Before the
National Telecommunications and Information Administration
Washington, DC 20230

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

COMMENTS OF PUBLIC KNOWLEDGE

Nicholas P. Garcia
John Bergmayer
Elise Phillips

Public Knowledge
1818 N Street, NW
Suite 410
Washington, DC 20036

March 27, 2024

INTRODUCTION & EXECUTIVE SUMMARY 3
DEFINITIONS & UNDERSTANDINGS 4
Dual-Use Foundation Models 4
Foundation models may not be a significant regulatory category. [Q 1b] 4
General purpose models are generally dual use models 5
Defining Openness for AI Systems. [Q. 1] 6
Openness should be defined for AI systems to promote free, open, and accessible technology 6
AI policy should recognize a broad spectrum of openness for AI systems 7
MARGINAL RISKS [Q2] 8
UNIQUE BENEFITS [Q3, 6] 9
Openness Reduces Barriers to Robust Competition. [Q. 3a, 6b] 9
AI markets have extremely high barriers to entry 10
Openness Enables Rapid and Inclusive Innovation. [Q. 6a] 11
Safety and Accountability. [Q. 3b, 6a] 12
RECOMMENDED POLICY MECHANISMS [Q7] 13
Regulations should be clear, scalable, and flexible 13
Clarity in regulations 14
Scalability of regulatory approaches 14
Flexibility in regulatory regimes 14
A combination of sector-specific and an expert regulatory authority would be best 15
Consider when it is best to target developers, deployers, or bad actors 15
Product Liability Frameworks 16
Traditional Software Liability 16
Reassessing Software Liability for AI Models 18
Certification and testing safe harbor for open foundation models 18
CONCLUSION 19

INTRODUCTION & EXECUTIVE SUMMARY

The discourse surrounding open foundation models—with a particular focus on those models characterized by widely available model weights—stands at the center of key policy and regulatory considerations about the risks and benefits of artificial intelligence (AI) technology. It is the hope of Public Knowledge that the National Telecommunication and Information Administration (NTIA) seeks through this proceeding to navigate the necessity of robust AI safety, security, and responsible development while protecting principles of dynamic competition, inclusive innovation, and access to open technology. Charting such a course necessitates a detailed examination of the definitions that frame our understanding of these models, a balanced consideration of their marginal risks versus their unique benefits, and the formulation of nuanced policy recommendations that support healthy development in the AI sector.

Evaluating Definitions.

To address the multifaceted challenges and opportunities presented by AI, especially open foundation models, it is imperative to clearly delineate the landscape. This includes understanding the nuances of the definitions of "dual-use foundation models" and openness in the context of AI systems, as well as the implications these definitions have for policy and regulation.

Balancing Marginal Risk with Unique Benefits.

The conversation around open foundation models is significantly enriched by a nuanced understanding of the marginal risks they pose compared to their closed counterparts and existing technologies. This understanding is vital for developing AI regulations based on evidence and consideration of the rapid pace of technological advancements. By comparing the ease of misuse in open models against the backdrop of closed models and traditional software, it becomes evident that while open models do introduce certain risks, they also offer unparalleled benefits in terms of fostering innovation, competition, and inclusion. These unique benefits highlight the necessity of evaluating open AI systems within a framework that appreciates their potential to contribute positively to society, especially in democratizing access to AI technologies and enabling diverse voices to shape the development of AI.

Policy Recommendations.

Informed by the evaluations of definitions and the delicate balance between risk and benefit, the path forward entails crafting policy interventions that are clear, scalable, and flexible. Policies must be understandable and predictable to ensure broad participation and investment in AI development. They should also be adaptable to the diversity within the AI ecosystem, acknowledging the varying capabilities and resources of different actors. Moreover, the policy framework must evolve in tandem with AI technologies, ensuring regulations remain relevant and effective. Among the specific recommendations is the reevaluation of product liability in the context of AI, suggesting a potential shift towards a framework that incentivizes the safe and responsible development of AI through mechanisms like testing and certification for open models. Such an approach would recognize the unique challenges and opportunities presented by AI, promoting innovation while safeguarding public welfare.

DEFINITIONS & UNDERSTANDINGS

Investigating the impact of open foundation models, defined in the RFC as "dual use foundation models with widely available weights," necessitates clearly defining and understanding both components of the definition: "dual-use foundation models" and "widely available model weights."

Dual-Use Foundation Models

The terms "foundation model" and "dual-use foundation model" as defined in the RFC and AI EO are critical terms that target the RFC at a specific class of large, general purpose AI models. However these definitions carry with them some implicit assumptions that are critical to unpack. First is that foundation models as a category of AI models are both distinct from other kinds of models and more significant targets for regulation. Second is that "dual-use" is a useful distinction that appropriately targets models which ought to raise significant concern. Both of these assumptions, unpacked further below, build towards the unspoken thesis of the AI EO's directive to investigate dual-use foundation models which is that the small number of powerful foundation models are the best target for regulation or oversight in order to mitigate risks. Considered completely aside from the question of openness, this implicit focus on dual-use foundation models as a class or category of model should be carefully scrutinized.

