Contribution ID: 9c2cf7be-07f0-4834-8b12-7b5368c2da9d
Date: 20/05/2025 17:18:24

EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

Consultation: Commission Guidelines to Clarify the Scope of the General-purpose AI Rules in the AI Act

Fields marked with * are mandatory.

1. Background and purpose of the consultation

The AI Office is launching a multi-stakeholder consultation to assist in the preparation of guidelines on general-purpose AI which aim at clarifying the scope of the rules for providers of general-purpose AI models in Regulation (EU) 2024/1689 ('AI Act'). Those rules will enter into application on 2 August 2025.

A working document by the AI Office forms the basis for this consultation. Stakeholders are invited to carefully read the document. Each question in the present survey is linked to dedicated sections in the working document.

Please reply to this survey by Thursday, 22 May, 12:00 (noon) CET.

The AI Office is dedicated to facilitate compliance of providers of general-purpose AI models with their obligations under the AI Act. To this end, the Commission guidelines on general-purpose AI are expected to clarify key concepts in the AI Act, such as what is a 'general-purpose AI model', a 'provider of a general-purpose AI model', a 'placing on the market of a general-purpose AI model', and how to estimate the computational resources used for training a general-purpose AI model. Beyond these conceptual clarifications, the guidelines are expected to clarify how the AI Office will work with providers who must comply with the general-purpose AI rules, to support them in their compliance.

The Commission guidelines on general-purpose AI will complement the General-Purpose AI Code of Practice ('Code') which will set out commitments to which providers of general-purpose AI models may adhere to ensure compliance with their obligations under the AI Act. Both the Commission guidelines on general-purpose AI and the final General-Purpose AI Code of Practice are expected to be published in May or June 2025.

This targeted consultation aims to gather a broad range of input and perspectives. We invite submissions from all stakeholders with relevant expertise and perspectives through this survey, particularly

from industry actors such as providers of general-purpose AI models and downstream providers of AI systems built on those models, civil society, academia, other independent experts, and public authorities.

We welcome full or partial replies from all respondents based on their expertise and perspective.

The responses to this consultation will provide important input to the Commission when preparing the guidelines. The working document does not prejudge the final decision that the Commission may take on the guidelines.

The AI Office will publish a summary of the results of the consultation. Results will be based on aggregated data and respondents will not be directly quoted.

In case you face any technical difficulties or would like to ask a question, please contact: CNECT-A3@ec.europa.eu

2. About you

* Do you represent one or more organisations (e.g., industry organisation or civil society organisation) or act in your personal capacity (e.g., independent expert)?
  - (●) Organisation (s)
  - ( ) In a personal capacity

* Please specify the name(s) of the organisation(s).

Institut des droits fondamentaux du numérique (iDFrights)

* First name

Jean-Marie

* Last name

CAVADA

* E-Mail address (this won't be published)

contact@jeanmariecavada.eu

* Is your organisation headquartered in the EU?
  - (●) Yes
  - ( ) No
  - ( ) Other (e.g. multiple organisations)

* EU member states

- Austria
- Belgium
- Bulgaria
- Croatia
- Cyprus
- Czechia
- Denmark
- Estonia
- Finland
- (•) France
- Germany
- Greece
- Hungary
- Ireland
- Italy
- Latvia
- Lithuania
- Luxembourg
- Malta
- Netherlands
- Poland
- Portugal
- Romania
- Slovak Republic
- Slovenia
- Spain
- Sweden

* What is the size of your organisation?

- (•) Micro (1 to 9 employees)
- Small (10 to 49 employees)
- Medium (50 to 249 employees)
- Large (250 or more employees)
- Other (e.g. multiple organisations)

* Which stakeholder category would you consider yourself in?

- ○ Provider of a general-purpose AI model, or acting on behalf of one
- ○ Downstream provider of an AI system based on general-purpose AI models, or acting on behalf of such providers
- ○ Other industry organisation, or acting on behalf of such organisations
- ○ Academia
- ◉ Civil Society Organisation
- ○ Rightsholder or a collective management organisation (CMO) or an independent management organisation (IME) or the representative of an organisation acting on behalf of rightsholders (other than a CMO or IME)
- ○ Public authority
- ○ Others

* Please briefly describe the activities of your organisation or yourself.

1000 character(s) maximum

We are bringing together lawyers, academics, and distinguished professionals whose work is focused on digital technology, AI, and intellectual property.

Please do not share any confidential information in your contribution.

