databricks

Databricks, Inc.

Response to the NTIA Request for Comment on
Dual-Use Foundation Artificial Intelligence Models with Widely Available Model Weights

Docket No. NTIA-2023-0009
March 27, 2024

Databricks, Inc. (Databricks) appreciates having the opportunity to respond to the National Telecommunications and Information Administration's request for comment on Dual-Use Foundation Artificial Intelligence Models with Widely Available Model Weights. This topic is extremely important because of the many substantial benefits provided by allowing powerful AI models to be openly available, particularly the positive effects openness drives in enhancing AI democratization, innovation, research, competition, economic productivity and economic growth.

Databricks' Unique Vantage Point with Respect to AI Open Models and Their Benefits

Databricks provides a high-performance cloud-based data processing and hosting platform, which is optimized, and heavily utilized, for AI services and applications. This platform and Databricks' extensive experience with AI have enabled Databricks to become a leading provider of AI model development, customization, hosting and monitoring services. As part of these services, Databricks has developed and released high capability open models, including DBRX, currently the most capable open model available.[1] Databricks has thousands of enterprise and government customers utilizing its AI related platform services around the world.

An extremely important benefit of permitting open models is to give businesses and other organizations the ability to cost effectively obtain, control and modify their own AI models and AI applications using their own proprietary data, which in turn greatly enhances their ability to innovate, conduct research, and improve the functions of their organizations. Databricks is heavily engaged in helping organizations obtain, customize, run and monitor open models for such purposes. We are observing a rapidly growing number of enterprises and public sector organizations turning to open models because closed models present challenges relating to cost of ownership and operation, constraints on modifiability, and risks around the access to, and security of, sensitive data used in training and inference.[2]

From this vantage point, Databricks has a distinctive perspective on the tremendous benefits open models provide to businesses and other organizations. The availability of open models offers organizations greater ability to use AI to significantly enhance their productivity, service and product quality, employee and customer experiences, and new product development, and these benefits from openness in AI are growing as the capabilities of AI improve.

---

[1] See *Inside the Creation of the World's Most Powerful Open Source AI Model*, Wired, March 27, 2024, https://www.wired.com/story/dbrx-inside-the-creation-of-the-worlds-most-powerful-open-source-ai-model/. See also *Introducing DBRX, a New State-of-the-Art Open LLM*, Databricks blog posting, March 27, 2024, https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm.

[2] For a survey-based discussion of the growing interest in customizable open models among Fortune 500 leaders, see *16 Changes to the Way Enterprises are Building and Buying Generative AI*, Andreesen Horowitz blog posting, March 21, 2024, https://a16z.com/generative-ai-enterprise-2024/.

Summary of Databricks' Views

Our views, which are presented in detail in our responses to many of the questions presented in the Request for Comment, are summarized as follows (our responses to selected questions follow this summary):

I. **The benefits of open models substantially outweigh the marginal risks, so open weights should be allowed, even at the frontier level -** The benefits of open models include their significant contribution to "democratizing" AI - i.e., broadening the dissemination of the advantages of AI to individuals, communities, small businesses, researchers and others, by making AI more affordable, accessible and transparent. Openness drives innovation, science, safety research, competition, economic productivity, and economic growth. The marginal risks of highly capable models that are open as compared to closed are not proven and are not incrementally significant. In contrast, the substantial benefits are clear and compelling, and are being amply demonstrated through the use of AI open models by numerous Databricks customers and others every day. Importantly, limiting frontier AI to closed models will concentrate the growing power of AI in the hands of a few dominant tech companies, restricting competition, innovation, transparency and breadth of adoption, as well as driving up costs and potentially limiting their application for beneficial uses.

II. **Regulation of highly capable AI models should focus on consumer-facing deployments and high risk deployments, with the obligations focused on the deployer -** Regulation should be based on the risks related to the actual use case, with appropriate guardrails required at the time of deployment based on the specific use case. Obligations should focus on the deployer rather than the developer. There are two basic categories of deployment: (a) broadly consumer facing deployments (such as subscription access to ChatGPT, online social media platforms, etc.), where consumer-protection guardrails should generally be required (applicable to the deployer); and (b) business-to-business applications where appropriate guardrails may be required (applicable to the deployer) if the use is high risk (e.g., loan qualification, access to important services, use of personally identifiable information, etc.), but may instead involve relatively low risk uses where mandated guardrails are not warranted, with deployers allowed to exercise their judgment as to appropriate safety measures. For open models, any obligations applicable to the developer should apply to the development stage, up to and including reasonable documentation as of the model release date.

III. **If policy makers determine at some point that access to the model weights of certain ultra-capable models should be gated, a registration "allow list" system should be utilized that strikes an appropriate balance, avoiding overly burdensome requirements that would deter innovation and competition except by a few dominant tech companies -** In Databricks' view, such a gating system is not reasonably necessary under current model architectures, but under future architectures, perhaps associated with artificial general intelligence (AGI) or higher capability levels, and if risky model behaviors are reasonably anticipated, such a system may be justified. The registration system Databricks contemplates is described in response to question 7(a). In summary, it would involve verification of identification, residency and contact information, standardized background checks, and checks against country, organization and individual deny lists. It would also require affirmation that: no distribution of model weights would occur except to other properly registered recipients; required safeguards would be maintained (or added, depending on use case); and use would not be in violation of any laws or

stipulated usage restrictions. The registration process would be comparable to the types of processes successfully used by FINRA for brokerage employees, for federal gun ownership background checks, and for the U.S. Global Entry (airport security clearance) program. Registration would be required of each individual given access to the model's weights. Violations would be subject to regulatory enforcement and potential criminal sanctions in the case of willful violations.

