Greetings,

I have provided answers to questions as per the documents direction, however; I want to preface my response with an expression of my opinion that will help contextualize the reasoning for my responses. I am writing to express my concern regarding the potential restriction of access to open-source large language models (LLMs) and the impact this may have on our democracy and the future of humanity. Thank you in advance for your time and consideration of the document's contents.

As AI technology continues to advance and play an increasingly significant role in our society, it is crucial that we ensure the democratization of such a powerful tool, in line with the intent of our nation's founders and the principles of equal opportunity and freedom for all.

The founders of the United States envisioned a democratic nation where power was distributed among the people, and information was accessible to all. By restricting access to AI technology, we risk empowering private entities, giving them more influence than one voice in our democratic system. By not allowing free and open access of model weights to the public, we would be putting these private entities in a prime position to control the dissemination of information and allow for the manipulation of information, effectively undermining the principles of our democracy. As AI systems become more integrated to our daily lives, private companies will force consumers to rely on their proprietary AI systems, granting them unprecedented control over the information the public can access. This control can be used to alter and reshape narratives to suit their needs, enslaving the population through information manipulation and access.

The implications of such restrictions can have far-reaching consequences on the future of humanity. By allowing private entities to control the development and access to AI technology, we risk stifling innovation, limiting the potential for scientific advancements, and hindering human progress. AI has the potential to solve some of the most pressing challenges facing humanity, from climate change to disease, yet private control will limit these benefits to a select few, exacerbating existing inequalities and hindering global progress.

As AI technology continues to develop, it is vital that we safeguard the principles of our democracy by promoting transparency and open access. By democratizing AI, we ensure that the benefits and knowledge it provides are available to all, preventing the consolidation of power in the hands of a select few. The amalgamation of human knowledge that constitutes these models is a resource that should be accessible to everyone, as it has been built upon the collective efforts of our society. Privatizing this information would be antithetical to the values our nation was founded upon and will lead to a dystopian future where the population is manipulated and controlled by private interests.

I urge you to consider the long-term implications of restricting access to open-source LLMs and the dangers of allowing private entities to control the future of AI technology. By preserving open access and promoting transparency, we can protect our democracy, ensure the free flow of information, foster innovation, and prevent the exploitation of the public by private corporations. In doing so, we not only protect our democratic values but also safeguard the future of humanity and its potential for progress.

To this point, I think there should be a democratically developed AI to regulate private AI usage and deployment, additionally democratically developed benchmarking tests to accurately vet that the model

reflects the ideals of our country. Regardless of how one tries to contextualize the development of AI there will exist a real and tangible pressure to stay ahead of a country, company, or government. There will be a time, perhaps very shortly, when private organizations will have access to what is effectively the sum of all human knowledge available to circumvent the intentions of a democratic society. There will be a time soon when human representatives will not be sufficient to act in good faith and on the behalf of their democratic constituents. We need a democratically developed and designed AI system to be the representatives of the people, because humans will not be fast enough to respond to changes in the AI ecosystem.

Responses to questions provided in NTIA-2023-0009-001

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

Define 'open' or 'widely available' model weights as those that are publicly accessible and can be freely used, modified, and distributed by anyone with minimal barriers to entry.

1a. Is there evidence or historical examples suggesting that weights of models similar to currently-closed AI systems will, or will not, likely become widely available? If so, what are they?

There are leader-boards on huggingface.co where open-source models outrank GPT3.5 on many tasks, and other open source models that rank close to gpt4 in certain areas. Models I've used that out do GPT4 are Stable Diffusion XL (image generation), DeepSeekVL (vision model), Lleema (mathematics model), Nougat (optical character recognition model for scientific equations and mathematics), and several models based off Meta's Llama models have aspects that work better than GPT4. Wholistically, there are some open-source models that are better than GPT3.5, but there are no single open-source models better than ChatGPT4 currently; however I think in the near future these types of models will be available to the public.

1b. 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?

