RESPONSE TO NTIA RFC ON DUAL USE FOUNDATION ARTIFICIAL INTELLIGENCE MODELS WITH WIDELY AVAILABLE MODEL WEIGHTS

Submitted by the Johns Hopkins Center for Health Security

Executive Summary

Thank you for the opportunity to provide comments in response to the National Telecommunications and Information Administration (NTIA) Request for Comment (RFC) on "Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights,"[1] related to NTIA's responsibilities under section 4.6 of the Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (EO).[2] The comments expressed herein reflect the thoughts of the Johns Hopkins Center for Health Security and do not necessarily reflect the views of Johns Hopkins University. Below, we provide information regarding biosecurity considerations for topics related to policy and regulatory approaches to "open" dual-use foundation models (ie, those for which the model weights are widely available).

The Johns Hopkins Center for Health Security conducts research on how new policy approaches, scientific advances, and technological innovations can strengthen health security and save lives. The Center has 25 years of experience in biosecurity and is dedicated to ensuring a future in which pandemics, disasters, and biological weapons can no longer threaten our world. Our Center is composed of researchers and experts in science, medicine, public health, law, social sciences, economics, national security, and emerging technology.

Section 4.6 of the EO tasked NTIA with preparing a report concerning the benefits and risks associated with dual-use foundation models with widely available model weights.[3] The EO expressed particular interest in the risks associated with users fine-tuning open dual-use foundation models or removing model safeguards.

The EO defines a dual-use foundation model as, among other things, any AI model that contains at least tens of billions of parameters, is "applicable across a wide range of contexts," and "exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety."[4] The first concerning capability highlighted by the EO is the ability to "substantially lower[] the

[1] *Dual Use Foundation Artificial Intelligence Models With Widely Available Model Weights*, 89 Fed. Reg. 14,059 (Feb. 26, 2024).
[2] *See* Executive Order No. 14,110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (Oct. 30, 2023) [hereinafter *EO*].
[3] *EO* § 4.6. Although model openness exists on a spectrum, for the sake of simplicity we refer to all models with widely available weights as "open." *See* Sayesh Kapoor et al., *On the Societal Impact of Open Foundation Models* (working paper, 2024), https://crfm.stanford.edu/open-fms/paper.pdf.
[4] *EO* § 1(k).

barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons."

We consider highly capable large-language models (LLMs) and broad biological-design tools (BDTs) as potentially covered by this definition. Although BDTs are more narrowly targeted than LLMs such as GPT-4 or Llama 2, some are capable of a broad range of biology-related tasks or can be adapted to perform such tasks. For example, the recently released Evo model can purportedly "generalize across the three fundamental modalities of the central dogma of molecular biology" to design novel DNA, RNA, and proteins.[5] Although Evo is a 7-billion parameter model, and so below the EO size threshold for a dual-use foundation model, current trends—including, in recent years, an exponential increase in compute used to train BDTs and rapid growth in biological sequence data that models can be trained on—indicate that BDTs will continue to rapidly scale up in size and capability.[6] For these reasons, our recommendations below apply to both frontier LLMs and broad BDTs.

We recommend that NTIA consider the following points in drafting its report:

(1) Open dual-use foundation models' current biological capabilities are a poor proxy for model capabilities in the near- and medium-term future.
- Increased model scaling and the rapid generation of usable data mean that LLMs and BDTs will likely grow substantially larger and more capable in the coming months and years. The United States should plan for that future rather than wait for it to arrive.

(2) The United States should set open dual-use foundation model policies that mitigate high-consequence biosecurity risks, which we judge to be the potential for a dual-use foundation model to do the following:
- Substantially accelerate or simplify the reintroduction of particularly dangerous extinct viruses, or dangerous viruses that only exist now within research labs, that have the capacity to start pandemics; or
- Substantially enable, accelerate, or simplify the creation of new or enhanced biological constructs that could start pandemics.

(3) Narrowly targeted export controls may mitigate high-consequence biosecurity risks stemming from open dual-use foundation models.
- The Department of Commerce (DOC) has significant export control authority to restrict the transmission of software and data that pose biosecurity risks.
- The federal government should consider whether narrowly targeted export control restrictions on dual-use foundation models with concerning biosecurity capabilities may be justified in the future.