Databricks' Responses to Specific Questions:

Question 1. *How should NTIA define "open" or "widely available" when thinking about foundation models and model weights?*

In evaluating the benefits, marginal risks and policy recommendations relating to open dual-use foundation models ("Open DUFMs") as compared to closed dual-use foundation models ("Closed DUFMs"), the most important thing to consider in determining what "open" or "widely available" means is whether the model weights are readily accessible by members of the public in editable (modifiable) form for use with the model code. Edit-enabled access to the weights facilitates the user's ability to make modifications to a model, whether the modification leads to a benefit or a risk.

Of the primary elements of an AI model (training data, code and model weights) the ability to obtain and modify the model weights is most important to assist a user in scrutinizing or making modifications to the model. The weights represent the acquired "knowledge" and behavior-setting of the model, and are the result of a very expensive and time-consuming investment in the training process. Changing the model weights allows for a broad, and at the same time very refinable, ability to impact the model's outputs. As long as the weights can be applied to the model code to generate output, having the ability to modify the weights allows for change in the output and behavior of the model without the need to modify the code or the original training data.

With use of the term "open" or "Open" throughout this document, we are referring to a model or AI system where the model weights are readily accessible by members of the public in editable (modifiable) form for use with the model code (whether or not the model code or training data are "open"). The term "readily accessible" as used in the prior sentence means that a person outside the organization that is developing, hosting or deploying the model can access and obtain the weights without the need for action by the publishing organization of the model specific to that individual and without willful circumvention of applicable technical constraints by the individual. In this document, we equate "widely available" with "open". Please also note that our use of the term "open" or "Open DUFM" does not mean to imply that the model is, or should be, considered pure open source under standard definitions of "open source software".

Question 1(b). *Is it possible to generally estimate the timeframe between the deployment of a closed model and the deployment of an open foundation model of similar performance on relevant tasks? How do you expect that timeframe to change? Based on what variables? How do you expect those variables to change in the coming months and years?*

Databricks believes that major open source model developers are not far behind the closed model developers in creating equally high performance models, and that the gap between the respective

development cycles may be closing.[3] There is no technical constraint preventing open foundation models from preceding closed models of similar performance. The competitive threat from open models is a primary reason that the large developers of closed models are so keen to see AI regulation implemented that will slow down open model development. Not only are the open models less expensive for users to obtain and deploy, they are modifiable by the user and more flexible to cover a range of usage needs. The growth in the use of open models will lead to more interest by investors in the open AI ecosystem, and increased funding for developers of open models will enable them to catch up and perhaps surpass closed models - assuming AI regulation does not slow them down. Allowing these trends to continue is important to enable more people to reap the benefits of AI, and to nurture innovation and competition, and avoid a concentration of power, in the AI sector.

Question 1(c). *Should "wide availability" of model weights be defined by level of distribution? If so, at what level of distribution (e.g., 10,000 entities; 1 million entities; open publication; etc.) should model weights be presumed to be "widely available"? If not, how should NTIA define "wide availability?"*

Please see our response above to the lead-in question of this Section 1.

Question 1(d). *Do certain forms of access to an open foundation model (web applications, Application Programming Interfaces (API), local hosting, edge deployment) provide more or less benefit or more or less risk than others? Are these risks dependent on other details of the system or application enabling access?*

Please see our response above to the lead-in question of this Section 1, which contemplates editable access to model weights as a requirement for a model to be considered "open." Databricks sees tremendous benefits from the use of open models tuned for special use cases by businesses and other organizations, and from open models being available for research and transparency, particularly for purposes of seeking AI safety and security. To achieve these benefits, the model weights need to be made fully available to the technical experts of organizations for modification and deployment and to researchers for experimentation and analysis. If access is restricted to occur only via an API, these benefits would not be available because the model weights could not be accessed, modified and tested.

Question 2. *How do the risks associated with making model weights widely available compare to the risks associated with non-public model weights?*

The most often cited marginal (incremental) risks of concern of Open DUFMs as compared to Closed DUFMs relate primarily to malicious use (intentional misuse or weaponization of the model), however it is important to note that virtually any technology can be weaponized or used maliciously in other ways, including Closed DUFMs and much simpler technologies as well. Examples with respect to Closed DUFMs include: the WormGPT, created by hackers to disable guardrails in GPT models to aid criminal activities[4]; a technique called AirPrompt developed to

[3] The DBRX open model introduced by Databricks on March 27, 2024, is more capable than GPT 3.5 based on widely accepted benchmarking, and challenges GPT 4 in several areas. See *Introducing DBRX, a New State-of-the-Art Open LLM*, Databricks blog posting, March 27, 2024, https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm.

