-opted by big industry players, and a long track record of companies drawing significant benefits from open source technology without re-investing into the communities that have developed those technologies. Still, promoting openness in AI can be a key building block for a more competitive AI ecosystem. A deliberate and well-considered approach to AI policy and openness should pay due attention to those risks while creating an incentive structure that steers AI research and development towards openness and rewards collaboration and knowledge sharing.

Additionally, concentrating cutting-edge research in ever-fewer research labs may also exacerbate phenomena like algorithmic monoculture and entrench (or increase the "stickiness" of) existing technological paradigms at the expense of pursuing new research directions.

4.3. Openness and accountability

Debates around safety and 'open source' AI should center marginal risk

Current debates around safety and openness in AI often focus on the *absolute* risk of making AI models openly available. Further, claims about these are often poorly evidenced (or not empirically supported at all), as a recent review of several frequently cited studies alleging such risks demonstrates. As the same study and a recent policy brief from some of its authors demonstrates, the debate around this topic also fails to adequately take stock of the *marginal* risk of making AI models (and other components) openly available.

To arrive at effective and practical policy solutions addressing clearly identified problems, more analytical rigor is needed in this debate. The threat landscape for emerging technologies doesn't exist in a vacuum. That is why, in assessing risks from any new technology, one must take stock of the marginal risk of this technology relative to information already readily available elsewhere. For example, as recent research on biological threats from RAND and OpenAI indicates, current advanced AI models don't pose a significant marginal risk in this respect compared to information available online and through ordinary online search. Similarly, the risk from openly available AI models should be assessed relative to that from proprietary models and from other technologies and available information. For anyone involved, this involves being open about the limitations of studies on the topic and not conducting (quasi-) experimental studies without a clearly identified control group or counterfactual (as is standard scientific practice). Only through a rigorous assessment of marginal risk will policymakers be able to effectively mitigate risks from AI without unnecessarily (or unintendedly) foreclosing the benefits enabled by openness in AI.

Openness supports independent scrutiny and regulatory oversight

Additionally, openness in AI is vital for regulators and civil society to be able to assess AI systems and other components used in AI development to ensure they appropriately conform to all applicable laws and regulations as well as broader concerns around safety and bias.

Openness in AI more broadly helps increase understanding of AI risks and harms among regulators and society more broadly by enabling an ecosystem of research and knowledge sharing on the topic. Much research on the risks of and potential harms from AI has been conducted on or with the help of openly available AI components. For example, some of the most extensive audits for harmful content of large-scale datasets used in machine learning have been studies of the openly available LAION datasets. This has sparked significant (and justified) criticism of the LAION datasets, but also underlines that comparable research on training data (and other components) used by most leading AI companies has not been possible due to the fact that neither the data itself nor sufficient information about said data have been made available by these companies.

This is reminiscent of, for example, Twitter's overrepresentation in social media research compared to other, larger platforms due to the fact that Twitter, for a long time, provided more favorable and open conditions for independent research about its platform. We're seeing similar dynamics play out in AI research as we did in social media research. For instance, in an open letter and research paper published in March 2024, leading AI researchers argue that AI companies don't provide equitable and robust access for independent AI safety and trustworthiness research, and that their policies create chilling effects on such research. In short, as a group of academic researchers from the U.S., Canada, and the UK have argued: "Black-Box Access is Insufficient for Rigorous AI Audits."

However, these conditions are exactly what is needed to adequately identify risks and harms, and to hold companies developing and deploying AI to account.

At the same time, openness increases the availability of the tools that regulators need to monitor and evaluate (large-scale) AI systems. For example, the Open Source Audit Tooling project supported by Mozilla, provides an extensive overview of existing openly available tools for audit practitioners.

Research based on 'open source' AI can drive progress in safety

Despite current debates' focus on risks stemming from 'open source' AI, research relying on openly available AI components has been instrumental to advancing safety, security, and trustworthiness across the entire AI ecosystem. Not only does the open availability of AI components allow researchers to dismantle 'the black box', but it also enables the responsible disclosure of vulnerabilities to developers and the development of tools and techniques for the prevention and mitigation of harms.