[5] Eric Nguyen et al., *Sequence modeling and design from molecular to genome scale with Evo* (working paper, 2024), https://www.biorxiv.org/content/10.1101/2024.02.27.582234v2.full.pdf.
[6] *See id*.; Nicole Maug et al., *Biological Sequence Models in the Context of the AI Directives*, Epoch (2024), https://epochai.org/blog/biological-sequence-models-in-the-context-of-the-ai-directives.

Considerations and Recommendations

Below, we discuss the above recommendations in more detail, including briefly surveying current and anticipated biological capabilities of leading AI models and the risks inherent in releasing open dual-use foundation models.[7]

Open dual-use foundation models' current biological capabilities are a poor proxy for future capabilities

Foundation models have enormous potential to address major challenges in medicine, public health, and the environment, and they offer other important benefits. However, AI capabilities that can improve health may also be used to cause harm. In analyzing the benefits and risks of open dual-use foundation models, NTIA should consider the serious biosecurity risks that emerging open dual-use foundation models may pose in the coming years over and above existing technologies such as the internet and preexisting biology modelling software.

Our concerns about such risks have been bolstered by two recent trends in AI development.

First, closed LLMs, such as GPT-4, have shown rapid progress in bioweapons-relevant tasks, including assisting with biological and chemical research design and testing.[8] Although public information related to the capabilities of LLMs suggests that the current generation of LLMs do not substantially assist in bioweapons planning today, their rapidly improving capacities are a cause for concern in the future.[9] And while open LLMs such as Gemma, Llama 2, and Mistral have lagged behind the technological frontier, their capabilities have advanced rapidly in recent months.[10] Meta, which has generally released open models, has announced plans to invest billions of dollars in creating models "that are at the state of the art and eventually the leading models in the industry."[11]

---

[7] Portions of our response draw in part on material provided in our recent response to NIST's Request for Information. *See* Center for Health Security, *Response to RFI Related to NIST's Assignments Under Sections 4.1, 4.5 and 11 of the Executive Order Concerning Artificial Intelligence* (Feb. 2, 2024), https://www.regulations.gov/comment/NIST-2023-0009-0138.

[8] *See* Daniil A. Boiko et al., *Autonomous chemical research with large language models*, 624 Nature 570 (2023); Brendt A. Koscher, *Autonomous, multiproperty-driven molecular discovery: From predictions to measurements and back*, 382 Science E1 (2023); Andres M Bran et al., *ChemCrow: Augmenting large-language models with chemistry tools*, (working paper, 2023), https://arxiv.org/abs/2304.05376.

[9] *See* Tejal Patwardhan et al., *Building an early warning system for LLM-aided biological threat creation*, OpenAI (2024), https://openai.com/research/building-an-early-warning-system-for-llm-aided-biological-threat-creation; Christopher A. Mouton et al., *The Operational Risks of AI in Large-Scale Biological Attacks: Results of a Red-Team Study*, RAND (2024), https://www.rand.org/pubs/research_reports/RRA2977-2.html.

[10] *See Mixtral of Experts: A High Quality Sparse Mixture-of-Experts*, Mistral (Dec. 11, 2023), https://mistral.ai/news/mixtral-of-experts/; Jeanine Banks & Tris Warkentin, *Gemma: Introducing new state-of-the-art open models*, Google (Feb. 21, 2024), https://blog.google/technology/developers/gemma-open-models/

[11] Alex Health, *Mark Zuckerberg's new goal is creating artificial general intelligence*, Verge (Jan. 18, 2024), https://www.theverge.com/2024/1/18/24042354/mark-zuckerberg-meta-agi-reorg-interview. *See also* Nathan Lambert, *Model commoditization and product moats*, Interconnects (March 13, 2024), https://www.interconnects.ai/p/gpt4-commoditization-and-moats ("There are countless individuals who can easily pay the price it takes to create a model like Claude 3 and release it to the world.").

Second, AI systems specifically focused on biological data and outputs—BDTs—have seen a similar rate of progress and model size expansion.[12] Many cutting-edge BDTs, such as Evo, RFdiffusion, and RoseTTAFold, are fully open. Moreover, advances in LLMs and BDTs are complementary. LLMs can now assist users in accessing and using BDTs to perform complex scientific tasks, such as designing proteins to bind to the SARS-CoV-2 spike protein.[13] Together, these advances are likely to lower the cost and decrease the skill required for researchers to use increasingly complex and powerful biological AI tools. Therefore, when assessing risks, dual-use foundation models should be understood to exist within the broader threat environment and not assessed singularly or within a vacuum.