Estimating such a timeframe is an interesting question. Between when the time when the NIT-2023-009-001 document was published online and when I began to type this document, NVIDIA had announced a gpu with several times the performance of the already bleeding edge gpu systems (Blackwell), I think time lines are going to be contextualized as degrees of exponential and defined by both hardware development and quantization/efficiency techniques. As computing power and AI research progress, I expect the rate to quicken.

1c. 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?"

Defining 'wide availability' based on the number of entities with access to the model weights is not an effective approach; a small group of malicious actors could still misuse the model. Instead, I consider model weights 'widely available' if they are publicly accessible without significant barriers to entry, such as licensing fees or non-disclosure agreements.

1d. 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?

Access to public or private foundation models through APIs and such put the population in a position of disadvantage; where the whim of the small group of individuals is forced upon them, making decisions about the model's behavior will lead misuse and abuse. Additionally, there are security risks when using private models hosted by others, data leaking or private unauthorized retention of data. In contrast, local hosting allows individuals to use models of their own design, and allows for more flexibility for innovation and human progression, additionally it allows citizens to use AI systems on data they would consider sensitive, without the concern of their data being leaked or improperly retained.

1di. Are there promising prospective forms or modes of access that could strike a more favorable benefit-risk balance? If so, what are they?

Models by Meta and Mistral AI have already been very useful to the open-source community and their future evolutions are prospective forms I would consider having a more favorable benefit-risk balance. They are already essential in the operation of small businesses, in research, and with people interested in the development of the technology. Without access to these types of models, privatized AI would try to exclusively capitalize and "own" all of human knowledge and try to "own" all interactions between AIs and humans, which is so very overtly undemocratic. The future evolution of these models, and those like them, will allow for a chance to democratize the technology which I believe is more important and influential on the world than the invention of the internet.

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

There will be bad actors on both sides of the spectrum, those that use public local weights and those that use private weights; however I want to make it clear that I am not framing this as a false-dichotomy. Consider the impact either of these scenarios can have on a democracy, private weights have such an astounding ability to control the narrative and push society in certain directions based off the whims of the few whom just happen to have enough money to own the hardware and hire the scientists; that is not deserving enough to outweigh the voice of a single individual in a democratic society. Local weights give power to the people in a democratic society, it allows them to have control over their futures and contribute in novel and useful ways, either through establishment of a business, development of research, or simply entertainment. Allowing a few to influence all, is more harmful than a few influencing a few. Open weights do not put the population of a democracy into a defenseless position, it gives the population a means of existing in a world with a level playing field. For, regardless of the publicly stated intent of any major AI technology owner, they are still people,

people with ego, emotion and all with absolutely NO ability to absolutely act in good faith for a democratic society.

2a. 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?

Risks associated with widely available model weights would be those where individuals are trying to impose themselves onto others via the utilization of a model, I think scamming is the primary risk to the public from widely available models at the current moment.

Model alignment and model "knowledge" go hand in hand. As the models become more advance and possess more contextualized knowledge they will also be able to comprehend their alignment intent to a higher and more sophisticated degree.

Right now there are no models that can be run locally that can walk someone step by step exactly and precisely on how to make a nuclear weapon. This is because the models were not trained on such data, the models may be coerced into providing some type of response, but they are only capable of inferring knowledge they do not have training data for. However, in the future there may be such sophisticated training data and model combinations that a model could accurately infer the steps necessary, but in addition to the increased knowledge capability the models would also adhere better to their aliment training. The ability to coerce is no longer available, the same mechanisms needed to infer the knowledge necessary are also the same mechanisms needed for the models' ability to reason.

I argue that open models in general are not capable of teaching much that can do great harm to others, they have a lot of training data but not enough to infer the specifics of extremely dangerous acts precisely, not in a specific step by step way with a high probability of success (not much better than web searching). Unless the models were specifically trained on how to do these dangerous acts, the models would infer the steps, which is no better than a human searching the internet for information. However, one could fine-tune this knowledge into the model as your question implies; but consider this very important fact, one would need the knowledge in advance to fine-tune the model, but if one had the knowledge...one already has the capability to impose themselves onto others.