For instance, research drawing on 'open source' AI has helped advance red-teaming and safety alignment work, measuring bias and mitigating toxicity, or removing protections from (or 'jailbreaking') aligned models (including state-of-the-art proprietary models). Such research has not only been critical to advancing safety and security research in the field and to making 'open-source' AI safer, but has also enabled providers of proprietary AI models to address crucial vulnerabilities in their products and services. In safety and security research, too, openness in AI promotes progress across the entire ecosystem.

At the same time, we're already seeing norms emerge around responsible practice in open research and development, for example around responsible disclosure of vulnerabilities. Similarly, more and more tools are made available for practitioners in the open community, for example documented in Hugging Face's Society & Ethics space or the Foundation Model Development Cheatsheet.

Openness helps bring more communities to the table

Openness in AI enables more communities to be able to understand, test, and trust the use of AI. For example, it helps enable researchers and journalists to be able to investigate how an AI system impacts different demographic groups, language communities, or geographic areas. Openness can also increase access for underserved groups to be a part of shaping the future of AI, and to benefit from value created through AI. In the current paradigm, underserved communities are often not a primary stakeholder for whom products are built, and increased access to AI resources can help diverse communities build and shape products for their specific needs. For example, decision-making in Mozilla's Common Voice project builds on ongoing engagement with the Common Voice community and a Language Reps Council representing the project's different language communities.

5. Regulating open AI

Responding to questions 6, 6a, 7, 7b, 7h, 8, and 8a

As Mozilla has argued again and again over the course of the past years, robust regulation to curb the risks stemming from AI is sorely needed. It is a necessity to keep those developing and deploying AI in check, particularly those dominating the industry.

Acting responsibly and not shying away from scrutiny are paramount across the technology industry. This applies to open source technology, too. Pursuing an open or 'open source' approach neither is nor should be a carte blanche. This was true for the open web, and it is true for 'open' AI as well. As we wrote in "Reimagine Open":

> *"Openness never meant the absence of all restrictions. The open Internet developed against the backdrop of technical features, social norms, and laws that influenced human behavior on the network. Together these created a set of accountability mechanisms that provided a backstop and framework for a successful human experience of the Internet."*

It is here where public policy can support the development of shared norms and help define a common standard of what responsibility in ‘open source’ AI means. The practice of openness in AI doesn’t happen removed from legal realities — in fact, good public policy in this space can complement good ‘open source’ development. And, conversely, more openness in the AI ecosystem can help identify and mitigate risks, and bring more scrutiny to both open and proprietary AI. It can be an asset to effective oversight and enforcement.

Against this backdrop, we believe there are several considerations policymakers must keep in mind in developing rules for AI that affect openness and ‘open source’ development:

1. **Provide definitional clarity:** As already discussed above, in debates around AI (including around the EU’s Artificial Intelligence Act), the term ‘open source’ has been used widely and loosely, with no consensus as to what exactly it captures in the context of AI and with widely diverging motives. As opposed to open source software, there is no widely accepted standard of what qualifies as open source, such as the Open Source Initiative’s repository of accepted licenses. In the absence of such a standard, policymakers must be careful in adopting specific language and not over-rely on widely known concepts that don’t translate to the AI context with sufficient clarity — especially where this will have legal ramifications for the AI ecosystem. They should also closely follow ongoing processes like the Open Source Initiative’s efforts to define ‘open source’ AI.
2. **Adopt an expansive view of openness:** As argued previously in this submission, a myopic focus on model weights falls short in capturing the complexity of the AI development lifecycle and of openness in AI in particular. Public policy should take into account the full breadth of AI components and all dimensions of openness in this regard, while considering the effects of different approaches — irrespective of whether they serve a restricting or supporting function — have on each of those. Only then can unintended consequences and potential harms to the broader (open) AI ecosystem be minimized, and its potential benefits can be captured.
3. **Consider the special nature and diversity of the ‘open source’ ecosystem:** The ‘open source’ ecosystem is no monolithic actor, be it in AI or traditional software. Its members comprise individual volunteer contributors to multinational corporations; development can occur entirely in-house or can be distributed and community-driven; and open source development can be part of non-commercial or profit-seeking endeavors. A catch-all approach that ignores this diversity and casts the respective concerns of different actors within the

ecosystem aside is unlikely to succeed. Policymakers should thus make efforts to better understand and navigate this landscape, taking into account, amongst other things, how ‘open source’ projects are embedded organizationally, how the technology is developed, and what the commercial (or non-commercial) context of development is.

