< encode justice >

Department of Commerce
Undersecretary of Commerce for Security and Industry

Bureau of Industry and Security
14th St NW & Constitution Ave. NW
Washington, DC 20230

Comment on Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters

Encode Justice, the world's first and largest youth movement for safe, equitable AI writes to express our support for the Bureau of Industry and Security's (BIS) proposed reporting requirements for the Development of Advanced Artificial Intelligence Models and Clusters. The proposed rule would create a clear structure and method of implementation for sections 4.2(a)(i) and 4.2(a)(ii) under Executive Order 14110.[1] In light of the massive potential benefits and risks of dual-use foundation models for American national security, it is critical that our security apparatus has a clear window into the activities of the companies developing these systems.[2]

Transparency for national security

There is no doubt that today we are leading the race to develop Artificial Intelligence. Overly burdensome regulations could stifle domestic innovation and potentially undermine national security efforts. We support the Bureau of Industry and Security's proposed rules as a narrow, non-burdensome method of increasing developer-to-government transparency without covering small entities. This transparency is key to ensuring that models released to the public are safe, the military and government agencies can confidently adopt AI technologies, and that dual-use foundation model developers are responsibly protecting their technologies from theft or tampering by foreign actors.

The military or government falling behind on the adoption of AI technologies would not only hurt government efficiency domestically but harm our ability to compete on the world stage. Any measures that can facilitate the confident military and government adoption of AI should be treated as critical to our national security and global competitiveness. Integrating these technologies is only possible when we can be

[1] U.S. Executive Order 14110. "Further Providing for the National Emergency with Respect to the COVID-19 Pandemic." 2020. Federal Register.
[2] Ryan Heath, "U.S. Tries to Cement Global AI Lead With a 'Private Sector First' Strategy," Axios, July 9, 2024, https://www.axios.com/2024/07/09/us-ai-global-leader-private-sector.

confident that the frontier of this technology is safe and reliable. Reliability and safety are *critical,* not counter, to maintaining our international competitiveness.

A nimble approach

As we have long stated, government reactions to AI must be nimble. This technology moves rapidly, and proposed rules should be similarly capable of swift adaptation. Because BIS maintains the ability to change the questions asked in surveys and modify the technical conditions for covered models, these standards will not become obsolete within two or three generations of model development.

We believe the timing of reports could also be improved. Generally, a quarterly survey should be adequate but there are circumstances in which BIS authority to request reporting out of schedule may be necessary. Recent reporting indicates that one of the largest frontier model developers provided its safety team just 9 days to test a new dual-use foundation model before being released.[3] After additional review post-launch, the safety team re-evaluated the model as unsafe.[4] Employee accounts differ as to the reason. There is currently no formal mechanism for monitoring such critical phases of the development process. Under the current reporting schedule, BIS may have gone as long as two and a half months before learning of such an incident. For true transparency, BIS should retain the ability to request information from covered developers outside of the typical schedule under defined certain circumstances. These circumstances should include a two-week period before or after a new large training run and a two-week period leading up to the public release of a new model.

Clarifying thresholds for models trained on biological synthesis data

One area for improvement is the definition of thresholds for models trained on biological synthesis data. While we support a separate threshold for such models, the current definition of "primarily trained on biological synthesis data" is ambiguous and could lead to inconsistencies. If read as being a simple majority of the total training data, there are models that should be covered that would not be. You may, for example, have a model where the training data is 60% biological synthesis data and another where it is only 40%. In this scenario, if the second model is trained on twice as much total data as the first model, the total amount of biological synthesis data the model is trained on may be higher than the first while evading the threshold as currently defined.

As an alternative, we would suggest either setting a clear percentage threshold on the ratio of data for a model to be considered “primarily” trained on biological synthesis data, or setting a hard threshold on the total quantity of biological synthesis data trained

[3] "OpenAI's Profit-Seeking Move Sparks Debate in AI Industry." The Wall Street Journal, October 5, 2023. https://www.wsj.com/tech/ai/open-ai-division-for-profit-da26c24b.
[4] *Ibid.*

on instead of a ratio. Both methods are imperfect. Setting the definition as a ratio of training data means that some models trained on a higher total quantity but a lower overall percentage of biological synthesis data may be left uncovered, while smaller models trained on less total data but a higher overall percentage may be unduly burdened. Shifting to a hard threshold on the total quantity of biological synthesis data would leave the threshold highly susceptible to advances in model architecture, but may provide more overall consistency. Regardless of the exact method chosen, this is an area in the rules that should be clarified before moving forward.

Regular threshold reevaluation

More broadly, BIS should take seriously its responsibility to regularly reevaluate the current thresholds. As new evaluation methods are established and standards agreed upon, more accurate ways of determining the level of risk from various models will emerge. Firm compute thresholds are likely the best proxy for risk currently available but should be moved away from or modified as soon as possible. Models narrowly trained on biological synthesis data well below the proposed thresholds, for example, could pose an equal or greater risk than a dual-use foundation model meeting the currently set threshold.[5] Five years from now, the performance of today's most advanced models could very well be emulated in models with a fraction of the total floating point operations.[6] Revised rules should include a set cadence for the regular revision of thresholds. With the current pace of advancements, a baseline of twice-yearly revisions should be adequate to maintain flexibility without adding unnecessary administrative burden. In the future, it may be necessary to increase the regularity of revisions if rapid advancements in model architecture cause high fluctuations in the computational cost of training advanced models.

Conclusion

The proposed rulemaking for the establishment of reporting requirements for the development of advanced AI models and computing clusters is a flexible, nimble method to increase developer-to-government transparency. This transparency will bolster public safety and trust, ensure the government and military can confidently adopt this technology, and verify the security of dual-use frontier model developers. In an ever-changing field like AI, BIS should maintain the ability to change the information requested from developers and the thresholds for coverage. The revised rules should include a clarified definition of "primarily trained on biological synthesis data" and the

---

[5] James Vincent, "AI Suggested 40,000 New Possible Chemical Weapons in Just Six Hours," The Verge, March 17, 2022, https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx

[6] Cottier, B., Rahman, R., Fattorini, L., Maslej, N., & Owen, D. (2024). The rising costs of training frontier AI models [Preprint]. arXiv. https://arxiv.org/pdf/2405.21015.

flexibility to request information from developers outside of the normal quarterly schedule under certain circumstances.

Encode Justice strongly supports BIS's proposed rule and believes that, with the suggested adjustments, it will significantly enhance both American national security and public safety.