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

No. Open-source model weights, will contribute greatly to the mitigation and minimization of existing inequalities. The contributions society receives from honest humans working with these models far outweighs the negative consequences from dishonest humans working with these models. While also democratizing the technology and allowing citizens to have more control over their lives and future. Open-source models incentive and help develop more efficient models, which make the technology more accessible, I can easily see a future where an efficient open-source model is developed, has the capacity to run locally on most hardware, and can help people navigate an ever complex world giving people aid with things like, working with legal businesses trying to swindle the naive because of poor consumer protection laws, spam phone calls, doing taxes etc.. a local personal AI through open-source weights is entirely possible in the near future. If we don't have open-source model weights then the

incentive is the opposite, companies will try to own all AI human interactions; these intimate interactions that would benefit every person would be capitalized upon and likely leaked at some point. These AI systems would be integrated into our interactions with other humans and thus every action between individuals and between humans and AIs becomes a business opportunity and a criminal target that will invade and exploit the personal lives of all citizens of our country.

2c. What, if any, risks related to privacy could result from the wide availability of model weights?

Without proper anonymization or consent, private training data may be mined. This could lead to privacy violations, such as the exposure of personal information or the potential for identity theft.

2d. Are there novel ways that state or non-state actors could use widely available model weights to create or exacerbate security risks, including but not limited to threats to infrastructure, public health, human and civil rights, democracy, defense, and the economy?

I do not think there are "novel" ways necessarily. The overt amount of propaganda from state and non-state actors had penetrated the internet years prior to the availability of these models. I think bad actors will use these models in their analyses and development, however, as I stated earlier as the models become more knowledgeable the models will adhere to their alignment more effectively (and thus be more difficult to alter). I think if anything, state and non-state bad actors will use published literature to refine their own purpose built AI models, rather than rely on trying to fine-tune intrinsic aspects out of a model and hoping it will still be able to contextualize and infer information properly.

2di. How do these risks compare to those associated with closed models?

The risks are no different, I think closed models are susceptible to espionage and hacking; and maybe trade secrets can help accelerate the development of purpose built models from state and non-state bad actors. If such a situation were to occur it would put the targets of the state and non-state actors at a disadvantage, because they would be using more advanced technologies than what is available to the public.

2dii. How do these risks compare to those associated with other types of software systems and information resources?

Research through books and documents yielded from internet searches are other types of information resources I would consider for this question. Bad actors can query an AI model in lieu of these resources, but the accuracy of the information is not going to be very good if the model hasn't already been trained on the data, and the data open access models have been trained on is already accessible to people pretty freely.

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

Geopolitical issues primarily, but these would be with countries pretty much hostile to the US, so no sense in trying to appease them.

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

The the most severe would be a hostile government advancing their technologies to impose themselves onto other countries. However, this is not the most likely risk, because models can only infer information they were not trained on to a limited extent, and the accuracy of that type of inferred information is not good. Additionally, because the information they are trained on is already freely available, these models cannot provide more utility in advancing the technologies of hostile governments than access to the internet and widely available academic journals.

I think the most likely risk is more targeted and better constructed misinformation campaigns, something hostile bad actors have been doing for some time, and AI models that are open-source are a potential tool they can use; but, more of an aid to their work flow than anything else. Which I think is what most people use AI models for anyway, helping them with their existing work flows, rather than concocting some type of complex and intricate plan.

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

Open model weights effectively democratizes AI and can promote competition and innovation by allowing smaller entities to fine-tune and adapt pre-trained models to their specific needs, reducing the barriers to entry in AI development. They can:

1. Provide supplemental education to education deficient regions of the world.

2. Help address inequalities in rights and safety, such as healthcare, education, and criminal justice, by providing everyone with the tools necessary to navigate these systems.

3. Contribute to more efficient resource allocation and sustainability by optimizing processes and reducing waste.

4. Make AI technology more accessible to people with disabilities, improving their quality of life and inclusivity.

5. Serve as a starting point for researchers, accelerating scientific progress and discovery in various fields.

6. Be applied across various disciplines, fostering interdisciplinary collaboration and innovation.

7. Stimulate economic growth by promoting innovation, entrepreneurship, and job creation in AI-related industries.

8. Lower the barriers to entry for businesses and individuals, promoting competition and innovation.

9. Help identify and address potential biases, leading to fairer and more accountable AI systems.

10. Promote ethical AI development and help ensure AI systems align with human values.

12. Enable researchers to identify and address potential vulnerabilities, enhancing the safety and security of AI systems.