Open dual-use foundation models create special risks and benefits. Such models could benefit safety by allowing open access for independent experts to test model characteristics and risks and understand their inner workings,[14] though such independent testing also could be done within closed systems that provide access to safety experts seeking to test the models. Openness, though, also poses serious risks. Researchers have shown that third parties can, at modest expense, strip out open dual-use foundation model safeguards and/or train open dual-use foundation models to create new (and potentially dangerous) capabilities. For example, scholars have trained Mistral 7B on the entirety of open-access content in the PubMed database to create "BioMistral," a model they report provides "superior performance compared to existing open-source medical models and [a] competitive edge against proprietary counterparts."[15] Researchers at MIT, meanwhile, fine-tuned Llama-2-70B—at the

[12] *See* Maug et al., *supra* note 6; Cassidy Nelson and Sophie Rose, *Examining Risks at the Intersection of AI and Bio*, Ctr. Long-Term Resilience (2023), https://www.longtermresilience.org/post/report-launch-examining-risks-at-the-intersection-of-ai-and-bio; Sarah R. Carter et al., *The Convergence of Artificial Intelligence and the Life Sciences*, NTI (2023), https://www.nti.org/analysis/articles/the-convergence-of-artificial-intelligence-and-the-life-sciences/; Jacob T. Rapp et al., *Self-driving Laboratories to Autonomously Navigate the Protein Fitness Landscape*, 1 Nature Chem. Engineering 97 (2024). *See also, eg,* Wei Feng et al., *Generation of 3D Molecules in Pockets via a Language Model*, 6 Nature Machine Intelligence 62 (2024); Google DeepMind Alpha Fold Team & Isomorphic Labs, *Performance and Structural Coverage of the Latest, In-development AlphaFold Model*, Alphabet (2023), https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/a-glimpse-of-the-next-generation-of-alphafold/alphafold_latest_oct2023.pdf; Minkyung Baek et al., *Accurate Prediction of Nucleic Acid and Protein-Nucleic Acid Complexes Using RoseTTAFoldNA* (working paper, 2022), https://www.biorxiv.org/content/10.1101/2022.09.09.507333v1; Joseph L. Watson et al., *De Novo Design of Protein Structure and Function with RFdiffusion*, 620 Nature 1089 (2023); Jiankun Lyu et al., *AlphaFold2 Structures Template Ligand Discovery* (working paper, 2023), https://www.biorxiv.org/content/10.1101/2023.12.20.572662v1.

[13] *See, eg, The Impact of Large Language Models on Scientific Discovery: A Preliminary Study using GPT-4,* Microsoft Research (working paper, 2023), https://arxiv.org/pdf/2311.07361.pdf.

[14] *See, eg,* Shayne Longpre et al., *A Safe Harbor for AI Evaluation and Red Teaming* (working paper, 2024), https://bpb-us-e1.wpmucdn.com/sites.mit.edu/dist/6/336/files/2024/03/Safe-Harbor-0e192065dccf6d83.pdf; Beren Millidge, *Open Source AI Has Been Vital for Alignment*, Beren's Blog (Nov. 5, 2023), https://www.beren.io/2023-11-05-Open-source-AI-has-been-vital-for-alignment/.

[15] Emmanuel Morin et al., *BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains* (working paper, 2024), https://arxiv.org/abs/2402.10373.

cost of only $200 in compute—to remove safeguards against providing virology-related answers in response to prompts that explicitly informed the model that the user was planning to release a bioweapon.[16] Finally, we note that the creators of Evo, a reportedly highly capable BDT, excluded viruses that infect eukaryotes from Evo's training set for safety purposes.[17] Because the model's weights are freely available, however, we are aware of no technical hurdle preventing a third party from doing that training themselves at a fraction of the cost it took to create the original Evo model (assuming data availability). Indeed, less than a month after Evo was released, it had already been fine-tuned on a dataset of adeno-associated virus capsids, ie, protein shells used by a class of viruses that infect humans.[18] As this case suggests, when a model's weights are publicly available, a developer's decision not to endow the model with dangerous capabilities is far from final.[19]

As Sayesh Kapoor and colleagues caution, it is important to consider the marginal risk that open models pose above preexisting technologies.[20] As of mid-2023, several small studies indicate that users with access to leading LLMs, even in one case a model with safeguards removed, were not statistically significantly better at planning biological weapons attacks than those with access to search engines alone.[21] And creating a competent plan of attack is quite different from having the skills or resources to carry it out.[22]

