



# **OpenAl**

October 11, 2024

From: Ben Rossen Member of Legal Staff OpenAl

To:

Sean Delahanty
Office of Strategic Industries and
Economic Security
Bureau of Industry and Security
Department of Commerce

RE: Proposed Rule on the Establishment of Reporting Requirements for the Development of Advanced Artificial Intelligence Models and Computing Clusters (Docket No. 240905-0231)

Dear Mr. Delahanty:

OpenAl values the opportunity to provide feedback on BIS's proposed rule amending 15 CFR Part 702 and its reporting process for advanced Al models and large computing clusters. Attached, please find OpenAl's comments and suggestions related to the proposed rule.

#### **Executive Summary**

OpenAl recognizes the challenges BIS is facing in addressing a complex and developing industry, and appreciates the effort to help build BIS's expertise through a collaborative process. In this submission, OpenAl suggests:

- Adjusting the frequency of reports to every six months to reduce the burden on both BIS and U.S. parties subject to these reporting requirements ("Respondents") by better aligning with development cycles and the time period covered by each notification:
- II. Specifying a mechanism to streamline notifications of previously-reported activities;
- III. Replacing the ambiguous and open-ended language for requesting "other information" with concrete and standardized requests that are within the scope of Executive Order 14110 and the Defense Production Act;
- IV. Ensuring safe and secure handling of information submitted throughout the notification and reporting process;
- V. Continuing progress on defining thresholds and technical standards to avoid sending Respondents potentially overbroad requests for information.

# I. Reporting Frequency and Burden

## A. Summary of Currently Proposed Reporting Cadence

The proposed rule sets forth a quarterly reporting schedule, to be initiated by Respondents, notifying BIS of covered activities. Following the initial report, BIS will send questions to Respondents. After Respondents answer BIS's initial questions, BIS may ask follow-up questions. The proposed rule does not specify BIS's timing, but Respondents' answers are due 30 days after BIS's initial questions and 7 days after any follow-up.

The text of the proposed rule includes an alternative notification for affirmations of "no applicable activities," but does not include the "simple notification" contemplated by the NPRM for streamlined responses when a Respondent has no additions, updates, or changes.

#### B. Areas of Concern

Under the proposed timelines, the quarterly reporting period would effectively create a continuous reporting process, burdening both Respondents and BIS. For example, a Respondent may receive BIS's questions two weeks after the initial notification and respond 30 days thereafter. A Respondent may then receive follow-up questions two weeks later, and must respond within seven days. This question-and-answer cycle would take more than two months of each three-month quarter. Under this cadence, Respondents would need to immediately begin on the next reporting cycle, and slight delays could result in overlapping cycles.

A six-month cadence is appropriate for large foundation models, which generally take months to train. Where applicable, a six-month cadence would also capture iterative releases, which similarly have multi-month development and testing pathways. The proposed rule already contemplates Respondents' reporting on a six-month period of planned activities. A six-month cadence for report submission would allow BIS to maintain desired visibility into planned activities and reduce unnecessary burden on both industry and BIS.

Further, the proposed rule does not contain the simplified notification described in the NPRM. A simplified notification, focused on new activities and substantive changes to those previously reported, would help BIS focus on updated information and reduce duplicative efforts.

These concerns are amplified by the ongoing and indefinite nature of the proposed rule. Unlike previous industrial base surveys, which were one-time snapshots of an industry's condition, the proposed rule is not time-limited and would pose a continuous burden, with the potential to grow larger with every cycle.

C. <u>Proposed Modification: Change from a quarterly notification</u> schedule to a semiannual schedule

#### **Proposed Changes to Rule Language**

§ 702.7(a)(2)(i) Notification of applicable activities. Covered U.S. persons subject to the reporting requirements in paragraph (a)(1) of this section must notify BIS of 'applicable activities' via email twice per year each quarter, identifying any 'applicable activities' planned in the six months following notification. Semiannual Quarterly notification dates are as follows: Q1—April 15 and Q2—July 15; Q3—October 15; Q4—January 15. For example, in a notification due on April 15, a covered U.S. person should include all activities planned until October 15 of the same year.

#### Rationale

This revision reduces the burden to both Respondents and BIS by extending the notification period to better accommodate the time required for questions and answers. A six-month reporting cadence aligns with the proposed rule's focus on covered activities during a six-month period, while also better aligning with product development cycles, and would reduce duplicative submissions.