4. **Develop proportionate rules:** In developing regulatory obligations for developers of AI (as well as for other actors along the AI value chain), regulators should take this diversity of actors into account as well. As we’ve argued previously, developers providing openly available AI components should adhere to recognized standards of good practice as much as possible. However, well-resourced multinational corporations engaging in open source development may be better positioned to comply with certain obligations than small non-profit or academic research labs, community-driven projects, or early-stage start-ups. Regulatory obligations should be designed in a way that doesn’t disproportionately disadvantage the latter to the benefit of larger developers of proprietary technologies. Similarly, in designing regulatory obligations, regulators should also take into account the inherent benefits of openly making available different AI components, for example the increased transparency and scrutability enabled by sharing training datasets or technical documentation. Finally, blunt instruments like imposing export controls on openly available AI components may run the risk of both creating substantial unintended consequences by obstructing innovation, research, competition, and transparency, and simultaneously fail at meeting their objectives due to significant practical obstacles to effective implementation and enforcement.

If policymakers take these principles to heart in developing new rules for AI, this can — again — serve as a productive foundation to steer AI development in a more open and trustworthy direction, underpinned by transparency and accountability.

6. Conclusion and recommendations

Responding to questions 5c, 7d, 8, and 8a

Openness is not an end in itself. It is a means to achieving more transparency, more accountability, and more innovation whose benefits are made available to more people, not concentrated in the hands of a small group of powerful and wealthy actors. Openness as a value has served us well in building the foundations of a thriving internet. It can help us do the same in AI.

But we’re already seeing the same erosion of openness that in the past has led the internet down a path marked all too often by concentration of wealth and power, and a striking disregard for people's agency and safety. In "Reimagine Open" we thus came to the conclusion that there is "an internet that we know can be better". Today, it is safe to say that there is an AI ecosystem that can be better, too.

As we've outlined throughout this submission, good policymaking on AI, and openness in AI in particular, requires a careful balancing of benefits and risks as well as analytical rigor in taking into account the various dimensions and actors in the AI ecosystem. Rash decisions and ill-considered supposed solutions may irreparably damage the 'open source' AI ecosystem, and with it both U.S. technology leadership in AI and the prosperity and safety of the American people. Instead, we recommend several steps that can help counter risks, harness the benefits of AI, and leverage openness as a key pillar for trustworthy AI and economic progress:

1. **Impose proportionate and carefully considered regulatory obligations relating to 'open source' AI:** In imposing new rules for AI, regulators should avoid definitional confusion and instead adopt a clear and expansive definition of what 'open source' means in AI, focusing not just on model weights but also consider other components and dimensions of openness in AI. New rules should similarly take into account the special nature and diversity of the 'open source' AI ecosystem and be tailored to account for different contexts of development and commercialization. Finally, rules for openness in AI must consider the marginal risk from openly available AI components.
2. **Support the 'open source' AI community in developing norms and practices around responsibly developing and openly releasing AI models and components:** While we're already seeing efforts across the AI ecosystem to advance the responsible development and public release of different AI components, more support is needed. Governments can play an important role in supporting the development of such norms and practices by promoting and funding research in this area, by enabling access to critical resources needed in AI research and development (see below), and by consulting with and providing guidance for the 'open source' AI community.
3. **Invest in and provide resources for the development and maintenance of 'open source' AI:** The U.S. government should leverage its position as a procurer of technology and a funder of research and development to promote openness in AI and advance the state of the open ecosystem. For instance, the National Science Foundation can fund research and development into open approaches for AI and prioritize research that openly shares technical artifacts (such as model weights, data, or code) created in the process. Similarly, the National AI Research Resource (NAIRR) should prioritize research on open approaches to AI and should be built out to enable better access to compute and other resources

for open research (while not further entrenching the position of those actors already dominating the industry). NIST can advance standards and provide guidance for the responsible development, evaluation, and release of openly available models and other components. And throughout the U.S. government, suitable datasets that can aid AI research and development and lower barriers to entry should be identified and made available.

