SUBMITTED VIA REGULATIONS.GOV

October 11, 2024

Ms. Thea D. Rozman Kendler

Assistant Secretary for Export Administration

U.S. Department of Commerce

1401 Constitution Avenue, NW

Washington, DC 20230

RE: Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters (RIN 0694-AJ55)

Assistant Secretary Rozman Kendler,

Intel appreciates the opportunity to provide comments on the BIS Proposed Rule on the Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters (RIN 0694-AJ55).

Intel plays an important role in Artificial Intelligence (AI). Intel's full spectrum of hardware and software platforms offer open and modular solutions which support AI workloads and fuel emerging usages such as AI at the edge. Intel is committed to advancing AI technology responsibly and contributing to the development of principles, international standards, best practices methods, tools, and solutions.

BIS states that this proposed rule is expected to apply to only a small number of entities – only those companies developing or intending to develop a dual-use foundation model and those companies, individuals, or other organizations or entities that acquire, develop, or possess potential large-scale computing clusters. Intel is concerned that the lack of clarity in certain definitions in the proposed rule could inadvertently include activities as "in scope" that otherwise would not be subject to these reporting requirements. We offer the following input on BIS' efforts related to topics outlined in the Agency's notice.

Definition of "Training" or "Training Run":

The proposed threshold for the reporting requirements in Section 2(a)(i) is:

> *Conducting any AI model training run using more than 10^26 computational operations (e.g., integer or floating-point operations)*

The proposed rule includes the following definition:

> *Training or training run refers to any process by which an AI model learns from data using computing power. Training includes but is not limited to techniques employed during pre-training like unsupervised learning and employed during fine tuning like reinforcement learning from human feedback.*

The use of the words "not limited to" could create confusion and inadvertently result in reporting beyond the intended scope of the proposed rule. For instance, "optimizing a trained AI model" refers to adjusting a trained AI model to achieve desired performance (e.g., faster inference), efficiency (e.g., lower computational costs, memory usage), accuracy (e.g., in predictions), or other operational metrics. Exemplary techniques include one or more of quantization, pruning, sparsification, knowledge distillation, and model compression. It is important to provide clarification that these types of optimization techniques are not intended to be within scope of the reporting requirements. A way to clarify this issue is to revise the definition of "training or training run" to exclude optimization and, if needed, create a new definition in the rule for "optimize trained AI models" such as the language provided below (proposed changes underlined).

> *Training or training run refers to any process by which an AI model learns from data using computing power. Training includes but is not limited to techniques employed during pre-training like unsupervised learning and employed during fine tuning like reinforcement learning from human feedback, <u>but excludes techniques employed to optimize trained AI models</u>.*
>
> *As needed:*
> <u>"Optimize trained AI models" refers to adjusting trained AI models without additional learning from data to achieve desired performance, efficiency, accuracy, or other operational metrics.</u>

<u>Definitions regarding Large-scale Computing Clusters:</u>

Section 2(a)(ii) of the proposed rule states:

(ii) Acquiring, developing, or coming into possession of a computing cluster that has a set of machines transitively connected by data center networking of greater than 300 Gbit/s and having a theoretical maximum greater than 10^20 computational operations (e.g., integer or floating-point operations) per second (OP/s) for AI training, without sparsity.

There are several aspects of this text that would benefit from additional clarity.

1) There is no definition of the term "transitively connected" and it does not appear to be an industry standard term. Clear requirements would facilitate consistency of reporting; without clarity, reporting entities may interpret their obligations differently.
2) Clarification of the term "theoretical maximum" is also important. For example, AI hardware may support various precision number formats such as Floating Point 32 (FP32), Floating

Point 8 (FP8), or Brain Floating Point 16 (BF16). The "theoretical maximum" should not be evaluated by summing every permutation of precision number formats supported. Rather, the "theoretical maximum" should include reference to the maximum MAC operations per second weighted by bit length as claimed by the cluster or processor manufacturer, as applicable. This would be consistent with guidance from a previous Interim Final Rule[1] as shown below.

> *The rate of 'MacTOPS' is to be calculated at its maximum value theoretically possible. The rate of 'MacTOPS' is assumed to be the highest value the manufacturer claims in annual or brochure for the integrated circuit. For example, the 'TPP' threshold of 4800 can be met with 600 tera integer operations (or 2 x 300 'MacTOPS') at 8 bits or 300 tera FLOPS (or 2 x 150 201 'MacTOPS') at 16 bits. If the IC is designed for MAC computation with multiple bit lengths that achieve different 'TPP' values, the highest 'TPP' value should be evaluated against parameters in 3A090.*

Updating of Technical Parameters:

The proposed rule states that "BIS will update these technical conditions as appropriate." Given the pace of innovation in AI, BIS should consider predictable, timely updates to the technical parameters such as raising the collection thresholds. BIS should also provide any additional clarifications needed if the language in the rule appears to apply more broadly than intended.

Intel welcomes the opportunity to discuss our feedback regarding this Proposed Rule. We look forward to continued engagement with the Department of Commerce on this and other matters relating to the safe and secure development and deployment of AI technologies.

Sincerely,

/s/ Jayne Stancavage

Jayne Stancavage
Vice President, Policy and Regulatory Affairs

[1] https://www.bis.doc.gov/index.php/documents/federal-register-notices-1/3353-2023-10-16-advanced-computing-supercomputing-ifr/file