These caveats may provide cold comfort in the time ahead. First, dual-use foundation model capabilities are rapidly improving. It is impossible to predict with certainty how substantially LLMs will eventually improve over search-enabled bioweapons planning. But the fact that experts with GPT-4 access had improved accuracy scores on all five metrics of bioweapons planning surveyed by OpenAI (albeit, not statistically significantly) suggests that future dual-use foundation models may provide marginal benefits over preexisting resources.[23]

---

[16] Anjali Gopal et al., *Will Releasing the Weights of Future Large Language Models Grant Widespread Access to Pandemic Agents?* (working paper, 2023), https://arxiv.org/abs/2310.18233.
[17] *See* Nguyen et al., *supra* note *5*.
[18] Kenny Workman, *Engineering AAVs with Evo and AlphaFold*, LatchBio (March 20, 2024), https://blog.latch.bio/p/engineering-aavs-with-evo-and-alphafold.
[19] *See also generally* Tom Davidson et al., *AI capabilities can be significantly improved without expensive retraining* (working paper, 2023), https://arxiv.org/pdf/2312.07413.pdf.
[20] *See* Kapoor et al., *supra* note 3.
[21] *See* Patwardhan et al. *supra* note 9; Mouton et al., *supra* note 9.
[22] For a discussion of the tacit knowledge requirements for creating biological weapons, *see* Sonia Ben Ouagrham-Gormley, Barriers to Bioweapon: The Challenges of Expertise and Organization for Weapons Development (2014) and Kathleen M. Vogel, Phantom Menace or Looming Danger?: A New Framework for Assessing Bioweapons Threats (2012).
[23] *See* Gary Marcus, *When Looked at Carefully, OpenAI's New Study on GPT-4 and Bioweapons is Deeply Worrisome*, Marcus on AI (Feb. 4, 2024), https://garymarcus.substack.com/p/when-looked-at-carefully-openais; Anjana Ahuja, *AI's Bioterrorism Potential Should Not Be Ruled Out*, Fin. Times (Feb. 9, 2024), https://www.ft.com/content/e2a28b73-9831-4e7e-be7c-a599d2498f24; Matthew E. Walsh, *How to Better Research the Possible Threats Posed by AI-driven Misuse of Biology*, Bulletin of the Atomic Scientists (Mar. 18, 2024), https://thebulletin.org/2024/03/how-to-better-research-the-possible-threats-posed-by-ai-driven-misuse-of-biology.

Second, none of the small studies in the field so far have evaluated how much dual-use foundation models purposefully trained on relevant data (eg, virology literature) will marginally improve bioweapons development or assessed the interaction between LLMs and BDTs.[24] Nor, to our knowledge, have there been any published evaluations of the marginal benefit BDTs like Evo or RFdiffusion could play in bioweapons design.

Third, tacit knowledge and resource barriers are likely falling even as AI capabilities are increasing. A growing proportion of wet-lab work can be conducted by machines, including machines that researchers can pay to access remotely on a part-time basis.[25] Dual-use foundation models, even those untrained for this purpose, have also shown facility at directing research robots to perform laboratory tasks.[26] Taken together, these facts suggest that informational capabilities may play an increasingly large role in enabling high-consequence biosecurity threats in the coming years.

More empirical research is certainly called for. But given the risks involved, and the direction of dual-use foundation model capabilities, the US government should plan for a future in which there is a reasonable probability that open dual-use foundation models could provide meaningful assistance to those seeking to design and deploy biological weapons.

The United States should set open dual-use foundation model policies that mitigate the highest-consequence biosecurity risks

Dual-use foundation models can excel at many tasks, and therefore create many forms of risk. These actual and potential risks range from assisting fraudulent behavior and inadvertently cementing bias to enabling mass-casualty attacks. All these dangers are worthy of serious attention. But given the limited time the federal government has to develop its initial approach to such risks—in light of fast-approach EO deadlines and the rapid advances in AI model capabilities—we believe US agencies should, at a minimum, set open dual-use foundation model policies that address the most catastrophic risks, such as foundation models substantially enabling the creation of pandemic-capable pathogens. In its report, NTIA should

