Comments of Matt Mittelsteadt in Response to Request for Public Comment on the Bureau of Industry and Security Framework for Artificial Intelligence Diffusion

I appreciate the opportunity to submit public comment on the Bureau of Industry and Security (BIS)'s Framework for Artificial Intelligence Diffusion. These comments do not represent the views of any party or special interest group but are intended to assist BIS as the agency formulates artificial intelligence (AI) export control policy.

The United States is the undisputed leader in AI technology. Yet, the recent "DeepSeek moment" demonstrates American AI leadership is far from assured. Key to enduring American AI success will be AI model diffusion.

Model diffusion is not merely an outcome but can itself be treated as a "resource" critical to fueling American AI success. *Exclusively* through global AI model diffusion can innovators collect R&D-critical, globe-scale usage data. Any limit represents a limit on training data. *Exclusively* through global diffusion can innovators ensure globe-scale revenue streams. Over the next four years, investors have pledged an unparalleled $1 trillion in new AI investments, roughly the size of 3% of the U.S. economy. Any barriers to model diffusion could cap the revenue streams needed to recoup these investments and ensure stable future R&D.[1] The strategic and economic importance of global AI diffusion cannot be understated.

Unfortunately, the Framework for Artificial Intelligence Diffusion's new global license requirement for the export of advanced AI model weights is a direct threat to this necessary national "resource." By requiring a license, BIS will create an AI economy centered on regulation as the default - and diffusion critical openness as the exception.

[1] Matt Mittelsteadt, "Comment on the Development of an Artificial Intelligence (AI) Action Plan," *Cato Institute*, March 17, 2025.

I urge BIS to be deeply skeptical of this level of towering regulatory ambition and control. Export controls are among the executive's most potent tools. Used in excess, they can act to restrain not our rivals, but American innovation.

In service to the administration's overall goal of American leadership in AI, I wish to highlight two key points:

1. Introducing broad based model export controls will yield significant regulatory uncertainty that could chill cutting edge innovation. Crossing this regulatory line is ill-advised.
2. The heavy deployment requirements of these rules will *significantly* burden the diffusion of advanced AI systems across all nations. This includes the eighteen allies currently subject to license exceptions.

Introducing broad based model export controls will yield significant regulatory uncertainty that could chill cutting edge innovation.

No matter its regulatory extent, the mere introduction of a model weight licensing regime will create immediate and sustained uncertainty. BIS must be clear eyed that this new licensing regime is not a final state. It is a starting gun. The ever-ballooning advanced computing export controls illustrates that once in effect, these rules will be disposed to significant revision and growth. As loopholes and blind spots are discovered, rules will need to morph to accommodate. Likewise, as new capabilities emerge, BIS will be pressured to shift regulations in tandem. Fundamentally, for these rules to achieve their stated goal of AI containment, constant, just-in-time changes are not optional, but required.

Once this regulatory starting gun fires, this process of change will begin, and business uncertainty will surge. A system that enables impromptu rules changes means entire product lines and investments can evaporate overnight. Even the best laid plans will bear heavy regulatory risk. For model makers, the recent export ban of the once regulatorily compliant Nvidia H20, and the company's resulting $5.5 billion in losses, will loom as a cautionary tale.[2]

[2] Anton Shilov, "Nvidia Writes Off $5.5 Billion in GPUs As U.S. Gov't Chokes Off Supply Of H20s To China," *Tom's Hardware*, April 16, 2025.

Even if the letter of the law is followed precisely, unfettered statutory power allows BIS to shift the rules and enforcement in an instant, slamming innovators with hefty costs.

Now that controls extend to model weights, BIS's mercurial regulatory approach may cause once free-wheeling developers to grow conservative. The present culture of risky frontier-pushing models, fast releases, and week-to-week iteration may fade as companies seek to avoid any release that tests BIS's comfort level.

Unfortunately, the text of the Framework makes clear these regulatory risks are not just the domain of large closed-source frontier labs. Per the regulation:

> "the U.S. Government will continue to actively monitor risks that could arise from open-weight models and assess what actions might need to be taken if heightened risks emerge."