D. <u>Proposed Modification: Provide a streamlined process for</u> reporting subsequent cycles

#### **Proposed Changes to Rule Language**

§ 702.7(a)(2)(v) Affirmation of no previously unreported applicable activities. For each of the three six-month periods seven quarters following the period quarter covered by a notification of 'applicable activities,' if the covered U.S. person has no 'applicable activities,' (i.e., an "applicable activity" that meets the criteria under paragraph (a)(1)(i) or (ii) of this section), or no applicable activities that were not included in a previous notification, to report, they must submit an affirmation of no activities or no previously unreported applicable activities (as appropriate) by emailing ai\_reporting@bis.doc.gov on the semiannual quarterly notification date. If the covered U.S. person submits an affirmation of no previously unreported applicable activities for three seven consecutive six-month periods quarters, they need not provide BIS with any affirmation thereafter until they have 'applicable activities' to report.

#### Rationale

This revision reduces the burden to both Respondents and BIS by reducing duplicative submissions, consistent with the NPRM's discussion of the rule indicating that the "reporting system will allow for companies that have no additions, updates, or changes since the last report to make a simple notification to that effect." With this revision, Respondents would be able to easily affirm that they have no new covered activities or updates to those previously reported, while efficiently notifying BIS of changes from the last reporting cycle. This revision is aimed at the notification process and ongoing reported activities would remain subject to the proposed rule, consistent with § 702.7(a)(ii) and (iv).

#### II. Scope of Reported Information and Requests from BIS

### A. Summary of Currently Proposed Scope

In its current form, the proposed rule specifies that Respondents will identify any applicable activities in the initial notification to BIS, then respond to a series of questions from BIS under the three categories specified in Executive Order 14110. The proposed rule also includes an ambiguous fourth category for questions broadly seeking "other information pertaining to" safety, reliability, and national security.

#### B. Areas of Concern

As written, the proposed rule would allow BIS to ask questions about nearly any topic or information plausibly related to safety, reliability or security, and does not provide Respondents with a concrete understanding of what information BIS is seeking. Particularly, the proposed language of § 702.7(b)(2)(iv) states that BIS's questions "must address... Other information pertaining to the safety and reliability of dual-use foundation models, or activities or risks that present concerns to U.S. national security." OpenAI recognizes the need for flexibility to address a complex and still-developing industry. However, the proposed language does not indicate what BIS will ask of Respondents and raises the risk that Respondents will need to address broad, bespoke questions about frontier models every few months. The open-ended nature of the proposed rule is potentially burdensome and would add costs to each reporting cycle.

Notably, the proposed language of § 702.7(b)(2)(iv) does not appear in Executive Order 14110, Section 4.2(a), which is limited to the three topics in proposed §702.7(b)(2)(i)–(iii), each of which recites specific categories of information. The broad, non-specific language of the added fourth category is also inconsistent with BIS's authority under the DPA, which requires "the scope and purpose" of an inquiry to be defined in advance (e.g., 50 U.S.C. § 4555(a)), and with historical uses of the DPA's reporting

requirements, which sought specific information rather than asking undefined, open-ended questions.

Without clarification, this language is likely to decrease efficiency and hamper the information-gathering purpose of this rule. Respondents will only be able to speculate as to how to prepare for this open-ended request and, when paired with the regular reporting timelines described above (whether quarterly or every six months), a lack of standardized or pre-determined questions will make it more difficult for companies to respond to BIS efficiently. Barring a more definite and standardized scope, companies will not have an opportunity to begin collecting and formatting relevant information unless and until BIS asks—at which point Respondents would have a relatively short timeframe to respond to broad BIS questions.

As discussed further below, the confidential information submitted to BIS through these requests is highly sensitive and critical to Respondents. Without knowing how BIS intends to use the information, the unconstrained language of the proposed rule compounds those concerns.

C. <u>Proposed Modification: Standardize reporting requirements</u> and inform Respondents in advance

#### **Proposed Changes to Rule Language**

§ 702.7(b)(2)(iv) Other standardized information pertaining to assessing the safety and reliability of dual-use foundation models, or assessing activities or risks involving dual-use foundation models that present concerns to U.S. national security. BIS will base these requests on industry feedback and publish the questions at least six months in advance of making such requests. For example, for requests following a notification due on October 15, BIS will publish the requests by April 15 of the same year.