Privacy statement

☑ I acknowledge the attached privacy statement.

Privacy_statement_-_Guidelines_consultation.pdf

3. Preliminary approach for the content of the guidelines

3.1 General-purpose AI model

The definition of “general-purpose AI model” is key to understanding whether an entity must comply with the AI Act’s rules for general-purpose AI models. **See section 3.1 of the working document.**

3.1.1 Conditions for sufficient generality and capabilities

See section 3.1.1 of the working document.

1) Many entities will have to assess the general-purpose nature of their models to determine whether they need to follow the obligations for providers of general-purpose AI models. A pragmatic metric is thus highly desirable to limit the burden, especially on smaller entities. Do you agree that training compute is currently the

best metric for assessing generality and capabilities, despite its various shortcomings?

- ○ Yes
- ◉ No

Please explain why and which alternatives may be preferable.

The size of a training calculation does not necessarily guarantee a high-performande model ; a model trained on a large set of targeted or poor-quality data will be less effecvtive. A model should be trained on a smaller but higher quality dataset. In addition, two models with the same calulation can have radically different performances depending on the arthitecture used.
In order to assess the general-purpose nature of their models, which is the cornerstone of the AI Act architecture, other criteria should be added in order to better reflect both the size, the performance and the central position in the value chain of these models.
Many experts refer to these models as being the « Operating Systems » of the AI value chain. Compute related metrics fail to embace this specific situation.

2) Is 10^22 FLOP a reasonable threshold for presuming that a model is a general-purpose AI model?

- ○ Yes
- ◉ No

Please explain why and which alternatives may be preferable.

Determining a precise threshold remains tricky, but to guarantee effective data management and intellectual property protection in AI, it is essential to:

1-Promote transparency and traceability
2-Strengthen control mechanisms
3-Regulate access and usage
4-Establish standards and agreements

The AI Act allows the European AI Office to designate “GPAI” models based on alternative criteria (e.g., parameter count, dataset size, or business impact) if they fall below the FLOP threshold but still pose significant risks. Given that exact FLOP counts are proprietary, models like DeepSeek R1 and V3, because of their market impact and open-source nature, should be candidates for such designation.

The threshold should be more nuanced and include additional parameters to complement the quantitative criterion. For instance, any model that generates significant risks to personal data protection or intellectual property should also be eligible for designation.

3) With the proposed threshold of 10^22 FLOP, or your alternative threshold suggested above, how many models and how many entities do you expect to be in scope of the AI Act, and why?

If we consider this threshold as merely a starting point, it's clear that it must evolve alongside technological advances. Among the AI models that have already surpassed it, we find all the digital giants, Google, Meta, Microsoft, OpenAI's offerings, ByteDance, and at least twenty others trained on a comparable scale. These models will fall under the AI Act's scope, being deemed systemically risky. However, a company might sidestep current legislation by licensing a high-performance model that demands less compute.

Because exact FLOP counts are proprietary, the Act permits alternative criteria (e.g., parameter count, dataset size, benchmarks) to designate GPAI models.

By August 2025, newer models are likely to appear, and compute costs continue to drop (for example, DeepSeek's V3 cost $6 million versus GPT-4's $100 million), potentially expanding the pool of models at or below 10^{22} FLOP.

4) In addition to the examples presented in section 3.1.1 of the working document, are there other examples for which it would be important to clarify whether the presumption of being a general-purpose AI model based on the training compute threshold may be rebutted?

It's not just computing power that matters but also access to unique, massive datasets, a significant advantage. Finally, companies below the threshold can still develop specific innovations with major industry impact.

In AI, many models are trained on web-scraped data, including copyrighted works. This raises legal questions around crawling/scraping and the use of licensed content.

Another reason to include criteria beyond compute thresholds is non-generative, optimization-focused AI. High compute usage for domain-specific tasks, common in logistics or energy, does not imply GPAI qualification. Clarifying this distinction ensures the Act targets truly general-purpose AI, maintaining proportionality and supporting sector-specific innovation.

Therefore, leveraging alternative criteria (e.g., dataset characteristics, parameter counts, and task generality) helps identify the most critical value chain players, those with the greatest potential risk and ecosystem impact.

3.1.2 Differentiation between new models and model versions

See section 3.1.2 of the working document.

5) Besides the criteria presented in Section 3.1.2 of the working document, are there other criteria that can be used to determine whether iterations, instances, or derivates of a model constitute distinct models for the purposes of the AI Act?