Foundation models may not be a significant regulatory category. [Q 1b]

The AI EO does not define foundation model apart from dual-use foundation model, but the RFC does independently define foundation models to be "powerful models that can be fine-tuned and used for multiple purposes."[1] Foundation models as a concept are at the heart of the current boom in AI innovation and investment, and so are also centered in these discussions of risks and regulation. Foundation models are currently imagined as very large (in terms of parameters), high cost (due to data and compute costs), and highly generalized but easily adaptable models. But this combination of features may not be durable or reliable as a regulatory target.

Just as quickly as AI has exploded into relevance, it continues to change and advance rapidly. Recent reporting indicates that performance of models at all sizes continues to improve and that smaller models are capable of rivaling the capabilities of the largest foundation models.[2] Similarly, there is increasing demand and development for small models that can run locally on

[1] RFC (citing Rishi Bommasani et al., On the Opportunities and Risks of Foundation Models, arXiv:2108.07258v3 (July 12, 2022). *https://arxiv.org/pdf/2108.07258.pdf.)*

[2] Lauren Leffer, When It Comes to AI Models, Bigger Isn't Always Better, Nov. 21, 2023, https://www.scientificamerican.com/article/when-it-comes-to-ai-models-bigger-isnt-always-better/.

consumer hardware—hardware which also will continue to improve and become more specialized and adapted for the reality of the AI ecosystem.[3] This means that while the foundation model may seem like the best target for high-risk concerns right now, there is no guarantee that the AI ecosystem remains organized around foundation models, cloud access, and fine-tuned deployments. Similarly, advancements in AI techniques and increasing investment and interest in AI may also mean that foundation models rapidly proliferate, which means that even if the current cloud-based paradigm continues to predominate, it isn't clear that foundation models will necessarily continue to exhibit the features that make them attractive regulatory targets right now (i.e. relatively few of them, best route to high levels of performance).

General purpose models are generally dual use models.

The rapid development and emerging capabilities of AI technology implicates critical questions about cybersecurity, national security, economic disruption, and other risks enabled or augmented by AI. Concerns about these uses lie at the center of the AI EO and are reflected in the concept of a "dual-use" model: one that has both benign and dangerous uses.[4] However, models with high performance in these potentially dangerous uses are not just dual-purpose, they are general purpose.

The distinction is significant: a true dual purpose model would be one where the specific beneficial purpose of the model necessarily also gives rise to the dangerous use. For example, an AI system designed to develop patches for software vulnerabilities requires a never similar set of characteristics to a system that is able to write programs to exploit software vulnerabilities; the duality of the benefit and risk in that case is clear, and the developer of such a system—theoretically one familiar with cybersecurity issues—would be in a position to design such a system in a safe and responsible fashion. Yet, what we are more commonly seeing in the current AI ecosystem is that models are capable of robust performance across a wide array of tasks, with the potential for dangerous misuses arising not from the design choices of the system in a specialized area that poses a high risk of dual-use, but from the relatively all-purpose power of the ability to manipulate language—both human and machine—or other forms of media.

It is critical that general-purpose tools, which may have the potential for misuse, are not overburdened or restricted because of potential for misuse. Many technologies, from email to AI tools, have both beneficial and dangerous applications. We can craft policies that incentivize responsible development and deployment, or directly target bad actors, while also not stifling innovation or unduly restricting the development and use of technologies with substantial legitimate uses. As Cory Doctorow argues in the context of general purpose computers:

[3] Clare Conley, Generative AI in 2024: The 6 most important consumer tech trends, https://www.qualcomm.com/news/onq/2023/12/generative-ai-in-2024-6-consumer-tech-trends-for-next-year; https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing

[4] *See* AI EO Sec3(k) available at https://www.federalregister.gov/d/2023-24283/p-25 ("an AI model that…exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety.")

> “[W]e don’t know how to build the general purpose computer that is capable of running any program we can compile except for some program that we don’t like, or that we prohibit by law, or that loses us money. The closest approximation we have to this is a computer with spyware. A computer on which remote parties set policies without the computer user’s knowledge, over the objection of the computer’s owner. And so it is that digital rights management always converges on malware.”[5]

LLMs and other robustly performing, highly generalized models are even more remarkably emergent in their capabilities than the general purpose computer, and it is very likely to be the case that attempts to impose very stringent requirements on AI systems will result in draconian policies and AI systems that create entirely new problems for privacy, free speech, and technological innovation.

Defining Openness for AI Systems. [Q. 1]

Openness should be defined for AI systems to promote free, open, and accessible technology.

The terms “open” or “openness” emerge from the open source software movement carrying with them a history of values, norms, and practices. Alternative terminology surrounding the openness of software, including “open source,” “open access,” "free," "libre," and other terms each reflect variations in underlying values. While “open” may be the most familiar or encompassing term, understanding these distinctions and nuances is crucial in the AI domain, building off of their history in the software industry. These terms convey different aspects of accessibility, rights to modify and redistribute, and the level of transparency, each contributing uniquely to the discourse on AI governance.

In the open-source software movement, "free" and "libre" emphasize the freedom to use, modify, and distribute software without restrictions, highlighting the rights of users to control their technology, not merely as consumers but as active participants. These terms, championed by individuals like Richard Stallman and organizations like the Free Software Foundation, assert an idealistic stance regarding technology, insisting that software should serve humanity, not the other way around. Since "free" often leads to ambiguity with "free of charge," the term "libre" has also been used interchangeably with “free” to emphasize the focus on liberty. This philosophical stance is embodied in the GNU Project, advocating for users' rights to use, study, modify, and distribute software.

