Name: Alex Punnen
email: alexcpn@gmail.com
Country: India

Question 1. and subparts:
How should NTIA define ‘‘open’’ or ‘‘widely available’’ when thinking about foundation models and model weights? Should ‘‘wide availability’’ of model weights be defined by level of distribution? If so, at what distribution level (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?’’

Comment:
Assuming the intention of not open-sourcing is to reduce risks, making weights available to download by anyone in public should term the access as essentially open-sourced and widely available. So the question of 1 or 1 million does not have much meaning But to answer, if the model weight is downloadable by anyone in public then it should be deemed as “open” or “widely available”

Note: A better way to frame this question would be whether it should be kept as a national secret (like nuclear plans) or not. And in that context define “open” or “closed”. For example, if GPT.X is an AGI system should those weights be classified as “open” or “closed”

Question 1: Subparts (a,b)
a. 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?
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?

Comment:
From recent history, it is clear that just after ChatGPT there were similar models trained and with training notes and weights released publicly.GPT NeoX, the BLOOM project and the Meta LLAMA project. The BLOOM project was from Europen collaboration.

If we backtrack to the fundamentals we can answer the above questions by answering two questions.

1. Is the technology behind the current LLMs accessible/open?
2. Is the resource needed for implementing the technology accessible?

The answer to the Question1 subparts then becomes self-evident and can be framed

1. Yes the technology related to Transformers and their implementations and optimisations are publicly accessible.
2 . Some of the resources are easily accessible, some are not
    1. The data needed for training is widely available as it is the current internet data and has contributions from multiple nations
    2. The hardware and software are not widely available
    3. The engineering skills needed are reasonably available in many countries and something all nations can build up if lacking

The major bottleneck in creating an Open model weight is the access to computational resources, But since information cannot be contained forever in a social system, it is matter of time before other companies and nations start making similar hardware and software, Maybe 20 years.

Question 1: Subparts (d)
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?

Comments
Access via API exposes only what the API exposes and it can be heavily controlled. Hence it is much less useful in extending the model weights but also much less risky as well

2. How do the risks associated with making model weights widely available compare to the risks associated with non-public model weights? 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? b. Could open foundation models reduce equity in rights and safety impacting AI systems (e.g., healthcare, education, criminal justice, housing, online platforms, etc.)? c. What, if any, risks related to privacy could result from the wide availability of model weights? d. 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, defence, and the economy? i. How do these risks compare to those associated with closed models? ii. How do these risks compare to those associated with other software systems and information resources? e. What, if any, risks could result from differences in access to widely available models across different jurisdictions? 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?

The Risks associated with making model weights widely available are unknown, but as long as the system capability is as is, the most would be to misuse this as a tool for misinformation and

cyber bots. Since misinformation can be used to fan the inherent hate among groups this is a serious threat to understand and mitigate.

The Risks will naturally increase with sharing the fine-tuning, pretraining and other scripts and source code and other information

Privacy concerns are not major as the system is already trained on data which if it is already compromised is implicitly part of the weights itself and nothing more is added by opening the weights

The Risk associated with non-availability is the risk that innovations and applications based on the fundamental models will be throttled. So much so that when these are used for good, things good for humanity example cure for cancer or fact-checking bots on social media that can auto-flag content etc. The good usually outweighs the bad from historical data. Even nuclear power has in net improved humanity and new research here of smaller and safer fusion reactors could be the future of clean energy or at least have some part there.

3. What are the benefits of foundation models with model weights that are widely available as compared to fully closed models? 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? 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? 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.)? 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? e. How do these benefits change, if at all, when the training data or the associated source code of the model is simultaneously widely available?

Comments are similar to question 2.

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.

Comments:
The main components are the architecture, the data and the resources. If compute resources are widely available it would change the risk equation or everyone will be able to replicate the weights.

5. What are the safety-related or broader technical issues involved in managing risks and amplifying benefits of dual-use foundation models with widely available model weights? a. What model evaluations, if any, can help determine the risks or benefits associated with making weights of a foundation model widely available? 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? c. What are the prospects for developing effective safeguards in the future? d. 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? e. What if any secure storage techniques or practices could be considered necessary to prevent unintentional distribution of model weights? f. Which components of a foundation model need to be available, and to whom, 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. g. Are there means by which to test or verify model weights? What methodology or methodologies exist to audit model weights and/or foundation models?

Comment: In case weights are widely available, I dont think any practical frameworks can prevent anyone from misusing those. All measurement matrics of LLMs are subjective compared to other simpler LLMs. This is because very large LLMs /weights behave us a weak AGI mode as of now and are generalist which makes it difficult to measure them, just like it is difficult to correctly access the other generalist species that is us.

6. What are the legal or business issues or effects related to open foundation models? 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? 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? 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? d. Are there concerns about potential barriers to interoperability stemming from different incompatible ‘‘open’’ licenses, e.g., licenses with conflicting requirements, applied to AI components? Would standardizing license terms specifically for foundation model weights be beneficial? Are there particular examples in existence that could be useful?

Comments: No comments

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? 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? b. How might the wide availability of open foundation model weights facilitate, or else frustrate, government action in AI regulation? 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? 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? i. Should other government or nongovernment bodies, currently existing or not, support the government in this role? Should this vary by sector? 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? f. Should there be different standards for government as opposed to private industry when it comes to sharing model weights of open foundation models or contracting with companies who use them? g. What should the U.S. prioritize in working with other countries on this topic, and which countries are most important to work with? h. What insights from other countries or other societal systems are most useful to consider? i. Are there effective mechanisms or procedures that can be used by the government or companies to make decisions regarding an appropriate degree of availability of model weights in a dual-use foundation model or the dual-use foundation model ecosystem? Are there methods for making effective decisions about open AI deployment that balance both benefits and risks? This may include responsible capability scaling policies, preparedness frameworks, et cetera. j. Are there particular individuals/ entities who should or should not have access to open-weight foundation models? If so, why and under what circumstances?

Comments: Should governments interfere or not? Can it interfere effectively? Has it in any way influenced the development of this technology? Since it has not, it is my view that it cannot interfere effectively and hence it should not interfere at all. Current laws are good enough to protect people, businesses etc from the risks.

8. In the face of continually changing technology, and given unforeseen risks and benefits, how can governments, companies, and individuals make decisions or plans today about open foundation models that will be useful in the future? a. How should these potentially competing interests of innovation, competition, and security be addressed or balanced? 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? c. Are there more robust risk metrics for foundation models with widely available weights that will

stand the test of time? Should we look at models that fall outside of the dual-use foundation model definition?

Comment: If all GPU chips henceforth have reduced resources or floating point precision, it will severely limit this technology and training in those countries where this is applied. Since other countries will soon have comparable GPUs it will just handicap the US. The question is if we should be afraid of very powerful AGI systems. I guess we should be as by definition higher intelligence will be hard to contain by lower intelligence. Should governments prevent AGI from fearing this? I believe it should not as this may turn out to be the thing that may even finally sustain our race, though it can be equally dangerous. But as a race we have survived very dangerous ideas before and confront such daily.I guess we will with this too.

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

Comments: None