Four alternative criteria could be used to assess whether a model constitutes a distinct instance:

Changes in Training Data Sources:
- Shifts in data sources, e.g., from public to proprietary or personal datasets—that introduce new IP risks (copyrighted material) or privacy risks (personal data).

- Risks in Output Generation:
Modifications in outputs that increase IP infringement potential (e.g., generating copyrighted text) or privacy breaches (e.g., leaking personal data).

- Fine-Tuning Involving Sensitive Data:
Retraining or fine-tuning on datasets containing IP-protected or personal data, thereby altering the model's risk profile compared to the original.
- Changes in Deployment Environment:
Redeployment in contexts that expose new user data (privacy risk) or integrate third-party IP-protected components (IP risk).

A model is considered "derived" when it presents substantial modifications that affect its autonomy relative to the original. Fine-tuning is treated as a derived modification, though the regulation is not entirely clear on this point. Furthermore, update management should be regarded as a continuation of the initial model. These examples are crucial in framing the regulatory scope of the AI Act.

6) In addition to the considerations presented in section 3.1.2 of the working document, are there other examples where it is unclear whether iterations, instances, or derivates of a model developed by the same entity constitute distinct models within the context of the AI Act?

3.2 Provider of a general-purpose AI model, including downstream modifiers

The definition of "provider" is key to understanding whether an entity must comply with the AI Act's rules for general-purpose AI models. **See section 3.2 of the working document.**

3.2.1 Examples of providers of general-purpose AI models

7) In addition to the considerations presented in section 3.2.1 of the paper, are there other examples for which it may not be clear which entity is the provider of a given general-purpose AI model?

There are several cases where the identity of a supplier of a general-purpose AI model may be ambiguous:
1° Models developed in collaboration (model trained jointly by a university and a private company)
2° modified open source models (when this model is modified and deployed by another entity
3° Models integrated into larger systems : integrated into platforms or software without being directly accessible
4° Models developed under contract (the case when a company develops a model for one of its customers)
5° Models with frequent updates.
These issues are discussed in the context of the AI Act, but there is no clear guidance on how to distinguish between these cases.

3.2.2 Downstream modifiers as providers of general-purpose AI models

This section of the guidelines is expected to clarify responsibilities along the AI value chain, by specifying the conditions under which an entity who modifies a general-purpose AI model ('downstream modifier') must comply with the obligations for all providers of general-purpose AI models, and the obligations for providers of general-purpose AI models with systemic risk respectively, in the AI Act.

The proposed approach of the AI Office is to set certain thresholds in terms of computational resources used for the modification, that, if met, mean the downstream modifier should be presumed to be the provider of the modified general-purpose AI model or general-purpose AI model with systemic risk, and subject to the relevant obligations in the AI Act.

This proposed approach draws a distinction between obligations for all providers of general-purpose AI models and obligations only for providers of general-purpose AI models with systemic risk.

See section 3.2.2 of the working document.

8) Many downstream modifiers will have to assess whether they need to comply with the obligations for all providers of general-purpose AI models and the obligations for providers of general-purpose AI models with systemic risk. A pragmatic metric is thus highly desirable to limit the burden on downstream modifiers having to make this assessment, especially on smaller entities. Do you agree that training compute is currently the best metric for quantifying the amount of modification, despite its various shortcomings?

- ○ Yes
- ◉ No

Please explain why and which alternatives may be preferable.

Downstream modifiers adapt GPAI models (e.g., fine-tuning Llama for medical applications) and must assess whether their changes cross the GPAI threshold (10^{22} FLOP) or the systemic-risk threshold (10^{25} FLOP), which would trigger obligations such as transparency, documentation, and risk mitigation.

A proportionate approach would be fairer: we shouldn't demand the same rigor from someone making minor tweaks as from a company undertaking major transformations. A minimum threshold should apply, especially when the modification doesn't fundamentally alter the model's capabilities or intended use.

Relying solely on compute ignores data quality and size, both critical to understanding a modification's impact (for example, fine-tuning on specialized medical datasets). Smaller entities often lack access to precise compute metrics for proprietary models, further complicating assessments. Moreover, functional changes (such as task specialization) or deployment optimizations may not correlate with compute at all.

In summary, training-compute (FLOP) is a poor proxy for measuring the scale or significance of modifications, since it doesn't capture their nature or impact.

9) Are there examples of modifications of general-purpose AI models that meet the proposed training compute threshold of 3x10^21 FLOP, yet which should not result in the downstream modifier being considered a provider?