[24] Gopal and colleagues studied a model that was altered to be more helpful in planning a bioweapons attack but did not formally evaluate its efficacy or compare its assistance to access to the internet alone.
[25] *See* Rapp et al., *supra* note 11 (reporting on "the Self-driving Autonomous Machines for Protein Landscape Exploration (SAMPLE) platform for fully autonomous protein engineering. SAMPLE is driven by an intelligent agent that learns protein sequence–function relationships, designs new proteins and sends designs to a fully automated robotic system that experimentally tests the designed proteins and provides feedback to improve the agent's understanding of the system."); Tianhao Yu et al., *In Vitro Continuous Protein Evolution Empowered by Machine Learning and Automation*, 14 Cell Sys. 633 (2023); Filippa Lentoz & Cédric Invernizzi, *Laboratories in the Cloud*, Bulletin of the Atomic Scientists (July 2, 2019), https://www.ft.com/content/e2a28b73-9831-4e7e-be7c-a599d2498f24; Tessa Alexanian, *Develop A Screening Framework Guidance For AI-Enabled Automated Labs*, Fed. Amer. Scientists (Dec. 12, 2023), https://fas.org/publication/bio-x-ai-policy-recommendations/.
[26] *See* Microsoft Research, *supra* note 12.

underscore the importance of prioritizing the mitigation of high-consequence biosecurity threats among other open dual-use foundation model risks.

As a group of civil society organizations and academics recently wrote to the Secretary of Commerce, model openness provides significant benefit to society.[27] This fact underscores the need for the US government to narrowly tailor rules and regulations on open dual-use foundation models to address the highest-consequence and best-supported safety concerns.

We are particularly concerned that future dual-use foundation models may make it easier for scientists, and perhaps even those outside the scientific community, to create, cultivate, modify, and disseminate new or existing pandemic-capable pathogens. We are also concerned that dual-use foundation models may lower bioweapon program costs for nation states or other high-capability actors or enable such entities to develop pathogens with greater transmissibility or virulence than would be possible using traditional approaches to synthetic biology. As discussed above, these dangers are exacerbated by the existence of open, highly capable models that malicious actors (or benign but insufficiently cautious actors) could modify to improve dual-use biological capabilities.

As we discuss at greater length in our recent response to the National Institute of Standards and Technology (NIST) request for information regarding its obligations under the EO,[28] we believe the US government should prioritize developing policies that will mitigate the following high-consequence biological risks:

1. **An AI system that substantially accelerates or simplifies the reintroduction of extinct viruses with pandemic potential or viruses with pandemic potential that only exist now within research labs or virus repositories.**
2. **An AI system that substantially enables, accelerates, or simplifies the creation of new or enhanced biological constructs that could start pandemics.**

At a minimum, this means the US government should develop evaluations that assess whether dual-use foundation models increase: (1) the possibility that users can synthesize pandemic-capable pathogens that are either extinct or are limited to being in a lab or repository; or (2) the capability of a user to create a novel variant of a pathogen that has the potential to initiate a pandemic.

The federal government should also consider policies that mitigate high-consequence biosecurity risks specific to open dual-use foundation models. In particular, the government should consider narrowly tailored limitations on dual-use foundation models that can substantially assist in enabling high-consequence biological attacks, since there may be settings in which it is appropriate for users to interact with dual-use foundation models with

[27] Accountable Tech et al., Letter to Gina Raimondo, Sec. Dept. of Commerce, March 25, 2024, https://cdt.org/wp-content/uploads/2024/03/Civil-Society-Letter-on-Openness-for-NTIA-Process-March-25-2024.pdf.
[28] *See* Center for Health Security, *supra* note 6.

dangerous capabilities. For example, vaccine developers and cell biologists may need to have access to a range of advanced BDTs, potentially including those with dual-use capabilities, to the extent that the public health benefits of access exceed the risks. But the fact that some access to dual-use foundation models is justified does not itself justify unlimited access to those models. The government should therefore consider mandatory limitations in cases in which the risks of open access to a dual-use foundation model exceed its benefits.