4. **Involve federal agencies responsible for protecting civil rights, promoting competition, and advancing scientific research in the development of any policy touching on openness in AI:** Openness in AI is not solely a national security issue. It has significant implications for other key priorities of the U.S. government. Any policy or public initiative that advocates constraining openness in AI should not be adopted without consulting with federal agencies whose work may be hindered as a consequence. The agencies protecting civil rights, for instance, will need to ensure that such policies do not hinder them in identifying and investigating potential civil rights violations involving the use of AI systems. Agencies advancing competition in the AI industry will desire certain levels of openness, both because of its value to startups in the AI marketplace and because it could be important for investigations of anticompetitive practices among industry players. Agencies promoting scientific research will care about openness in the context of supporting a robust R&D ecosystem fueled by open and inclusive science. The U.S. Government would be ill-advised to view this solely as a national security issue — or to take a myopic view of national security given the significant implications for U.S. economic and scientific leadership.

We hope these measures can help construct a better AI ecosystem — one that is more innovative, more competitive, and more accountable. As an independent and trusted voice with a long history in the open source movement and firm footing in both the technology industry and civil society, Mozilla stands ready to support these efforts.

Annex

Overview of components across the AI stack and the respective benefits and downsides of their open release.

AI Stack Component	Benefits	Drawbacks
Code ● Data pre-processing ● Pre-training ● Training libraries ● Fine-tuning ● Inference ● Distributed computing ● Infrastructure frameworks	● Enhances reproducibility, auditability, and transparency; ● Enables better independent assessment of quality and fairness; ● Facilitates community participation in model training and understanding training strategies; ● Supports open science and knowledge; ● Makes inference more efficient and cheaper.	● May expose proprietary technologies or intellectual property; ● Could lead to misuse of open-source tools in malicious applications.
Datasets ● Pre-training ● Supervised fine-tuning ● Preference ● Evaluation	● Allows examination of biases and fairness; ● Promotes understanding of model specialization and alignment; ● Increases transparency in evaluation; ● Supports privacy-preserving approaches.	● Risks privacy breaches and security concerns, when the dataset contains personally identifiable information (PII); ● Potential for harmful content propagation proliferation if datasets are not carefully managed.

Model Weights ● Pretrained weights ● Intermediate checkpoint weights ● Downstream task adaptation ● Compressed and adapter ● Reward	● Fosters development ecosystems; ● Aids in linguistic and cultural model tuning; ● Enhances auditability and transparency by enabling independent research and testing; ● Supports research in mechanistic interpretability and architecture efficiency.	● Could compromise model integrity if weights are altered maliciously; ● Potential for unauthorized use and exploitation of pretrained models.
Documentation ● Datasheet ● Model cards ● Evaluation ● Red teaming results ● Publications	● Promotes understanding of data representation and model design choices; ● Aids in designing efficient models and tracking AI carbon footprint; ● Facilitates fair model comparisons; ● Increases foresight into potential model misuse.	● Might reveal sensitive or proprietary information; ● Could create a false sense of trust in a system's capabilities and safety precautions.
Distribution ● License ● Type of Release ● Acceptable Use Policy/Use Restrictions	● Enables ethical and legal risk management; ● Supports staggered scrutiny and societal impact evaluation; ● Documents acceptable uses and feedback mechanisms for rights and redress.	● Legal and ethical implications of widespread access; ● Strict terms may deter safety research.

Guardrails (applies throughout the stack) ● Aligned weights ● Programmable weights ● Safeguard models/safety classifiers ● System prompts	● Improves community testing of safety techniques; ● Enables transparent and auditable content moderation; ● Facilitates evaluation of content moderation robustness; ● Aids in collective determination of effective prompts.	● Risk of enabling adversarial misuse through detailed knowledge of guardrails; ● Potential for creation of bypass strategies that undermine safety measures.