The risk is quite the opposite. Without appropriate constraints, downstream providers could make high-impact modifications (e.g., fine-tuning with sensitive data, altering outputs to infringe IP or breach privacy) without triggering provider obligations if they fall below the 3×10^{21} FLOP threshold or evade market-placement criteria.

Examples of risky modifications:
- Fine-tuning Llama with proprietary data to generate copyrighted content, bypassing IP-transparency requirements.
- Retraining GPT-4 variants on personal data, risking GDPR violations without privacy-risk assessments.
- Modifying DeepSeek R1 for critical applications (e.g., healthcare) without systemic-risk oversight, despite significant societal impact.

Such loopholes could undermine the Act's goals of ensuring trustworthy AI, protecting fundamental rights, and mitigating systemic risks.

10) Are there examples of modifications of general-purpose AI models with systemic risk that do not meet the proposed training compute threshold of one third of 10^24 FLOP, yet which significantly change the systemic risk profile of the model in ways that could not have been reasonably foreseen by the upstream model provider?

Yes, there are several cases where downstream modifications to a general-purpose model can lead to unforeseen systemic risks, even without exceeding the 10^{24} FLOP threshold.
For example, a general language model may be fine-tuned on confidential medical data, which raises risks of data leakage, misinformation, and non-compliance with data-protection regulations.

3.3 Placing on the market of a general-purpose AI model and the open-source exemptions

The definition of “placing on the market” is key to understanding whether an entity must comply with the AI Act’s obligations for providers of general-purpose AI models.

See section 3.3 of the working document.

3.3.1 Examples of placing on the market of general-purpose AI models

11) In addition to the examples presented in section 3.3.1 of the working document, are there other examples of when a general-purpose AI model should be considered as being placed on the market?

Yes, there are other cases:
- If a model is integrated into a SaaS platform and made accessible to users without the provider directly offering the model.
- If an AI model is exposed via a public API that developers can use to build applications, even though the company doesn’t sell the model itself.
- If an open-source model is adopted at scale by commercial entities.
- If a model is embedded in marketed software or devices (e.g., voice assistants), even when the AI component isn’t sold separately.

3.3.2 Exemptions from certain obligations for certain open-source releases

12) What are examples of ways in which open-source general-purpose AI models can be monetised?

There are already workarounds used by open-source models to monetize them, including:
- “Premium access,” offering a free version and charging for advanced features
- Integration into advertising-funded platforms through strategic partnerships

The risk is that the creators of the models will see their work used without compensation, while companies profit, and intellectual property issues persist, since some models are trained on protected data, potentially leading to legal problems.

13) What are examples of ‘information on usage’ as stated in Articles 53(2) and 54 (6) AI Act for open-source models?

14) What are examples of free and open-source licenses in the sense of the AI Act that allow for the access, usage, modification, and distribution of general-purpose AI models, and also require the release of publicly available information on the model parameters, including the weights, the model architecture, and model usage?

Yes, some open-source licences allow this, some require publication of changes made to the model (e.g., Apache License 2.0, Creative Commons licences), while others are more permissive and do not require publication of changes (e.g., MIT License).

3.4 Estimating the computational resources used to train or modify a model

The AI Act includes a threshold involving computational resources (compute) used to train a general-purpose AI model which can lead the model to be classified as a general-purpose AI model with systemic risk. Sections 3.1.1 and 3.2.2 introduce thresholds which also involve compute. To know whether a model meets any of these thresholds, potential providers must estimate the amount of compute used.

See section 3.4 of the working document.

3.4.1 Estimating the amount of compute used for training or modification

15) Do any of the formula provided in section 3.4.1 require further clarification? If yes, please specify.

This formula should be more precise, particularly to improve its applicability and accuracy in assessing AI training computation. The current formula is not fully adapted to hybrid models, and an adaptation would better reflect the diversity of approaches used. For open-source models, the formula should also include specific parameters to better assess the real impact of these models.

16) Are there any cases where a potential provider or downstream modifier would be unable to estimate the relevant amount of compute using any of the formula provided in this section? If so, why?

The formulas proposed in Section 3.4.1 may not be suitable for all model types, particularly those employing hybrid, multi-architecture approaches or advanced optimization methods, and some providers or modifiers lack complete visibility into the resources used to train the model.

17) In addition to the approaches presented in the respective section of the paper, are there other ways for providers to estimate the amount of computational resources used for training?