In contrast, "open" or "open source" stresses accessibility and the practical benefits of sharing source code, including collaborative development and innovation. The open-source movement, led by figures like Linus Torvalds with the Linux kernel, focuses on the practical benefits of accessible source code for development efficiency and innovation. While "open source" conveys the technical openness and collaborative potential, "free" and "libre" stress the

[5] Cory Doctorow, The Coming War on General Purpose Computing, http://opentranscripts.org/transcript/coming-war-general-computation/

user's freedoms, underlining a distinct ideological stance on software freedom and rights. Mozilla, a proponent of "open source," exemplifies how openness can foster innovation and collaboration, as seen in the Firefox browser project that competes with—and even outcompetes—the products of the largest tech companies like Google and Apple.

In AI, these principles translate into debates over the accessibility of AI models and datasets ("open"), the rights to modify and use AI for any purpose ("free" and "libre"), and the transparency of AI operations. OpenAI's shift from total openness to a more controlled model release with GPT-3 illustrates the tensions between innovation, commercial motive, risk mitigation, and openness. Currently, the terms “open” or “openness” may represent the best terminology for understanding the gradient of accesses and transparency into AI systems and their components, but understanding the nuances and distinctions between these other terms—and what the different values and policy goals these alternative terms bring with them—is worthy of consideration in the AI context just as it is for software.

AI policy should recognize a broad spectrum of openness for AI systems.

The Administration’s report should reflect the complexity and diversity of AI systems when it comes to openness, acknowledging that the concept of openness in AI differs fundamentally from that in traditional open-source software. Unlike software, which primarily comprises code, AI systems are multifaceted, encompassing not just the algorithms but also the data on which they are trained, the infrastructure they run on, and the processes that develop them. This complexity necessitates a nuanced understanding of openness, one that transcends the binary perspective of open versus closed systems. A truly open AI system involves much more than providing access to model weights. By way of analogy, releasing model weights alone is closer to the practice of releasing compiled software as freeware, than as free or open source software. Just as true openness in software requires that source code be available, true openness in AI systems requires a greater degree of transparency than simply providing a finished product for download.

First and foremost, model architecture documentation is essential. This involves providing detailed descriptions of the AI model's design and structure, enabling others in the field to understand, replicate, and potentially innovate upon the existing model, thus fostering a collaborative environment for growth and advancement. Relatedly, model cards play a crucial role in summarizing essential information about the AI model. These summaries should include details about the training process, the compute resources used, and other characteristics that give a comprehensive overview of the model. Model cards serve as a bridge between developers and users, offering a concise yet informative snapshot of the model's features and facilitating a better understanding of its potential applications and limitations.

Furthermore, the software code used throughout the model's lifecycle—from data pre-processing, training, validation, to testing—must also be accessible and, ideally, permissively licensed. This transparency allows for a deeper understanding of the model's development process, offering insights into how data is prepared, how the model learns, and how its performance is evaluated. Such openness enhances the model's credibility and reliability. Permissively licensing this code under existing open source licenses will take this a step further, not only allowing examination but also facilitating improvements and adaptations by the broader community. Similarly, the inference process, or how the model makes predictions, is another critical area where openness is necessary. Sharing the code used for model inference provides users with a clearer picture of how outputs are generated, which in turn builds trust in the AI system's capabilities and decisions.

Perhaps most importantly, access to the contents of—or at least detailed information about—the training, testing, and evaluation datasets is critical. By making these datasets available, developers ensure that the model's performance can be independently verified. This not only helps identify and correct biases or errors but also contributes to the collective understanding of the model's strengths and limitations.

Every step taken towards more openness should be recognized and rewarded, as it contributes to a culture of transparency and accountability in AI development. Advocating for maximal responsible openness means encouraging the AI community to share as much information as possible about their AI systems, while also respecting privacy, security, and ethical considerations.

Conversely, policies must also acknowledge that systems which remain fully or mostly closed—thereby avoiding public scrutiny—present significant risks. These risks range from perpetuating biases and errors in AI systems to hindering innovation by restricting access to potentially groundbreaking technology. As such, maintaining a closed system should come with additional responsibilities and oversight to mitigate these risks. This approach ensures that while the AI community is encouraged to move towards greater openness, those who choose to keep their systems closed are held to a higher standard of accountability, ensuring that their AI systems are developed and used responsibly and ethically.

MARGINAL RISKS [Q2]

Public Knowledge recently joined with a diverse coalition of other civil society organizations and academics in a joint letter to the NTIA regarding this proceeding, emphasizing the need for a balanced evaluation of the risks and benefits associated with both open and closed AI models.[6] The Civil Society Letter's message is straightforward: it's crucial to develop AI regulations that are grounded in evidence and take into account the fast-paced nature of technological advancements. The goal is to create a regulatory environment that supports the positive development of AI technologies while mitigating their risks, ensuring that AI can be a beneficial tool for all members of society.