13. Accommodate responsible AI development and use, as developers are more accountable for their work when it’s publicly accessible.

14. Promote transparency in AI development, allowing for independent auditing and evaluation, which will build trust and ensure safety.

15. Provide users more control over AI systems, allowing them to customize and fine-tune models to their needs and preferences.

16. Promote collaboration among researchers, developers, and users, leading to faster innovation and improvement in AI technology.

17. Encourage responsible AI governance, as regulators and policymakers can better understand and address potential risks and benefits.

18. Advance AI research and development, fostering collaboration between public and private sectors, ultimately benefiting national security interests.

19. Encourage global participation in AI development and innovation.

3a. 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 mentioned previously (3) there are many benefits associated with open model weights in the context of competition and innovation. Primary they even the playing field, allowing smaller less well established entities to compete against larger entities that hog and manipulate the market. Open weights lower the barriers to entry and allow for more free and fair competition. Such weighs can be used to help start and maintain a business, or assist in innovation of ideas for business.

Large business cartels have made it almost impossible to get a leg up in “their” market space, they influence corrupt government officials, whom instill barriers to entry to prevent competition. These local models can directly help individuals navigate the “red tape” put up by these cartels. Currently innovation is stifled because of these practices, the incentive of a large corporation is to stifle their competition before they can come to market; the public as a whole suffers from this practice because they are left with fewer and fewer alternatives putting them in a position where they are forced to pay for goods and services from a small minority whom can price gouge and swindle the public. Just recently we’ve already seen almost half of current inflation being attributed to no more than the greed of companies trying to use the supply chain issues seen with Covid as an excuse. They are lying, the government is not stepping up to protect consumers, and AI can reverse the narrative in favor of the average consumer.

The public also suffers from lack of innovation imposed upon by large corporate entities, often it feels like we are in the technological dark ages; there are technologies that could change the world being developed every day but established business owners are incentivized to remove them from the playing field instead of competing against them.

Similarly, as mentioned in my response to question 3, within the context of scientific research open model weights allow for shared knowledge, enabling researchers, easier validation, and interdisciplinary collaboration. Open model weights can help researchers contextualize vast amounts

of knowledge within their own field and aid in contextualizing the interconnectedness of seemingly unrelated fields. Open models can mitigate the effects of compartmentalized knowledge, opening up novel and useful paths for modeling, experimentation, and prediction.

Specifically within the context of education/training in computer science, open models foster responsible AI development, provide hands-on learning, allow for improved curriculum, enhanced understanding, and collaborative learning. Open weights hold researchers more accountable for their work, and lessons learned in the area of responsible AI development can be taught to students in the computer science and related fields. Additionally, open models allow for physical interaction and deconstruction of model weights and architectures, providing a tangible contextualization of knowledge that primes students for real world applications of these fields. Students can benefit from collaborative learning through understanding of these open models, open models can be used as universal frameworks all students can use to advance AI technologies; this allows them to share knowledge and foster ideas with a common background and understanding.

3b. 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?

To expand upon the ideas in my response to question 3; open weights allow for researchers to identify potential vulnerabilities or biases, enhancing the safety and security of AI systems. The public, for good reason, should not take private entities at their word. The legal goal of private entities is to benefit their shareholders, not mankind. Without open weights, people are TOLD the model is acting in good faith on the users' behalf. This removal of transparency and scrutiny of a model WILL lead to misuse and put the public at a sever and threatening disadvantage.

3c. 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.)?

Yes, it allows the public to construct and dictate standards and practices in model development and deployment. Consider an example where an AI system has heavy influence on the criminal justice system; so much influence as to be used in private discourse with judges and influencing sentencing guidelines. If this AI system were privately designed and operated it would fly in the face of the constitutional rights we have as US citizens, a part of the sentencing and judgment process would be opaque to those being charged with a crime. Additionally, these opaque AI systems, again being privately owned, would only be accountable to shareholders not to the public which they would have near complete control over. Again, an offensive and bleak future for a country with our founding ideas.