[4] *How Criminals are Getting Help from AI Hacking Tools*, Financial Review, March 25, 2024, https://www.afr.com/technology/one-ai-hacking-tool-has-fallen-others-will-rise-to-take-its-place-20240305-p5fa1l

circumvent safety measures in closed models (and open models) to effectively trick the AI into responding to queries it was designed to reject, such as instructions on building bombs and making counterfeit money[5]; and a security flaw identified by researchers in closed vision language models allowing users to successfully make inquiries as to how to make a bomb[6].

Question 2(a). *What, if any, are the risks associated with widely available model weights? How do these risks change, if at all, when the training data or source code associated with fine-tuning, pretraining, or deploying a model is simultaneously widely available?*

The biggest risks Databricks sees are the risks that would be created by prohibiting the wide availability of model weights: i.e., the risks to economic productivity benefitting a larger swath of society, innovation, science, competition, and AI transparency if Open DUFMs were not widely available. In particular, the greater competition provided by Open DUFMs will drive down costs, allowing a greater number of people and communities to participate in the benefits of AI. Databricks sees the advantages of open source AI every day. Databricks has well over a thousand enterprise and government customers using modified open models to achieve a wide range of important benefits. Moreover, from this vantage point, it is clear to Databricks that open models are having a direct and significantly positive impact on our economy's overall productivity. Access to a model's weights is necessary for a researcher to explore how the model works and for a user to make modifications to the model. For organizations wanting to re-train a model for a specific organizational purpose, the ability to modify the model is required. Without the ability to modify a model using open model weights, businesses and other organizations will be reliant on expensive, less flexible and opaque closed models. If the highest performance models are required to be closed, the benefits of AI will be significantly curtailed while AI's risks may be increased, including harms from data leakage, inaccuracy, hallucination, environmental impact, etc.

A recent study by Stanford HAI on the risks and opportunities presented by open models concluded that there is limited evidence of meaningful marginal risk in areas that are often cited as risks (such as bio terrorism, and enhanced phishing and other cybersecurity threats) from open models as compared to closed models or other existing (non-AI) technologies[7]. With respect to concerns about non-consensual imagery or child sexual abuse material, equally offending models can be much smaller (in terms of compute training thresholds) than any proposed 'frontier model' thresholds. The study also concluded that similar safety mechanisms for closed models are also vulnerable. If the marginal risk of open models vs. closed models or existing technology is not significant, then any regulation restricting open models would have the effect of impairing the substantial benefits they offer while achieving no significant gain in terms of reduced risk.

With open models, a community of users and researchers work to identify and address issues and to understand how the software or application works. This benefits the entire user base, and outweighs the comparative risks associated with closed models regarding transparency and model explainability. Enabling a much wider audience to look inside the model to conduct

---

[5] *Researchers jailbreak AI chatbots with ASCII art -- ArtPrompt bypasses safety measures to unlock malicious queries,* March 7, 2024 https://www.tomshardware.com/tech-industry/artificial-intelligence/researchers-jailbreak-ai-chatbots-with-ascii-art-artprompt-bypasses-safety-measures-to-unlock-malicious-queries.

[6] *Scientists identify security flaw in AI query models*, January 10, 2024, https://techxplore.com/news/2024-01-scientists-flaw-ai-query.html#google_vignette.

[7] *On the Societal Impact of Open Foundation Models*, Stanford HAI, February 27, 2024, https://crfm.stanford.edu/open-fms/.

research helps advance the safety of both that model and future models, and helps advance the capabilities of future models. As an example, in its own model development, Databricks has significantly benefited from third party research enabled by the availability of the model weights and code of Llama 2 and other open models, assisting Databricks in developing greater safety and capability for its own models. The availability of such third party research means a greater diversity of input is incorporated into AI models, which is particularly important with respect to safety issues.

In addition, with closed models, an organization desiring to modify the model must share its fine-tuning data with the provider of the model. This exposed data may include valuable and sensitive confidential information, creating safety, security and privacy issues. With open models, the modifier can maintain control over its proprietary and sensitive data.

If all DUFMs are required to be closed, there will be a substantial risk of extreme market concentration in the hands of a few big players and a lack of competition, limiting the availability of the advantages of AI and increasing its costs. This market concentration would also decrease incentives for improvement, limiting the pace of innovation.

Question 2(b). *Could open foundation models reduce equity in rights and safety-impacting AI systems (e.g. healthcare, education, criminal justice, housing, online platforms, etc.)?*

Because of the heightened scrutiny from the open community with access to an open model, we believe bias and other flaws of the model will be identified and addressed with greater speed. As with open software generally, an open model will have an army of private investigators looking for issues and raising them quickly. Arguably this attribute of openness is more important than ever with highly performant AI given its power and growing involvement in so many aspects of our lives.

Question 2(e). *What, if any, risks could result from differences in access to widely available models across different jurisdictions?*

Policy relating to the availability of powerful open models should be aligned internationally. Provision of powerful AI is a global business with limited ability to prevent transfer of AI model functionality across borders. Stricter standards in one jurisdiction places a compliance burden on good actors while not effectively preventing bad acts by bad actors. Similar concerns apply within the U.S. A national approach at the federal level with explicit preemption is important to avoid a patchwork of state laws that would unnecessarily increase compliance burdens while not meaningfully restricting the conduct of bad actors.