In particular, the Civil Society Letter emphasizes the importance of evaluating the risks of open foundation models using a marginal risk framework. The letter reflects a growing consensus that it is important to evaluate the risks of open models in comparison to the risks and

[6] https://publicknowledge.org/policy/ntia-joint-letter/ ("Civil Society Letter").

benefits from closed models and pre-existing technologies like the internet.[7] Policy should be based on clear evidence of marginal risks that open models pose compared to closed models. Put another way, what is the difference in risk created by the existence of the open model, given the available capabilities of closed models and other technologies?

Taking a closer look at concerns such as disinformation campaigns, one must weigh the potential impact of open models against the backdrop of existing tools and closed models. For instance, the accessibility of open models for being theoretically modified to create disinformation should be balanced against the use of existing software like Photoshop and closed AI models which also possess capabilities for misuse. This comparison is crucial, not only to assess the relative ease of misuse but also to highlight the broader context in which these technologies operate.

Meanwhile, open models can often provide unique benefits compared to closed models, as outlined further below. We therefore urge the Administration to be rigorous in evaluating and targeting the specific risks from openness in AI.

UNIQUE BENEFITS [Q3, 6]

While open foundation models generally present only marginal risks compared to closed systems, they present unique benefits that cannot be replicated with closed systems.

Open systems reduce barriers to robust competition whereas closed systems further entrench existing incumbents. The nascent AI industry is vulnerable to concentration to capture and history has proven that openness and competition has always proven to be an asset in the American technology sector.

Making model weights and other elements of AI systems widely available also promotes rapid and inclusive innovation, not just through new competitors in the market, but by allowing existing competitors to rely on shared resources that improve for all as they are improved in common. Open innovation also means inclusive innovation: making the building blocks of powerful technology like AI available to individuals and communities means that more diverse and marginalized voices, ideas, and perspectives can contribute to the development of AI technology.

Finally, making AI more open also will make it safer, more accountable, and more trustworthy. Public access to model weights and other components of AI systems allow for individual, academic, and civil society investigation, testing, and oversight of AI systems.

Openness Reduces Barriers to Robust Competition. [Q. 3a, 6b]

A market already entrenched in favor of incumbents is less accountable and less innovative. AI is poised for integration across systems and industries, which may gravely impact essential environmental, healthcare and labor markets. Companies can help balance the scales and ensure that AI markets clear a path to innovation, opportunity and greater consumer choice.

[7] *Id.*; Sayash Kapoor et al., “On the Societal Impact of Open Foundation Models,” Center for Research on Foundation Models (CRFM), Stanford University, February 2024.

Through open systems, developers can have the chance to compete fairly both on and against Big Tech platforms and help foster a healthier, innovative future.

Competition and openness in tech historically made the American tech sector vibrant, innovative, flourishing.

Dominant companies often utilize gatekeeper power to further their own market power and cut off new entrants from the chance to compete. Open technologies may serve to counteract this exclusionary conduct and lower barriers to entry for innovative, up-and-coming rivals. Historically, we've already seen how open access to technology patents had competitive benefits, leading to a wellspring of innovative products.

The 1956 consent decree that settled the seven-year antitrust lawsuit against Bell Labs boosted tech development, mostly by small and young companies building on Bell's established technologies. The decree contained a provision that Bell Systems patents be licensed to competitors on request.[8] While these patents did not serve as the sole counterweight to gatekeeper power,[9] it may be concluded that this remedy played an important role in increasing product innovation.[10] As such, accessibility to open technologies can clear the path to opportunity and greater consumer choice. Much like these patents, open model weights are poised to lower barriers to entry in the AI marketplace.

AI markets have extremely high barriers to entry.

AI markets currently possess unique and extremely challenging barriers to entry, exacerbated by the existing market power of large companies. These obstacles make it increasingly difficult for developers to participate in an evolving economy that aims to integrate AI into the status quo.

The computing power needed to build the most highly performing AI systems from scratch is massive—only a select few companies possess the computing capabilities necessary to do so. Open source models can remove the need for that compute-intensive pre-training stage and give potential competitors a head start, at minimum freeing them from needing special relationships with cloud providers to access huge amounts of computing power and develop these systems.[11]

[8] The consent decree contained two main remedies. The Bell System was obligated to license all its patents royalty free, and it was barred from entering any industry other than telecommunications. As a consequence, 7,820 patents, or 1.3% of all unexpired US patents, in a wide range of fields became freely available in 1956.

[9] https://cepr.org/voxeu/columns/how-antitrust-enforcement-can-spur-innovation-bell-labs-and-1956-consent-decree (arguing that compulsory licensing is found to be ineffective in markets where dominant firms have other means of market foreclosure)

[10] https://economics.yale.edu/sites/default/files/how_antitrust_enforcement.pdf ("We conclude that antitrust enforcement can play an important role in increasing innovation by facilitating market entry. Several antitrust scholars have argued that antitrust enforcement should pay special attention to exclusionary practices because of their negative influence on innovation.")