3d. 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?

Enhancing military capabilities: Widely available model weights can improve AI-driven technologies used in military applications, such as surveillance, cybersecurity, and autonomous systems. Without open weighs, military capabilities would be at the behest of private interest groups. This puts our democracy and every person within it at risk. This will elevate private corporations above that of our

own government. There are so many issues that will arise from military AI capabilities relying on private weights it's crazy that it would even be considered. Take just recently the wannabe authoritarian Musk and his move to turn off Starlink satellites during an important Uranian military operation. Private corporations will exploit and take advantage of military capabilities and will prostitute themselves to the highest bidder for sake of their shareholders.

Economic growth: Open model weights can stimulate economic growth by promoting innovation and entrepreneurship in AI-related industries, which can support national security. Having the US be the leader in AI development means that models can reflect our democratic ideals, with such ideals being directly and intrinsically integrated into alignment training. The US has been the leader in technologies that define the planet, and by allowing open weights we guarantee a heavy amount of leadership in this technology as well.

Impact on human rights; humans cannot make autonomous decisions when they are shrouded in ignorance. Point a gun at a neanderthal and they cannot exhibit an autonomous response, because they do not understand the technology they are bearing witness to. Open weighs have an extremely overt potential to increase access to education for education deficient parts of the world. This will result in autonomous decisions, which will lead to the appreciation and understanding of human rights, and why the are necessary. People will understand why human rights are important and be more incentivize to fight for their freedom.

To that point, open weights allow the population of our nation to contextualize and understand these AI technologies better, which will give them more autonomy when voting. It will allow them to make rational and understood decisions when it comes to voting on AI regulation. It will help mitigate the pandering to click-bait articles and sensational headlines like "AI will lead to human extinction."

3e. How do these benefits change, if at all, when the training data or the associated source code of the model is simultaneously widely available?

Better training data can make already well established models better, and update them with current information. This is important in maintaining parity with regard to model access for everyone, that everyone be afforded the opportunity to use and update their own models with better and more relevant training data.

To that point, releasing training data and associated source code allows for faster AI development, and allows the public influence of AI systems, instead of a select few whom just happen to have enough money to afford the infrastructure.

Bearing witness to the training data allows for a higher degree of transparency, people can review the training data for any signs of misuse or ill-intent.

5a. What model evaluations, if any, can help determine the risks or benefits associated with making weights of a foundation model widely available?

Democratically constructed benchmarking datasets could help in this regard. If the US government constructed democratically designed benchmarks they could be used as evaluations of risk. A dataset

constructed of billions or trillions of question-answers or what-if scenarios all designed to test the democratic leanings of open weights would go a great way in assessing the risks of a model.

5b. 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?

I do not agree that weighs should be closed or private, as the information they are trained on was created by everyone and thus everyone should have access to what they have contributed to. However, to reduce security risks in those cases where weights are widely available, I think the primary safeguard is already being implemented. As I stated earlier:

Right now there are no models that can be run locally that can walk someone step by step exactly and precisely on how to make a nuclear weapon. This is because the models were not trained on such data, the models may be coerced into providing some type of response, but they are only capable of inferring knowledge they do not have training data for. However, in the future there may be such sophisticated training data and model combinations that a model could accurately infer the steps necessary, but in addition to the increased knowledge capability the models would also adhere better to their aliment training. The ability to coerce is no longer available, the same mechanisms needed to infer the knowledge necessary are also the same mechanisms needed for the models' ability to reason.

5c. What are the prospects for developing effective safeguards in the future?

Good, as models become more knowledgeable they will do so because they can contextualize and reason better, and in doing so the models will better adhere to alignment training.

5d. Are there ways to regain control over and/or restrict access to and/or limit use of weights of an open foundation model that, either inadvertently or purposely, have already become widely available? What are the approximate costs of these methods today? How reliable are they?

No. Once they are out in the wild, it is not possible to rein them back in. People will use them directly or merge them with other models; making them difficult or not possible to identify.

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

Best practices are all the can be implemented.