Question 2(f). *Which are the most severe, and which are the most likely risks described in answering the questions above? How do these sets of risks relate to each other, if at all?*

Databricks believes the biggest risk considered in this discussion is the economic risk that would arise from giving a small number of closed model providers an effective oligopoly by banning or creating onerous regulatory hurdles to Open DUFMs. This risk of reduced competition is certain, and substantial in its potential impact. The concentration of power would not only increase costs, it would limit accessibility to powerful AI and its benefits, and it would slow the pace of innovation, as well as potentially facilitating the development of “walled gardens” / gatekeeper systems where the Closed DUFM owners stifle competition within the broader AI ecosystem. In terms of safety and security, we believe that the clear advantages of transparency and earlier

identification of risks and mitigation pathways offered by Open DUFMs largely offset the uncertain perceived safety risks of Open DUFMs, and that the likelihood of these perceived safety risks being incrementally addressed in a meaningful way by making model weights not widely available is very small.

Question 3. *What are the benefits of foundation models with model weights that are widely available as compared to fully closed models?*

Open models provide substantial benefits, including: (a) broad accessibility for research, transparency, explainability, and risk identification/mitigation that significantly contribute to improving safety and security - what amounts to "crowd sourced" quality control; (b) accelerated innovation through greater accessibility, broader input, collaborative improvements and a more rapid innovation cycle; (c) democratization of AI by virtue of broader availability, reduced costs, and allowing user modifications for special use cases and experimentation; (d) creating competition and avoiding the concentration of power in AI in the hands of a few big closed model providers; (e) allowing organizations to use models they own and control to fine tune them with their secured proprietary data without exposing it to a closed model provider over the web via an API or in other manners; (f) the potential ability to modify models in ways that could moderate environmental impact by reducing the amount of compute used for inference for special use cases; and (g) greater potential diversity in terms of demographic groups (ethnic, gender, geographic location, etc.) that will have the chance to contribute to (and benefit from) model innovation, scrutiny and risk identification. All of these benefits of open models are applicable at the frontier ("dual-use") level in addition to being applicable to less capable open models. With respect to (f) (modifications that could limit use of compute that may reduce environmental impact), a very large, capable model could be modified in ways that could reduce compute usage by bypassing or deleting certain elements of the full model, but such modification would generally only be practical if the modifier has access to the model weights.

Question 3(a). *What benefits do open model weights offer for competition and innovation, both in the AI marketplace and in other areas of the economy? In what ways can open dual-use foundation models enable or enhance scientific research, as well as education/training in computer science and related fields?*

As with open source software generally, open models foster collaboration and rapid experimentation by the open model community members. Collaborative research is easier, and innovative breakthroughs are achieved sooner. The avoidance of API access fees lowers costs, thereby increasing competition. Faster innovation, greater ability to customize for special use cases and lower costs will improve productivity and competitiveness in parts of the economy outside the AI sector. Databricks is seeing this improvement in productivity and innovation in working with thousands of customers on AI implementations in virtually every sector. As just a few examples, Databricks customers are using AI based on open models to: significantly reduce software product development timeframes; accelerate pharmaceutical R&D lifecycles; increase the efficiency of telecommunications, transportation and freight networks; positively impact climate change by increasing efficiency in the energy sector; enhance end-customer experiences in gathering information about products and services; and improve employee job quality by reducing repetitive, tedious work that can lead to high worker burnout and turnover. From a macroeconomic point of view, a recent study by McKinsey concluded that generative AI models

could generate $4 trillion or more in additional annual global growth[8]. The heightened capabilities of dual-use foundation models when they become available will cause these productivity and innovation benefits to grow to even greater levels.

Permitting a business or other organization to control their own models, including the model weights, lets them move their models from one vendor data platform to another, avoiding the problem of vendor lock-in and increasing competition within the AI sector. If the model weights are not available, the model cannot be moved to a new vendor. Even where a Closed DUFM provider facilitates some sort of modifiability of a model, the model is still controlled by that Closed DUFM provider and the customer does not have the ability to transport the model as modified to a new vendor platform of its choice. The more powerful AI becomes, the greater the dangers to society from having the power over it concentrated in a few hands.