Narrowly targeted export controls may mitigate high-consequence biosecurity risks stemming from open dual-use foundation models

NTIA should consider whether narrow, targeted export controls can serve as a useful regulatory tool to mitigate high-consequence biosecurity risks associated with open dual-use foundation models. The United States has used its broad export control authorities for more than 70 years to reduce access to biological weapons worldwide.[29] The US also has for decades participated in an arrangement known as the Australia Group to maintain multilateral controls on advanced technology, including software, that could be used to develop biological weapons.[30] Congress has recently reinforced this mandate, directing the DOC in 2018 to regulate the export of physical goods, software, and technical data, as well as the actions of US persons, in order to limit access to biological weapons anywhere in the world.[31]

Export controls, despite their name, do not only regulate physical goods shipped abroad. They can also be used to control dual-use technical information shared in the United States. In rare and controversial instances, the government has used export controls to prevent the publication of software or computer files it deemed threatening to national security.[32] The DOC conceivably could use its authority to restrict parties from making the weights of dangerous advanced dual-use foundation models freely available for download.[33]

---

[29] *See Proclamation 3038, Enumeration of Arms, Ammunition, and Implements of War*, 18 Fed. Reg. 7505, 7505 (Nov. 25, 1953).
[30] *See About Us – History*, Australia Group (2023), https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/origins.html. *See also Control List of Dual-use Biological Equipment and Related Technology and Software*, Australia Group (2022), https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/dual_biological.html. Notably, the Australia Group's software controls do not apply to software "in the public domain."
[31] *See* 50 USC. §§ 4811(2)(A)(i); 4812(a)(2)(C); 4813(d). For regulatory instantiations, *see, eg,* 15 C.F.R. Part 774, Supp. No. 1 (Commerce Control List), at ECCNs 1C351, 1C352, 1C353, 1E001, 2B352.e, j; 15 C.F.R. § 744.6(b). For a longer discussion, *see* Doni Bloomfield, *Export Controls and Artificial Intelligence Biosecurity Risks* (working paper, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4741033. *See also* Doron Hindin et al., *The Role of Export Controls in Regulating Dual Use Research of Concern: Striking a Balance between Freedom of Fundamental Research and National Security*, National Academies (2017), https://nap.nationalacademies.org/resource/24761/Strosnider-Hindin-Trooboff_Paper_012017.pdf.
[32] *See* Bloomfield, *supra* note 31; Steven Levy, Crypto (2001); *Defense Distributed v. Dep't of State*, 838 F.3d 451 (5th Cir. 2016).
[33] *See* Bloomfield, *supra* note 31.

NTIA should consider whether and under what circumstances the US—and potentially other members of the Australia Group or even a broader set of nations—should update export control rules to reduce high-consequence biosecurity risks associated with open dual-use foundation models. In doing so, NTIA should address what capabilities would justify export controls on model weights or actions, how controls might be narrowly tailored to apply only to the most concerning set of dual-use foundation models, and whether alternative models posing analogous risks or dangers would be readily available from sources the US government (or other Australia Group members) cannot control.

In a comment to NTIA, a group of civil society organizations and academics, including the Federation of American Scientists and the Electronic Frontier Foundation, have cautioned the DOC against the application of broad export controls to "general purpose" models.[34] We agree that open models can confer significant social benefits, including to public health, and that prior attempts to control open software provide a cautionary tale about the legal and practical challenges of applying export controls in this domain.[35] As those commenters acknowledge, however, "there are some situations where openness may exacerbate risks from AI." Given the comparative ease with which users can modify open models to remove safeguards or confer additional capabilities, we believe that openness may in some circumstances exacerbate the biosecurity risks associated with highly biologically capable models.[36]

For these reasons, any export controls on open dual-use foundation model weights should be narrowly tailored to address high-consequence threats to safety, such as the high-consequence biosecurity risks we outline above. In ongoing work, the Johns Hopkins Center for Health Security is studying specific model capabilities that could increase high-consequence biosecurity risk on the margin. We look forward to sharing that research with NTIA and other US government agencies when appropriate.

Given the uncertain nature of current and future open model capabilities, and the importance of open software, we are not suggesting that the DOC should impose export controls on dual-use foundation models today. Rather, the risks posed by open, biologically capable dual-use foundation models are grave enough for the US government to prepare such policy options so they can be deployed when and if they become relevant.

[34] *See* Accountable Tech et al., *supra* note 27.
[35] *See* Craig Jarvis, Crypto Wars: The Fight for Privacy in the Digital Age (2021); *Bernstein v. United States,* 176 F.3d 1132 (9th Cir.), *reh'g en banc granted and opinion withdrawn*, 192 F.3d 1308 (9th Cir. 1999); *Junger v. Daley*, 209 F.3d 481 (6th Cir. 2000).
[36] *See supra* notes 15—26 and associated text for a discussion of potential risks.