This ominous assertion will rightfully be treated as a signal that exceptions for small and open models are subject to rescission. Amid such uncertainty, this air of conservative innovation may extend to small and under-resourced teams. In some cases, developers may choose to bow out of innovation today, if market success demands a robust legal department tomorrow.

By creating a model weight license, BIS crosses a critical regulatory Rubicon. No matter how modest, the *mere introduction* of such controls bears an uncertainty cost that will hamper both commerce and innovation.

The heavy requirements of these rules will *significantly* burden the deployment of advanced AI systems across all nations.

Under the Framework, BIS lays out substantial new processes, requirements, and restrictions on the diffusion of advanced AI model weights, defined as closed weight models trained on more than 10^{26} computational operations or significantly fine-tuned open-source models. Under the Framework, a license exception enables entities in the United States to export models to a select list of eighteen nations- close allies such as the United Kingdom, France, and Japan - hereafter referred to as "Tier 1" nations. Meanwhile, exports to users in nations not subject to U.S. arms embargoes - hereafter referred to as "Tier 2" nations –will require a license with a presumption of denial. Tier 1 and U.S. headquartered entities, however, will be relatively free to deploy

models in both Tier 1 and tier 2 nations subject to rigorous security standards. Finally, entities in the roughly 150 Tier 2 nations need a presumption of denial license to develop frontier models – effectively banning frontier development across most of the world.

While no model surpasses BIS's compute thresholds, such models can be anticipated soon. According to EpochAI, models that trigger these export controls may come as early as late 2025 based on industry trends and open-source intelligence.[3] Given the vastness of these new rules, the full effects, once triggered, cannot be predicted. Even still, I wish to highlight two significant burdens these rules will likely have on global model diffusion:

Deployment will be limited.

While the Framework preserves the freedom for Tier 1 providers to freely *transfer* models to their own data centers in Tier 1 and 2 nations, the regulations still dramatically limit the freedom to *deploy* these technologies.

Under the regulations, deployers in non-U.S. nations will be required to implement a truly significant range of security controls. To host a model, end users must use dedicated servers, bolster data centers with physical security standards drawn from the Department of Defense SCIF standards, deploy a range of strict cybersecurity controls, and install a "24/7/365 roving guard patrol or Perimeter Intrusion Detection System (PIDS) with a 15-minute response time." On top of these substantial constraints, host data centers must also submit to annual compliance audits by select "Third-Party Assessment Organizations" accredited by the FedRAMP Program Management Office.

This security will not be cheap. The Government Accountability Office (GAO) notes that cost estimate data for the security requirements are incomplete, but suggests a possible upper cost bound in the millions.[4] As for the annual audits, the GAO reports third party assessment can cost as much as $367,000 per year. Naturally, this assumes such assessors are internationally available, a deeply questionable assumption for a U.S. Government specific accreditation. It is indeed possible assessor supply will fail to meet these new international-scale assessment

[3] Luke Frymire, "Frontier Open Models May Surpass 10^{26} FLOP of Training Compute Before 2026," *Epoch.AI*, January 15, 2025.
[4] U.S. Government Accountability Office, "Federal Authorization Program Usage Increasing, but Challenges Need to Be Fully Addressed," *GAO-24-106591*, January 2024. https://www.gao.gov/assets/gao-24-106591.pdf

demands nurturing cost increases, and perhaps service delays, for not only the global AI industry but domestic firms drawing from this common pool of labor.

These costly requirements will both restrict and slow diffusion. By demanding specific data center designs and dedicated servers, the regulations will cap the number of globally deployed AI instances. For firms looking to overcome this manufactured supply constraint, the added costs of security-fitting may limit the number of new service centers. Meanwhile, the hundreds of thousands spent on annual audits will function as an effective "diffusion tax," raising the cost-of-doing-business while favoring large incumbents. Finally, diffusion restrictions for Tier 2 nations will be especially stringent as the rules block Tier 2 entities from helping fill any compute gaps.