#### **Rationale**

This revision better aligns the proposed rule with the mandate of Executive Order 14110 to gather information on dual-use foundation models and its focus on measurable risk factors like "physical and cybersecurity protections." This is also consistent with the authority granted under § 4555(a) of the DPA, which requires that information requests occur "only after the scope and purpose of the investigation, inspection, or inquiry to be made *have been defined* by competent authority."

By publishing standardized requests in advance, BIS provides Respondents ample time to understand the information being sought and prepare for the upcoming reporting cycle. This also provides BIS with flexibility to refine its requests, as it collects and analyzes information during each cycle.

As an example, such standardized requests could include information pertaining to the following categories for each frontier model over the reporting threshold:

- 1. Model name:
- 2. Qualifying threshold;
- 3. Amount of operations used;
- 4. How the amount was calculated;
- 5. Bit length of operations;
- 6. Training start date;
- 7. Training (or checkpoint) end date; and
- 8. Anticipated or actual results on common benchmarks.

#### III. Report Handling and Security

A. <u>Summary of Currently Proposed Language regarding</u>
Reporting, Sharing, and Securing Information

The proposed rule specifies that notifications would be emailed to BIS and further responses would be "in the form and manner that BIS will prescribe in instructions sent to the covered U.S. person." Notably, the proposed rule does not indicate how any information obtained will be used, shared, or secured by BIS.

#### B. Areas of Concern

As recognized in the NPRM, information submitted may be extremely sensitive. This information is often confidential to the Respondent and its secrecy critical to business plans and competition. However, the proposed rule does not indicate what BIS will do with such sensitive information or how that information will be protected from exposure, intentional or otherwise. Executive Order 14110 cites the DPA as authority for the reporting requirements "to ensure and verify the continuous availability of safe, reliable, and effective AI in accordance with the Defense Production Act, . . . including for the national defense and the protection of critical infrastructure." This mandate sets a clear boundary, that the reporting requirements are solely to assess the availability of safe, reliable, and effective AI—and no more.

Section 4555(d) of the DPA requires that confidential information gathered under its authority "shall not be published or disclosed unless the President determines that the withholding thereof is contrary to the interest of the national defense." Under this clear direction, BIS must take efforts to protect any submitted information from disclosure. As an initial matter, the proposed rule begins its notification process with sending information by email rather than a more secure method that would reflect the sensitivity of such information. Steps can, and should be, taken to

ensure that the reporting process provides adequate safeguards for Respondents.

Consistent with the Executive Order's limitation on assessing availability and the confidentiality requirements under the DPA, BIS should make it clear that any information obtained through this reporting process will not be shared in any way beyond that process and should be fully transparent with Respondents about how BIS will safeguard the information it receives. Sharing the reporting information it receives would pose risks that far exceed the stated goal of the proposed rule. For example, if BIS shares this information—even within the Department of Commerce—that may inadvertently expose the information outside of the intended scope of the reporting requirements. Put simply, the disclosure of reported information in any capacity outside of the reporting process could irreparably harm Respondents.

C. <u>Proposed Modification: Include obligations for BIS to protect</u> <u>received information and limit sharing beyond the intended</u> scope of the reporting requirement

#### **Proposed Changes to Rule Language**

§ 702.7(a) Reporting requirements. (1) Covered U.S. persons are required to submit a notification to the Department by <u>secure submission emailing ai reporting@bis.doc.gov</u> on a quarterly basis as defined in paragraph (a)(2) of this section if the covered U.S. person engages in, or plans, within six months, to engage in 'applicable activities,' defined as follows:

. . .

(a)(2)(i) Notification of applicable activities. Covered U.S. persons subject to the reporting requirements in paragraph (a)(1) of this section must notify BIS of 'applicable activities' via <u>secure submission</u> <u>email</u> each quarter, identifying any 'applicable activities' planned in the six months following notification. Quarterly notification dates are as follows: Q1—April 15; Q2—July 15; Q3—October 15; Q4—January 15. For example, in a notification due on April 15, a covered U.S. person should include all activities planned until October 15 of the same year.\*

. . .