Collaboration with the academic community on AI science is crucial, to enhance innovation and efficiency but also importantly to help identify and understand safety and security issues posed by AI and find ways to address AI risk. A recent research paper on AI model security vulnerabilities, shows how access to model weights enables scrutiny of model vulnerabilities and can lead to security improvements for both closed and open models[9]. Another recent research paper on AI model privacy vulnerabilities demonstrates the same advantage with respect to data privacy[10]. In a recent research paper from Stanford HAI, 25 leading AI experts across industry, academia, and civil society concluded that "model weights are essential for several forms of research across AI interpretability, security, and safety" and that "model weights enable external researchers, auditors, and journalists to investigate and scrutinize foundation models more deeply."[11]

If cutting edge AI models are required to be closed, the ability for academics to conduct such research will be severely limited and the world will have to rely largely on the commercial frontier model developers themselves to address safety and security issues posed by cutting edge models, subject to whatever incentives and disincentives may apply to them. Inevitably, commercial developers will be more likely to prioritize factors they deem important to commercial success, whereas academic researchers will have relatively more interest in focusing on a broader range of areas, including safety, security, and other risks. If we deem the risks of the most capable models to be particularly high, we should prioritize the openness of such models to better enable research by academics and others outside the realm of what will likely be a very small number of large developers of such models. As more than 1,800 people involved in AI research, analysis, policy and entrepreneurship have declared in a recent "Joint Statement on AI Safety and Openness", "If our objectives are safety, security and accountability, then openness and transparency are essential ingredients to get us there."[12]

---

[8] *The economic potential of generative AI: the next productivity frontier*, McKinsey Digital, June 14, 2023, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction.

[9] *MasterKey: Automated Jailbreak Across Multiple Large Language Model Chatbots,* October 25, 2023, https://arxiv.org/abs/2307.08715.

[10] *Alpaca against Vicuna: Using LLMs to Uncover Memorization of LLMs*, March 5, 2024, https://arxiv.org/abs/2403.04801.

[11] *On the Societal Impact of Open Foundation Models*, Stanford HAI, February 27, 2024, https://crfm.stanford.edu/open-fms/.

[12] *Joint Statement on AI Safety and Openness*, October 31, 2023, https://open.mozilla.org/letter/.

As one example of the value of open models weighed against their risks, in an important area of scientific research in the medical arena where AI risks are often highlighted, more than 150 leading scientists recently signed a statement affirming their view that "the benefits of current AI technologies for protein design far outweigh the potential for harm," noting that "many researchers in our community benefit from open-source scientific software, which has enabled rapid innovation and broad collaboration."[13]

Question 3(b). *How can making model weights widely available improve the safety, security, and trustworthiness of AI and the robustness of public preparedness against potential AI risks?*

Open access to model weights increases transparency, allowing a broader community of researchers, AI, cybersecurity and other experts, developers, and the public to understand how models make decisions and to participate in identifying and fixing security vulnerabilities. Without the model weights, and limited to any code or data that might be available, these parties will be less able to understand the model's functionality and to run experimentation. By making it easier for external parties to scrutinize AI models, the AI community can work collaboratively to address and mitigate potential risks, including biases, errors and vulnerabilities, leading to safer and more reliable AI systems.

Please also see our response to the prior question 3(a) regarding the importance of open collaboration with academic researchers looking into AI safety and security issues.

Question 3(c). *Could open model weights, and in particular the ability to retrain models, help advance equity in rights and safety-impacting AI systems (e.g. healthcare, education, criminal justice, housing, online platforms etc.)?*

The heightened scrutiny and modifiability of open models will lead to earlier detection and mitigation of bias and other flaws of the model. An open model will have an army of private researchers looking for issues and raising them quickly. This also adds the benefit of greater demographic and geographic diversity in terms of who is scrutinizing the model for bias and other issues. Arguably this attribute of openness is more important than ever with AI, given its power and growing involvement in so many aspects of our lives. The more powerful AI becomes, the more important this widespread and diverse scrutiny becomes.

Additionally, the ability to readily retrain an open model lets the deployer more effectively, efficiently and swiftly implement modifications and other measures that can address discovered bias or other flaws in the model.

Question 3(d). *How can the diffusion of AI models with widely available weights support the United States' national security interests? How could it interfere with, or further the enjoyment and protection of human rights within and outside of the United States?*

The availability of highly capable open, modifiable models permits U.S. government agencies to affordably own and control their deployments of such technology in a manner that lets the agencies tailor them to their needs and maintain security for classified information. High end AI technology is expected to be very helpful in analyzing national security information. If all high end models are closed, the U.S. government would incur much greater costs in being able to tap such technology in a manner that would maintain adequate security for classified information. Without

[13] *Community Values, Guiding Principles, and Commitments for the Responsible Development of AI for Protein Design*, March 8, 2024, https://responsiblebiodesign.ai/.

Open DUFMs, the U.S. government would either have to devote substantial resources to building its own such models, or it would have to incur substantial costs in procuring them (or obtaining sufficiently secured access) from one of the few large technology companies able to offer such models.

Question 4. *Are there other relevant components of open foundation models that, if simultaneously widely available, would change the risks or benefits presented by widely available model weights? If so, please list them and explain their impact.*

Making the training data widely available provides the benefits of facilitating analysis to identify potential bias, harmful content and other risks, while creating little if any incremental risk, so there should be no regulatory restrictions on the availability of training data.

Making the model code widely available in addition to the model weights provides the benefits of incremental transparency in evaluating the model, including to identify and mitigate potential risks, and it can facilitate the ability to modify the model for beneficial purposes. It can also facilitate research, innovation and the democratization of AI. Although providing access to the code if the model weights are already available does marginally increase risk by making it easier to use, modify and potentially misuse the model, availability of the model weights is the crucial part of the model required for its independent use and modification. The incremental risk of allowing access to the code when the weights are available is made even less important from a practical point of view due to the fact that the type of skilled operators able to use open model weights to modify a model are likely also skilled enough to reverse engineer the code with use of the model weights. On balance, the transparency and other benefits of allowing open code significantly outweigh the marginal risk prevention afforded by preventing access to the code.