These regulations also risk distorting deployment patterns. By increasing costs, the regulations could push providers to focus AI service provision on only the wealthiest nations and only high population centers to ensure profitability. Further, the hosting security controls will act to deeply limit – or more likely ban- any edge or self-hosted deployments of full, distilled, or quantized versions of these models. For Tier 2 entities, such deployments will be impossible in most cases due to the presumption of denial license required.

Overall, these effects could promote significant manufactured "AI scarcity" outside of the United States. For U.S. AI firms, this could mean missed market opportunities while potentially ceding market share to foreign firms unbound by such rules. For impacted nations, it could also mean lost productivity.

The impacts of any local AI scarcity will be most acute for emerging latency-sensitive applications. These include autonomous vehicles, AI agents, advanced cyber-defensive systems and other AI tools that require real time operation and likely cannot be effectively served by API. For such systems to provide real time intelligence, geographic latency means model weights must be physically close – a demand that will be hard to meet given security limitations on data centers and discouraged edge and self-hosted deployment. It must be stressed that these latency sensitive systems are among the most promising and beneficial AI use cases. By hampering local deployment, BIS could bind American multinationals from competing in these product categories.

In the specific case of real time cyber-defense, these risks are most significant. Through an AI scarcity, BIS may deprive allies of latency-dependent cyber defensive technologies including anomaly detection, vulnerability analysis, and just-in-time security patching. Scarcity could be greatest for Tier 2 NATO allies, including Poland and the Baltic States, risking significant holes in NATO's overall cyber defenses. As the billions in domestic losses from the foreign born NotPetya worm illustrated: foreign cyber weaknesses are a doorway to harming the United States. Any denial of AI cyber defenses abroad could risk cyber defense at home.

Model controls also bear serious diplomatic risks. By broadly restraining frontier AI deployment, the United States will impose a strict system of manufactured global AI inequality. Any perceived slowing or outright denial of AI – even if temporarily – by the United States will actively tarnish the global image of our AI sector. Further limits on the ability of Tier 2 nations to both develop and deploy frontier models through homegrown entities will only add to a sense of inequality, heightening diplomatic risk. Spited by U.S. policy, international users may gain a lasting preference for the comparatively permissionless Chinese alternatives.

Limitation of future AI possibilities.

The final risk of these controls is to deny countless emergent AI form factors and possibilities. These rules make myriad assumptions about the basic form of AI that may not apply to future applications. Anything diverging from the established regulatory norms may be effectively banned.

Autonomous logistics driven by driverless vehicles, trucks, and ships is the most significant. For Tier 1 and 2 nations, the assumption that models must be datacenter hosted necessarily means autonomous systems operating in-country cannot run either full versions of advanced AI models, or advanced models distilled or quantized to match edge-device requirements. For autonomous systems operating between nations, a possibility in the shipping industry, the assumption that models remain stationery creates even greater challenges. If porting between nations with varied restrictions, shipping may be snarled by a web of controls and licenses. Such misfit regulations will deeply hamper if not deny the United States leadership in a future autonomous logistics industry.

Autonomous systems are just one predictable example of technologies these regulations fail to account for. Numerous other possibilities – such as distributed AI training and inference - may also be limited. While such rules can certainly update to accommodate emergent AI form factors, BIS must be clear eyed that such technologies are unlikely to emerge if firms believe investment success hinges on regulatory change.

Conclusion

Amid intensifying international AI competition, rapid, responsible AI diffusion is more important than ever. While the impulse to control such powerful technology is understandable, the downsides of limiting model weight diffusion cannot be understated. By creating a global model weights license requirement, BIS will immediately foster significant business uncertainty, potentially depressing American led innovation. Further, the stringent specifics of these rules will slow advanced AI diffusion in both Tier 1 and Tier 2 nations and restrict countless future technologies unmatched to current regulatory assumptions.

Curbing this nascent technology could dramatically shape future success. While concerns about adversarial access to AI's unique potential are valid, diffusion is an unavoidably necessary "resource." By instead pursuing continued openness, the Administration can ensure innovation continues and American commerce leads the AI age.