[11] https://publicknowledge.org/challenging-big-tech-in-the-age-of-ai/

Additionally, large tech firms with existing troves of user data have an outsized market advantage, using said data to create and train AI products. Moreover, access to this training data may be restricted from the public. Open source model weights, commercially available data warehouses, and public compute resources would enable many new model developers to use the data to develop and train new models. In addition, foundation models and APIs could also be opened, so that developers have reliable access to these resources.[12]

Openness Enables Rapid and Inclusive Innovation. [Q. 6a]

Openness in AI, particularly through the wide availability of model weights and other foundational elements, acts as a catalyst for rapid and inclusive innovation. This openness not only paves the way for new entrants to challenge Big Tech incumbents but it also empowers existing entities to leverage a collective pool of resources. These shared resources, enhanced collectively, become increasingly robust, driving improvements across the board. Moreover, open innovation inherently embodies the principle of inclusivity. By democratizing access to the critical building blocks of AI, we ensure that a broader spectrum of individuals and communities, especially those marginalized or underrepresented, can have a voice in shaping the future of technology. This approach enriches the AI landscape with diverse ideas, perspectives, and solutions, fostering a technology that is truly reflective of the communities it serves.

The economic implications of open source in software development have long been recognized, with recent estimates attributing over $8 trillion in value to open-source software, which constitutes 96% of commercial software.[13] This immense value underscores the transformative potential of openness, a principle that the U.S. government, among the world's largest users of open-source software,[14] actively supports. Through funding open-source initiatives that range from enhancing cybersecurity to combating cancer, the government acknowledges the strategic advantage and societal benefit inherent in open development practices.[15]

Applying these principles to AI, open models significantly lower the entry barriers for innovators, startups, and small businesses, particularly those from diverse communities. These models facilitate scientific advancement by being more affordable, easier to customize, and

[12] https://cdn.vanderbilt.edu/vu-sub/wp-content/uploads/sites/281/2023/12/19183408/Policy-Brief-2023.10.08-.pdf

[13] Manuel Hoffman et al., “The Value of Open Source Software,” Harvard Business School, January 2024; 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.”).

[14] 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.

[15] 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.

conducive to reproducible research. This aspect of open development is crucial in serving varied interests that might not align with immediate commercial priorities. Large corporations, driven by profit motives, may overlook individuals or niche markets, however profitable they might be, in favor of more lucrative opportunities. This market dynamic often leaves underserved communities without the benefits of AI technologies tailored to their unique needs and circumstances.

Open AI systems, by their nature, challenge this status quo by enabling stakeholders outside of the tech giants to influence AI's trajectory. This inclusivity ensures that AI services and applications are developed with a broader range of needs and perspectives in mind, promoting diversity, equity, and inclusion. Such a broad-based participatory approach to AI development ensures that the technology is not only economically valuable but also socially equitable, offering solutions that cater to the wider spectrum of human experience and need. In this way, the ethos of open AI systems aligns with the larger goal of creating technology that benefits all of society, fostering innovations that are as diverse and dynamic as our nation.

Safety and Accountability. [Q. 3b, 6a]

The journey toward greater safety and accountability in AI is significantly enhanced by fostering openness within AI systems. By making AI models more accessible—particularly through the public availability of model weights and other integral components—these systems become not only more transparent but also more amenable to scrutiny and evaluation. This level of openness is crucial for enabling a wide array of stakeholders, including individuals, academics, and civil society organizations, to conduct thorough investigations, perform tests, and maintain oversight of AI technologies. While closed systems may seem to present security benefits, it is openness that inherently promotes safety, security, and trustworthiness, offering a pathway to a more accountable and reliable AI ecosystem.

Open AI systems allow for independent and external analysis, crucial for identifying and addressing potential risks and vulnerabilities. This capability has profound implications for cybersecurity and safety, drawing parallels to the positive impact open-source software has had in these areas. Community scrutiny of systems, testing for vulnerabilities or bugs, and repair of problems can greatly enhance the security of a system when a proprietary, closed model would struggle. By adopting a similar ethos of openness in AI, the community can inherit the same beneficial qualities of open-source software—while also learning from its challenges and building in best practices from the start.

Additionally, open AI models facilitate a more straightforward process for regulators and civil society to assess AI systems' compliance with laws aimed at protecting civil rights, privacy, consumers, and workers. This increased transparency not only elevates the level of public education and testing but also enhances trust in AI technologies. It empowers researchers and journalists to audit AI systems and scrutinize their impact on various demographic groups, providing a critical check on the power of AI developers and deployers.

The advancement of safety and security through open models is a dynamic process. It involves accelerating our collective understanding of AI's capabilities, risks, and harms through the lens of independent research, collaboration, and knowledge sharing. Such an approach is invaluable for regulators and researchers who rely on the latest methodologies, tools, and insights to effectively monitor and test large-scale AI systems. This ecosystem of open innovation and evaluation is especially vital given the current lack of binding safeguards, guardrails, or testing and accountability standards for closed AI models. The current investigation highlights the paradox of imposing requirements that might deter openness when, in reality, openness serves as the sole mechanism for any level of publicly accountable transparency in AI systems in the current regulatory environment.

While transparency and openness are not a panacea for all the challenges posed by AI, they establish a foundation or baseline whereby this key technology can in some way be scrutinized, challenged, and improved. At the same time, developers and deployers, particularly those with substantial resources and risk profiles, can and should bear responsibility for their products; openness is not a free pass to innovate irresponsibly. As discussed further below, the responsibility placed on developers and deployers should be proportional to their resources, the specific risks posed by their products, and their degree of openness. Those who actively contribute to a common ecosystem where safety, accountability, and trust in AI are paramount should be rewarded, while those who keep their systems closed to capture greater commercial value must also take greater responsibility to ensure their systems meet the highest standards. In this way, the push for open AI systems aligns with broader societal goals, ensuring that AI technologies serve the public good while safeguarding against potential harms.