Question 5(b). *Are there effective ways to create safeguards around foundation models, either to ensure that model weights do not become available, or to protect system integrity or human well-being (including privacy) and reduce security risks in those cases where weights are widely available?*

Safeguards can be, and are, built into both open and closed models. Employing a system that provides unified governance across the disparate elements related to the AI lifecycle is important in enabling robust, seamless access control as part of keeping model weights and other aspects of the AI system secure prior to intended release. As an example of such a system, Databricks offers Unity Catalog[14], an advanced unified governance, permissioning, versioning and lineage tracking system that integrates with the data and AI workflow. The typical non-unified approach creates potential gaps in governance coverage where governance is shifted from one governing element to another.

Although making model weights available can assist a bad actor in disabling the safeguards built into an open model, it is worth noting that the safeguards in closed models can also be "broken" by a determined and knowledgeable actor. However, with open models it is easier to quickly build a collective understanding of what might have gone wrong, and to make fundamental changes to address the problem.

Question5(c). *What are the prospects for developing effective safeguards in the future?*

Databricks believes measures to maintain safeguards as part of open models and/or to disable open models when attempts are made to use the models for malicious acts may be achievable in

[14] See https://docs.databricks.com/en/data-governance/unity-catalog/index.html.

the future, and that research in these areas should be encouraged. Databricks believes that encouraging this research is a better policy focus than restricting access to model weights in light of the substantial benefits open models provide to society.

Question 5(e). *What if any secure storage techniques or practices could be considered necessary to prevent unintentional distribution of model weights?*

Unified governance is important. Please see our response to question 5(b).

Question 5(f). *Which components of a foundation model need to be available, and to whom, in order to analyze, evaluate, certify, or red-team the model? To the extent possible, please identify specific evaluations or types of evaluations and the component(s) that need to be available for each.*

For research aimed at risk mitigation, it is beneficial to provide researchers access to the model code in addition to the model weights. Between these elements, the availability of model weights has the greatest impact on risk. While the additional availability of the code doesn't significantly increase risk, such availability would aid in research efforts to identify and address potential risks. To fully leverage the benefits of open models, both components should be accessible. Assuming model weights are provided, the incremental risk posed by making the code available is minimal. Accepting this minor risk is worthwhile considering the significant benefits achievable.

Question 6(a). *In which ways is open-source software policy analogous (or not) to the availability of model weights? Are there lessons we can learn from the history and ecosystem of open-source software, open data, and other "open" initiatives for open foundation models, particularly the availability of model weights?*

Much of the technology we benefit from every day is built upon open source software because the open source ecosystem has fostered significant innovation, system quality, and ease of adoption. Open models should be allowed so that this contribution to innovation, quality and growth can continue in the field of AI, as AI rapidly becomes an ever more important part of our economy.

One legal issue to address is the allocation of liability when use of an open model, including an Open DUFM, causes harm, in cases where the end user is not responsible. Generally, the deployer of the model or AI system, rather than the model developer, should be responsible because the deployer is selecting the use case, configuring the application, has better visibility into the actual and potential risks, and has responsibility for making sure the appropriate safeguards are in place. With a consumer-facing application, the deployer should bear the burden of responsibility for any heightened standards applicable to consumer-facing deployments, including addressing any risk of harmful content being generated by the underlying model. Exposing open model developers to liability for such downstream harms would constrain innovation in open model development and limit realization of many of the significant benefits of open models. The developer should have the ability to disclaim liability and seek indemnities as part of the license it provides for open use of the model. Without the ability to allocate such risks within the AI supply chain, open model development and the substantial benefits it brings would be impaired.

Question 6(b). *How, if at all, does the wide availability of model weights change the competition dynamics in the broader economy, specifically looking at industries such as but not limited to healthcare, marketing, and education?*

The availability of open models significantly enhances competition among AI providers, which benefits users in all sectors of the economy. If the most capable models were restricted to closed

models, the concentration of market power in the AI sector would be significant, which would be dangerous not only from a competition and innovation point of view, but from a societal perspective. The power of AI should not be held exclusively in the hands of a few giant technology companies. In addition, society as a whole benefits if costs are reduced and productivity is enhanced. The "democratization" of AI that open models provide will help extend the benefits of cost reduction and productivity to everyone. Given how important AI will become in workflows throughout our economy, these benefits will have substantial impact.

The availability of open models gives organizations that don't develop highly capable models themselves the ability to actually own and securely control the model they use in making their organization more productive, giving them more control, including over their sensitive data used in retraining and fine-tuning. Limiting the most powerful AI to closed models would put a few large model developers in ultimate control of the cutting edge of AI, with everyone else dependent on their judgment and oversight, while at the same time allowing that small handful of parties to exclusively extract the value from applying AI across a broad swath of industries. Ultimately we face a choice between centralized control over the cutting edge power of AI in the hands of a few big commercial players, versus decentralized control.