3.4.2 Estimating the cumulative amount of computational resources used for training

18) What are examples of activities and methods used during training or directly feeding into training that are intended to enhance the capabilities of the model prior to its deployment, beyond pre-training, fine-tuning and synthetic data generation?

Yes, there are several methods used, for example, fine-tuning and transfer learning, which allow a pre-trained model to be adapted to a specific domain by adjusting its parameters with additional data, or reinforcement learning, which trains models to make optimal decisions based on rewards or penalties, but regardless, these methods can raise legal issues, particularly regarding the use of training data and the reuse of pre-trained models from an intellectual property perspective.

19) Are there examples of activities and methods that are specifically aimed at making the model safer, but which do not at the same time change the model's capabilities, and what would represent a rigorous justification that this is the case?

Yes, there are methods that target the classification of models without altering their capabilities, and these are certainly more suitable for the protection of intellectual property:

• Model evaluation and certification, which enables models to be tested and classified according to performance and compliance criteria without altering their operation, thereby guaranteeing their reliability and compliance with regulations.
• Indexing and traceability of models, which allows tracking the origin and evolution of models without modifying their structure, and is essential for protecting intellectual property and preventing unauthorized use.

20) What other activities and methods used during training should not be counted as part of cumulative training compute, and why?

These exclusions should make it easier to distinguish the stages that have a direct influence on learning the model from those necessary for its development and deployment, as they are generally omitted for methodological reasons.

In particular, data loading and processing are not considered part of cumulative training because they do not directly modify model parameters, and validation and testing, used to evaluate model performance, are not included in the training calculation because they do not contribute to learning the model itself. These methods provide a better framework for model use and ensure respect for partners. By excluding certain activities from the cumulative training calculation, it becomes easier to track the origin of data and algorithms, which strengthens protection against copyright infringement.

21) How may providers reasonably and in a practically feasible way estimate the amount of computational resources used for synthetic data generation when the generating model is not their own model (for example a closed-source model accessed via API) or when the synthetic data set has been obtained from a third party (taking into account the possibility that the data set may not represent the entirety of the synthetic data generated to produce the data set, for example if a selection process was conducted), and how accurate would these estimations be?

There are several ways of doing this, notably by analysing logs and API metrics. When models are accessible via an API, service providers can generate usage logs to estimate resource consumption, and API providers can furnish metrics on the volume of data processed and processing time, enabling an indirect estimate of resources used.

Another example is tracking cloud costs and hardware resources, since service providers can use invoices from cloud providers to estimate consumption. There is also the selection process and its impact, to determine whether data filtering (e.g. removing synthetic data) has been conducted, which may affect resource-use estimates.

22) When might a provider reasonably be expected to know how much compute they will use in post-training?

In general, the provider can supply a reliable estimate after validation tests, once the model's performance has stabilized, the computational requirements for the benchmark are well understood, and optimization is complete.

Several factors influence this estimate, in particular the model's complexity, since more complex models demand more computation, the type of training, whether supervised or unsupervised, has different requirements, and finally algorithm optimization, as post-training adjustments such as model compression or parameter tuning help reduce the computational load required during deployment.

23) Is further clarification required regarding any of the aspects discussed in section 3.4.2 of the working document?

Clarification is critical to address ambiguities in systemic risk assessment, transparency obligations, and downstream provider attribution, which could otherwise lead to inconsistent compliance, excessive burdens on SMEs, or regulatory loopholes. Systemic risk criteria need standardization to ensure providers like OpenAI or DeepSeek consistently mitigate high-impact risks. Transparency requirements must be precise enough to enable modifiers, especially smaller entities, to comply without proprietary data access, as for example in Llama fine-tuning. Provider attribution needs clear criteria to prevent evasion in open-source or third-party scenarios, ensuring accountability for modified models like Mixtral 8x7B. The AI Act's mechanisms, Article 51's alternative benchmarks, regulatory sandboxes (Article 57), SME support (Article 55), and the European AI Office's consultation (until May 22, 2025), provide tools to deliver these clarifications. This approach, driven by trends in efficient architectures (e.g., Mixture-of-Experts), open-source customization (e.g., Llama), and third-party services (e.g., Hugging Face), ensures the Act balances robust oversight with innovation, fostering trust and safety in a dynamic AI ecosystem.

Thank you

Thank you for participating in the consultation. Please don't forget to click on submit.

The AI Office will publish a summary of the results of the consultation. Results will be based on aggregated data and respondents will not be directly quoted.

Background Documents

GPAI_guidelines_consultation.pdf

Contact

CNECT-A3@ec.europa.eu