RECOMMENDED POLICY MECHANISMS [Q7]

Rules, regulations and policies should be clear to promote certainty, scalable to account for the wide range of actors in the AI ecosystem from individual hobbyists to the largest technology companies, and flexible to respond adaptably to the fast-paced nature of technological change in this sector.

Legal liability and the burden of rules should be targeted carefully depending on the risks and harms that need to be mitigated. Data gatherers or providers, model developers, AI system deployers, and end users themselves may each prove to be the right target depending on the issue. Placing responsibility too far upstream or downstream risks not only ineffective mitigation that will allow harms to persist or come to pass, but also may create new harms to innovation, adoption, or user rights.

Regulations should be clear, scalable, and flexible.

As the Administration turns its attention towards policy recommendations for regulatory frameworks for open foundation models, or AI systems more generally, it should keep in mind how to safeguard innovation, broad participation, and user rights. For any regulatory efforts to be both effective and conducive to those foundational principles, they must be anchored in three features: clarity, scalability, and flexibility.

Clarity in regulations.

Clear and predictable rules are not merely administrative niceties; they are vital cogs in the machinery of innovation and participation. Clarity in regulation provides a stable foundation upon which individuals, academics, and civil society organizations can build without the looming shadow of legal risk. This clarity is particularly crucial for those with limited resources, for whom legal ambiguity poses a disproportionate threat. In contrast, large, well-resourced firms might navigate or even exploit regulatory uncertainties to their advantage, potentially stifling competition and innovation. Clear regulations level the playing field, encouraging investment and participation across the spectrum of society.

Scalability of regulatory approaches.

To foster an environment where a variety of open AI systems can thrive, regulations must be adeptly scaled to reflect the diversity of entities operating within this space. This scalability ensures that the regulatory burden does not disproportionately impact smaller players or stifle innovation. Factors to consider in scaling regulations include the commercial versus non-commercial nature of model development, the access to compute power available to a developer, and the size of a deploying company (which could be measured in terms like user base or market capitalization). Such a nuanced approach to scalability supports broad participation, robust competition, and encourages public engagement in AI development, ensuring that the benefits of AI are widely accessible.

Flexibility in regulatory regimes.

The rapid pace at which AI technology advances demands regulations that are not only technology-neutral but also adaptable to unforeseen changes. A flexible regulatory approach acknowledges the limitations of our current understanding and anticipates the need for adjustments as technology evolves. This humility in regulatory design is essential to avoid the pitfalls of past efforts, which either failed to evolve with technological advancements or, conversely, imposed rigid structures that quickly became obsolete. As Public Knowledge has previously advised the Administration, "It is easy to look back on regulatory efforts from the early days of computers and the internet and see laws and policies—many of which we are still living with—that missed the mark; not because of misaligned intentions or lack of state-of-the-art knowledge, but because of a narrow view of the technology as it currently existed and reliance on rigid or overly prescriptive models that failed to evolve with the times and technology. In other domains, policymakers failed to take action at all, allowing practices like commercial surveillance, addictive attention-based design, and anticompetitive consolidation to take hold."[16] Flexibility in regulation allows for a responsive and dynamic approach to AI governance, one that can adjust to new developments and challenges as they arise.

A combination of sector-specific and an expert regulatory authority would be best.

The path to effective AI regulation, therefore, lies in balancing these three principles. It involves crafting rules that are clear enough to provide stability and predictability, scalable to ensure fairness and encourage wide participation, and flexible enough to adapt to the rapid pace of technological innovation. Such a regulatory approach demands a combination of deep sector-specific expertise and the capacity for broader, ecosystem-wide oversight. It suggests a hybrid model where sector-specific regulators address immediate and foreseeable harms, complemented by a centralized AI regulatory body equipped to adapt alongside the technology. This dual approach maximizes the potential for AI to benefit society while minimizing risks, ensuring a future where AI systems are developed and deployed responsibly, ethically, and inclusively.

Consider when it is best to target developers, deployers, or bad actors.

To foster an environment where innovation thrives alongside robust accountability mechanisms, it's pivotal to delineate the responsibilities of different actors within the AI ecosystem. This demarcation is crucial not only for addressing harms effectively but also for providing academic researchers, market newcomers, and users with a clear understanding of their obligations and assurances regarding the risks involved. The distinction between the responsibilities of developers, deployers, and bad actors must be well-defined, reflecting on the nature of potential harms and the best stage for intervention—whether at the model's development, its deployment, or in curbing misuse by malicious end users.

Understanding where to address potential harms in the lifecycle of AI technologies is essential. Certain risks are inherently linked to the model itself and are best mitigated during the development phase. Conversely, other harms may emerge more prominently during deployment or usage, necessitating a focus on deployers or users. Keeping these different scenarios in mind allows for a more nuanced approach to AI governance, ensuring interventions are both timely and effective. In any event, transparency plays a fundamental role in ensuring accountability and facilitating the identification and rectification of biases or other issues. This transparency needs to extend across the entire AI development and deployment chain.