In a world where only closed highly capable models exist, an organization wishing to operate its own such model, with the ability to securely modify for its own use cases over time, would have to build it from scratch, which would likely be prohibitively expensive and time consuming, meaning it could not practically do so. The result would be to force the organization to pay expensive access fees to an oligopolist provider of highly capable models, and to forgo the ability to freely make modifications and realize the other benefits of being able to control the model it wants to deploy.

Question 6(c). *How, if at all, do intellectual property-related issues—such as the license terms under which foundation model weights are made publicly available—influence competition, benefits, and risks? Which licenses are most prominent in the context of making model weights widely available? What are the tradeoffs associated with each of these licenses?*

Open model licenses vary somewhat, with some (e.g., the Llama 2 Community License Agreement - the "L2CLA") restricting commercial use to a certain extent. The L2CLA prohibits use of Llama 2 to train other models, and requires certain extremely large companies (those with over 700 million monthly active subscribers) to obtain a separate license from Meta. The Databricks Open Model License ("DOML"), under which the open model DBRX was released on March 27, 2024, has similar provisions. Despite these restrictions, Llama 2 and DBRX in large part provide the key advantages of open source: they are generally available free of charge, to virtually anyone, including for most commercial purposes; the model weights and code are open for research, modification, and fine-tuning for specialized use cases; and their openness promotes innovation and democratization of AI. Although not considered by purists to be true "open source", these models generally provide the benefits of openness we have discussed in our responses. Unlike typical open source licenses, these licenses require the licensee to indemnify the developer (Meta or Databricks) for third party claims arising from downstream use under the licenses. The L2CLA and the DOML address certain concerns of developers while preserving much of the benefit of open source. We expect other open model developers will consider using licenses similar to the L2CLA and the DOML going forward.

Prohibitions on specified harmful uses contained in Responsible AI Licenses (RAIL) and Acceptable Use Policies (AUP) (like the ones provided by Meta and Databricks as part of the L2CLA and the DOML, respectively) can be helpful in lessening the chance of harmful use, although they do not provide technical barriers against determined bad actors.

Question 7. *What are current or potential voluntary, domestic regulatory, and international mechanisms to manage the risks and maximize the benefits of foundation models with widely available weights? What kind of entities should take a leadership role across which features of governance?*

Because of the vast advantages of open models, as discussed in response to prior questions, Databricks feels AI regulation and requested voluntary commitments should be very carefully tailored to not impede open model development. Regulation of highly capable foundation models should be directed to requirements applicable during the pre-release development stage, including documentation requirements as of release date. Databricks believes that regulatory scrutiny may be appropriate at the time of deployment of an AI system that constitutes or is built upon an Open DUFM, with appropriate additional regulatory requirements applied if the deployment represents a high risk use case, but the regulatory obligations applicable at the time of deployment should be the responsibility of the deployer rather than the developer since the deployer (which in some cases may also be the developer) has better control and visibility regarding the actual usage of the model or system. Deployment of a system or application incorporating a DUFM (whether closed or open) in a consumer-facing manner should be subject to requirements placed upon the deployer appropriate to consumer deployments, including the implementation of standard safeguards preventing generation of harmful content and any user-facing transparency requirements. In addition, consumer-facing applications incorporating Open DUFMs should not allow consumer users to access the model weights through the application. The deployer should be required to test the application sufficiently prior to release to confirm that any retraining or fine-tuning did not remove or deteriorate preexisting safeguards built into the model by the original developer.

Question 7(a). *What security, legal, or other measures can reasonably be employed to reliably prevent wide availability of access to a foundation model's weights, or limit their end use?*

Databricks believes regulation should allow open models at all capability levels because the substantial benefits of openness meaningfully outweigh the largely unproven marginal risks openness may present. However, Databricks realizes that at some point AI systems will arise that are so powerful that policy makers are likely to require that broad access to the inner workings of the AI systems (whether model weights or some other future crucial element) be gated in some manner. We think the day gating might be warranted is far in the future, and likely to involve AI architectures that are substantially different from architectures used in today's leading generative models, so crafting requirements currently would be premature and unlikely to address the challenges presented at that time. That said, if a gating requirement ever is considered, Databricks feels it should take the form of an "allow listing" process whereby those provided access to the gated AI system's crucial element(s) (e.g., model weights, if still the relevant factor) must verifiably register to gain access. This registration should apply to anyone with access to a qualifying model's weights, whether the model is open or closed. Registration should involve verification of identification (including a fingerprinting process similar to that conducted for employees of FINRA-registered broker dealers), collection of detailed contact information,

background checks (similar to the federal background checks to permit gun ownership), verification of residency, a check against official federal "deny lists" (e.g., sanctioned or otherwise prohibited countries, entities or individuals, etc.), and disclosure of intended use. The registrant would also be required to affirm their commitment to: (a) maintain any built in safeguards which were documented by the developer at time of release; (b) adhere to any restrictions on use contained in the applicable open model license, including any Acceptable Use Policies; (c) limit any redistribution giving access to the crucial element(s) to only parties who have verifiably registered pursuant to the same process; (d) limit any deployment to other internal users or customers solely to implementations that do not provide access to the crucial element(s) (e.g., model weights), and only if proper security safeguards are in place; and (e) not deploy in any specified high risk manner or consumer-facing manner unless any separate regulations relating to such deployments of highly capable models are fully adhered to. Both developers and distributors (including open source repositories) would be required to verify proper registration by all potential distributees prior to distribution of a qualifying model's weights to the potential distributee. Willful violations of the registration process would be subject to penalties severe enough to constitute a meaningful deterrent. An organization would need to register every employee having access to the crucial element(s). A developer would have an obligation to register at the point it reasonably becomes aware that a model in development meets whatever the model qualification standard is at the time.