(b)(1) All information submitted under this section shall be filed with BIS in the <u>secured</u> form and manner that BIS will prescribe in instructions sent to the covered U.S. person after BIS has received a notification of `applicable activities.' <u>BIS</u> will treat any information received as confidential under 50 U.S.C. §4555(d) and take appropriate measures to safeguard any information received and prevent its dissemination or use outside of this notification and reporting process. Such information will be used for the sole purpose to ensure and verify the continuous availability of safe, reliable, and effective AI.

\* See also proposed changes to 702.7(a)(2)(i) reporting frequency in Section Labove.

#### Rationale

These revisions reflect BIS's recognition "that the information collected through these reporting requirements is extremely sensitive" and establish measures that will be taken to ensure that Respondents' information is handled securely. For example, even reporting activities at a high level via unencrypted email could pose a risk of exposing Respondents' business plans or competitive position. BIS should provide a secure method for submitting notifications, such as an encrypted web portal, and do so for each step of the reporting process.

Likewise, once BIS has a Respondent's information, that information should be handled in light of its extremely sensitive nature and BIS should limit its use of the information to the purpose of this rule, namely collecting information on potential dual-use foundation AI models and large computing clusters. Further, information collected under authority of the DPA "with reference to which a request for confidential treatment is made by the person furnishing such information shall not be published or disclosed unless the President determines that the withholding thereof is contrary to the interest of the national defense." Accordingly, BIS is obligated to maintain the confidentiality and security of information provided in this reporting process, even within government. In addition, recognizing that the government has endeavored to collaborate with industry on various topics related to AI models and compute clusters, BIS should make clearly defined and transparent efforts to ensure that the collected information is not shared in those forums.

#### IV. Compute Threshold and Technical Criteria

# A. <u>Summary of Currently Proposed Compute Thresholds and</u> Technical Criteria

The proposed rule would require notifying BIS of AI models and compute clusters that meet certain compute thresholds, derived from the technical requirements in Executive Order 14110. The proposed rule differs from those requirements by raising the computing cluster networking requirement to 300 Gbit/s and specifying that the operations per second metric for such clusters is for AI training without sparsity.

#### B. <u>Areas of Concern</u>

The technical thresholds in the proposed rule remain largely unchanged from those announced in Executive Order 14110 nearly a year ago, which instructed the Secretary of Commerce to define and regularly update these requirements. However, it remains unclear the extent to which compute thresholds are in fact linked to risk and frontier model developers increasingly believe such thresholds will not be a durable

proxy. We believe a more appropriate threshold would be based on capabilities but recognize that this remains an area of active research. We encourage BIS to work with us and other stakeholders to develop an appropriate threshold as this research advances.

Even if BIS maintains a compute threshold, steps can be taken to reduce ambiguity in the available definitions, which lack detail necessary for Respondents to understand whether models or clusters meet these thresholds. For example, the thresholds reference floating-point operations without specifying which level of precision to assess this metric under.

Further, the technical thresholds in the rule depart from the Executive Order, broadening to "any Al model training run using more than 10^26 computational operations" rather than applying only to dual-use foundation models, as contemplated in Section 4.2(a) of the Executive Order. As discussed above, the scope of information reported under the proposed rule, as written, is indefinite and a lack of clarity in the technical thresholds increases the uncertainty for those covered by the proposed rule.

C. <u>Proposed Modification: Clarify that the compute thresholds</u> <u>are based on double-precision operations</u>

#### **Proposed Modification**

§ 702.7(a)(1)(i) Conducting any AI model training run using more than 10^26 computational operations (e.g., integer or double-precision (64-bit) floating-point operations); or

(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 double-precision (64-bit) floating-point operations) per second (OP/s) for Al training, without sparsity.

#### Rationale

This revision provides more clarity on how the technical thresholds are calculated. Using double-precision for floating point operations is consistent with other regulations, such as the EAR definition of "supercomputer" (15 CFR 772.1), and standard practices for describing compute cluster performance, such as reported to the <a href="Top500 List">Top500 List</a> or used to describe Department of Energy clusters like <a href="Frontier">Frontier</a> at Oak Ridge National Laboratory.

We appreciate the opportunity to engage with your office on this matter and your consideration of these comments and suggested revisions to the proposed rule.

Sincerely,

Ben Rossen Associate General Counsel OpenAl