For example, issues such as ingrained biases may be best identified and addressed at the earliest stage within the model. Open models provide an advantage here, as they allow for the examination and reevaluation of the data and algorithms in light of bias concerns. This level of openness not only promotes corrective actions but also fosters a culture of accountability from the outset. On the other hand, if the biases of a model are known and documented, they may be more easily mitigated at the deployment stage, where the deployer has a better understanding of the ultimate use case and risks.

[16] PK NTIA AI Accountability Comments

Deployers of AI technologies carry a responsibility to ensure that the models they utilize do not result in unlawful or dangerous outcomes. This obligation raises questions about the potential burden it places on deployers. However, it posits a principle: if an entity has the capacity to deploy a complex AI model, especially commercially, it should inherently possess—or ensure access to—the necessary resources to vet the model for risks and ensure it is compliant with laws and regulations.

In yet other instances, it may be the case that placing responsibility on either developers or deployers to prevent certain misuses of remarkably general purpose technology like an LLM is simply not viable. Such requirements could suppress innovative development, restrict beneficial use cases, and impose onerous and unnecessary restrictions on non-malicious users. In those cases, enforcement is best targeted at bad actors and other malicious end users, or at the harmful effects of their conduct, rather than at the tool itself.

Ultimately, developing an effective accountability system in AI requires a delicate balance. It necessitates a framework where the responsibilities of developers, deployers, and users are clearly defined and aligned with the goal of minimizing harms while maximizing the potential for positive societal impact. Through a combination of transparency, proactive mitigation of risks at the development level, and thorough vetting by deployers, the AI ecosystem can evolve into one that not only innovates but does so responsibly and ethically.

Product Liability Frameworks.

One example of an existing legal liability regime that may prove particularly significant to reimagine in the context of AI models is product liability. The rapid advancement and integration of artificial intelligence into our daily lives, including through the proliferation of open AI models, presents a complex challenge for existing product liability legal regimes.

Traditional Software Liability.

Traditional software has often been shielded from strict liability due to its intangible nature and complex functionality.[17] Courts have generally permitted software developers to disclaim liability through license conditions, even though general disclaimers of liability for foreseeable defects and harms are typically ineffective for tangible consumer products.[18] When it comes to these products—from power tools to children's toys to household appliances—manufacturers and vendors typically do not get to simply decide the scope of their liability.

The common law of torts as described in the Restatement (Second) of Torts § 402A, imposes liability on manufacturers for defective products that cause harm, irrespective of any disclaimers. Restatement (Second) of Torts § 402A (1965). Under the updated Third

[17] Frances E. Zollers et al., *No More Soft Landings for Software: Liability for Defects in an Industry That Has Come of Age*, 21 Santa Clara Computer & High Tech. L.J. 757, 769 (2004).

[18] Jacob Kreutzer, *Somebody Has to Pay: Products Liability for Spyware*, 45 Am. Bus. L.J. 61, 73 (2008) ("product liability claims are not barred by disclaimers or license agreements accompanying consumer goods.").

Restatement, “strict liability continues to apply to cases involving manufacturing defects.”[19] The Fifth Circuit has stated,

> [M]anufacturers [are held] to the knowledge and skill of an expert. They are obliged to keep abreast of any scientific discoveries and are presumed to know the results of all such advances. Moreover, they each bear the duty to fully test their products to uncover all scientifically discoverable dangers before the products are sold.[20]

Similarly, the Uniform Commercial Code (UCC) has established implied warranties of merchantability and fitness for a particular purpose, which cannot be easily waived.[21]

By contrast, sweeping liability disclaimers are standard in the software industry. The current license terms for Microsoft Windows state,

> Neither Microsoft, nor the device manufacturer or installer, gives any other express warranties, guarantees, or conditions. Microsoft and the device manufacturer and installer exclude all implied warranties and conditions, including those of merchantability, fitness for a particular purpose, and non-infringement.[22]

These disclaimers are common in free and open source software as well. The GPL 2.0 states (in all caps):

> BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.[23]

Similar broad disclaimers are considered standard practice in software.[24]

[19] Michael D. Scott, Tort Liability for Vendors of Insecure Software: Has the Time Finally Come?, 67 Md. L. Rev. 425 (2008).

[20] *Dartez v. Fibreboard Corp.*, 765 F. 2d 456, 461 (5th Cir. 1985).

[21] U.C.C. § 2-314 (implied warranty of merchantability); U.C.C. § 2-315 (implied warranty of fitness for a particular purpose).

[22] Microsoft Software License Terms for Windows 11, https://www.microsoft.com/en-us/UseTerms/Retail/Windows/11/UseTerms_Retail_Windows_11_English.htm.

[23] Free Software Foundation, Inc., GNU General Public License, Version 2, https://www.gnu.org/licenses/old-licenses/gpl-2.0.en.html.

[24] Law Insider, Disclaimer of Software Sample Warranty Clauses, https://www.lawinsider.com/clause/disclaimer-of-software-warranty.

Reassessing Software Liability for AI Models.