This registration process would increase the costs and risks faced by potential bad actors, increasing the chances of interdiction, enforcement and penalty. It would therefore be expected to significantly reduce the probability of malicious use. It is true that a bad actor could successfully register and then simply download the model weights to a thumb drive and sell them to a black market buyer, but note that the same could occur with employees of a closed model developer. In either case, the registration and enforcement system would lessen the probabilities of the occurrence of malicious acts.

Question 7(c). *When, if ever, should entities deploying AI disclose to users or the general public that they are using open foundation models either with or without widely available weights?*

Based on the balance of burdens versus benefits, it would not be justified to require a special obligation to disclose to users that they are dealing with a model or AI application where the model weights are widely available, except potentially in the case of a consumer-facing application (where special requirements should apply to both Closed and Open DUFMs), and any such requirement should be placed on the deployer and not the developer. Any such requirement placed on the deployer should be carefully constructed to apply only where reasonably necessary.

Question 7(d). *What role, if any, should the U.S. government take in setting metrics for risk, creating standards for best practices, and/or supporting or restricting the availability of foundation model weights?*

It is important that any required standards and restrictions be national in scope, to avoid a patchwork of regulations at the state level. AI applications and services are typically cloud based, and are delivered on a widespread basis, often globally. The necessity of dealing with a patchwork of regulatory requirements would be an onerous burden for deployers. Databricks supports the principles and guidance set forth in the NIST AI Risk Management Framework as a basis for national AI policy and standard setting.

Question 7(e). *What should the role of model hosting services (e.g. HuggingFace, GitHub, etc.) be in making dual-use models with open weights more or less available? Should hosting services host models that do not meet certain safety standards? By whom should those standards be prescribed?*

If a registration system like the one referenced in our response to question 7(a) is ever contemplated, the model hosting services such as HuggingFace and GitHub would be required to limit distribution to verified registrants, with respect to those open models where registration is required.

Question 8(a). *How should these potentially competing interests of innovation, competition, and security be addressed or balanced?*

Open models offer substantial benefits relating to innovation, safety and security research, competition, lower costs, prevention of the concentration of the power of AI in the hands of a few, broadening the benefits of AI, and ease of customization for a large range of beneficial purposes. These significant benefits should be taken into consideration in weighing whether to take steps to restrict open models to achieve theoretical reductions in marginal risk. Ultimately the question is whether it is worth losing these many benefits, and leaving the power of AI in the hands of a few large technology companies, when the addressable marginal risk of open models is uncertain and can potentially be mitigated by other less disadvantageous means.

Question 8(b). *Noting that E.O. 14110 grants the Secretary of Commerce the capacity to adapt the threshold, is the amount of computational resources required to build a model, such as the cutoff of 1026 integer or floating-point operations used in the Executive Order, a useful metric for thresholds to mitigate risk in the long-term, particularly for risks associated with wide availability of model weights?*

When setting thresholds, it is crucial to ensure that open models fine-tuned or customized for the specific needs of an enterprise or government organization not become classified as high risk simply due to the additional computing resources used for such modifications. Any calculation of compute on a "cumulative" basis should explicitly not include compute used in modifying an open model. Customization for specialized use generally improves the accuracy of the model in a focused area and does not increase risk. Such customization is more likely to reduce risk, in particular the propensity to "hallucinate".

Question 9. *What other issues, topics, or adjacent technological advancements should we consider when analyzing risks and benefits of dual-use foundation models with widely available model weights?*

Research should be encouraged to explore means of reducing the likelihood of malicious use of open models. One such area that appears promising is 'task blocking' - a means by which an open model that has been modified can be blocked from being used in a harmful manner[15]. In evaluating the relative risks of open vs. closed models, it is also worth paying attention to research that has shown that closed models are not necessarily meaningfully less risky[16].

---

[15] *Self-Destructing Models: Increasing the Costs of Harmful Dual Uses of Foundation Models*, August 9, 2023, Stanford Research presented at AIES (AI, Ethics & Society) Conference on August 9, 2023, https://arxiv.org/pdf/2211.14946.pdf.

[16] *ChatGPT can be Jailbroken as Easily as Llama 2*, Stanford HAI, November 2, 2023, https://hai.stanford.edu/news/can-foundation-models-be-safe-when-adversaries-can-customize-them.

* * * * *

Thank you for the opportunity to provide comments on this important subject. Databricks looks forward to additional opportunities to discuss the important advantages of permitting the open availability of highly capable AI models.