The characteristics of AI models, in particular their increasing integration with and control of products where traditional liability principles hold, demands a reassessment of this approach. AI models, with their ability to "learn," adapt, and make decisions in ways that are complex and challenging to audit, introduce a novel set of risks that necessitate the application of traditional product liability principles, albeit tailored to their distinct features.

The potential impact of AI models is vast, as they are increasingly being deployed across various sectors, from healthcare and finance to transportation and education. As AI becomes more deeply embedded in critical decision-making processes, the potential harms are greater than before. A malfunctioning AI model could lead to significant harm, whether it's a self-driving car causing an accident or a medical diagnosis system providing incorrect treatment recommendations.

In light of these risks, we should hold the developers of AI models to a higher standard of accountability than applies to the broader software industry. This is not to suggest that innovation should be stifled; rather, it is a call for responsible innovation that prioritizes the safety and well-being of the public.

Various commenters have argued that traditional liability principles should apply to software that has the potential to cause physical injury or other serious harms.[25] Approaches like this are not only fairer, but more economically efficient, incentivizing the reduction of accident costs by those who are best positioned to do so.[26]

AI is complex and carries risks of these kinds. But it is also potentially transformative, creating broad economic and social benefits. A balanced approach to liability is needed—one that encourages innovation while ensuring public safety.

Certification and testing safe harbor for open foundation models.

As discussed above, open models provide significant benefits in terms of innovation, safety, and accountability through the capacity for rapid iteration and improvement, transparency, and testability. Nevertheless, reliance on community oversight, development, and accountability also poses a significant challenge for applying liability. To accommodate this challenge—while incentivizing the development and adoption of open models— liability regimes can be adapted to provide safe harbor for open models that meet pre-release standards through a testing and certification process for open AI models,[27] similar to the safety standards that exist in other

[25] Frances E. Zollers et al., *No More Soft Landings for Software: Liability for Defects in an Industry That Has Come of Age*, 21 Santa Clara Computer & High Tech. L.J. 757, 769 (2004).

[26] T. Randolf Beard et al., *Tort Liability for Software Developers: A Law & Economics Perspective*, 27 J. Marshall J. Computer & Info. L. 199, 206 (2009); Susan W. Brenner, *Toward a Criminal Law for Cyberspace: Product Liability and Other Issues*, 5 Pgh. J. Tech. L. & Pol'y 1, 88 (2004) ("The doctrine of product liability evolved in the context of civil tort litigation between consumers and manufacturers...to give manufacturers incentives to design safer products.").

[27] *See* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353 (2015), for one such proposal.

industries, such as automotive[28] and pharmaceuticals.[29] This process would serve as a quality control mechanism, ensuring that AI models meet safety and ethical standards before they are released to the public. In return, under this approach, open AI models that successfully navigate testing and certification processes have met their duty of care and would be subject a more lenient liability regime, akin to that of traditional software. This creates a powerful incentive for developers to prioritize the safety and reliability of AI systems, as they would be able to demonstrate due diligence and adherence to best practices. Developers of open models would not necessarily be required to certify their products—but they would then be strictly liable for reasonably foreseeable harms their products might cause.

There are several benefits to implementing a certification process for open AI models. First, it would establish a clear benchmark for what constitutes a "safe" AI, providing a framework for determining liability in the event of an accident or harm. Second, it would help to build public trust in AI technologies by demonstrating a commitment to transparency, accountability, and safety. Third, it would foster a culture of continuous improvement and monitoring within the AI community, as certification would likely require periodic review and renewal to ensure that AI models remain safe and reliable as they evolve.

By adopting an adaptive regulatory regime and continuously evolving testing standards, we can keep pace with the rapid advancements in AI technology. This approach has proven successful in other industries, such pharmaceuticals,[30] where products are subject to rigorous testing and monitoring throughout their lifecycle.

While traditional software has enjoyed a unique liability regime that often shields it from strict liability, the distinctive features and potential risks associated with AI demand a different approach. By embracing traditional product liability principles, coupled with a comprehensive testing and certification process, we can strike a balance between fostering innovation and safeguarding the public interest–or rather, ensuring that innovation and the public interest are aligned. This approach aligns with the principles of fairness, protection of public welfare, and responsible innovation, providing a pathway for the safe and beneficial use of AI.

CONCLUSION

The path forward involves a balanced approach that recognizes the potential of open foundation models to drive innovation and inclusion while addressing the inherent risks through targeted, evidence-based policy interventions. By embracing the principles of openness, transparency, and responsible innovation, policymakers can ensure that AI technologies evolve

[28] 49 U.S.C. §§ 30101-30169.

[29] 21 U.S.C. §§ 301-399f.

[30] FDA, FDA-TRACK: Center for Drug Evaluation & Research - Post-Approval Safety Monitoring, https://www.fda.gov/about-fda/fda-track-agency-wide-program-performance/fda-track-center-drug-evaluation-research-post-approval-safety-monitoring ("After a drug is approved, that same drug can be taken by thousands or even millions of patients. With this large-scale use, new risks and new information about the drug's effectiveness are often found. FDA maintains a system of postmarketing surveillance and risk assessment programs to identify adverse events that did not appear during the drug approval process.").

in a manner that is both beneficial and safe. We thank the Administration for the opportunity to engage with these critical issues.