Contribution ID: 168512a9-e480-48b5-b561-a5a6718b70ae
Date: 08/06/2025 16:57:40

Targeted stakeholder consultation on classification of AI systems as high-risk

Fields marked with * are mandatory.

Targeted stakeholder consultation on the implementation of the AI Act's rules for high-risk AI systems

Disclaimer: This document is a working document of the AI Office for the purpose of consultation and does not prejudge the final decision that the Commission may take on the final guidelines. The responses to this consultation paper will provide important input to the Commission when preparing the guidelines.

This consultation is targeted to stakeholders of different categories. These categories include, but are not limited to, providers and deployers of (high-risk) AI systems, other industry organisations, as well as academia, other independent experts, civil society organisations, and public authorities.

The Artificial Intelligence Act (the 'AI Act')[1], which entered into force on 1 August 2024, creates a single market and harmonised rules for trustworthy and human-centric Artificial Intelligence (AI) in the EU.[2] It aims to promote innovation and uptake of AI, while ensuring a high level of protection of health, safety and fundamental rights, including democracy and the rule of law. The AI Act follows a risk-based approach classifying AI systems into different risk categories, one of which is the high-risk AI systems (Chapter III of the AI Act). The relevant obligations for those systems will be applicable two years after the entry into force of the AI Act, as from 2 August 2026.

The AI Act distinguishes between two categories of AI systems that are considered as 'high-risk' set out in Article 6(1) and 6(2) AI Act. Article 6(1) AI Act covers AI systems that are embedded as safety components in products or that themselves are products covered by Union legislation in Annex I, which could have an adverse impact on health and safety of persons. Article 6(2) AI Act covers AI systems that in view of their intended purpose are considered to pose a significant risk to health, safety or fundamental rights. The AI Act lists eight areas in which AI systems could pose such significant risk to health, safety or fundamental rights in Annex III and, within each area, lists specific use-cases that are to be classified as high-risk. Article 6(3) AI Act provides for exemptions for AI systems that are intended to be used for one of the cases listed in Annex III, but which do not pose significant risk since they fall under one of the exceptions listed in Article 6(3).

AI systems that classify as high-risk must be developed and designed to meet the requirements set out in Chapter III Section 2, in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security. Providers of high-risk AI systems must ensure that their high-risk AI system is compliant with these requirements and must themselves comply with a number of obligations set out in Chapter III Section 3, notably the obligation to put in place a quality management system and ensure that the high-risk AI system undergoes a conformity assessment prior to its being placed on the market or put into service. The AI Act also sets out obligations for deployers of high-risk AI systems, related to the correct use, human oversight, monitoring the operation of the high-risk AI system and, in certain cases, to transparency vis-à-vis affected persons.

Pursuant to Article 6(5) AI Act, the Commission is required to provide guidelines specifying the practical implementation of Article 6, which sets out the rules for high-risk classification, by 2 February 2026. It is further required that these guidelines should be accompanied with a comprehensive list of practical examples of use cases of AI systems that are high-risk and not high-risk. Moreover, pursuant to Article 96(1)(a) AI Act, the Commission is required to develop guidelines on the practical application of the requirements for high-risk AI systems and obligation for operators, including the responsibilities along the AI value chain set out in Article 25.

The purpose of the present targeted stakeholder consultation is to collect input from stakeholders on practical examples of AI systems and issues to be clarified in the Commission's **guidelines** on the classification of high-risk AI systems and future guidelines on high-risk requirements and obligations, as well as responsibilities along the AI value chain.

As not all questions may be relevant for all stakeholders, respondents may reply only to the section(s) and the questions they would like. Respondents are encouraged to provide **explanations and practical cases** as a part of their responses to support the practical usefulness of the guidelines.

The targeted consultation is available in English only and will be open for **6 weeks starting on 6 June until 18 July 2025.**

The questionnaire for this consultation is structured along 5 sections with several questions.

Regarding section 1 and 2, respondents will be asked to provide answers pursuant to the parts of the survey they expressed interest for in Question 13, whereas all participants are kindly asked to provide input for section 3, 4 and 5.
Section 1. Questions in relation to the classification rules of high-risk AI systems in Article 6(1) and the Annex I to the AI Act

- This section includes questions on the concept of a safety component and on each product category listed in Annex I of the AI Act.

Section 2. Questions in relation to the classification of high-risk AI systems in Article 6(2) and the Annex III of the AI Act. This category includes questions related to:

- AI systems in each use case under the 8 areas referred to in Annex III.
- The filter mechanism of Article 6(3) AI Act allowing to exempt certain AI systems from being

classified as high-risk under certain conditions.
- If pertinent: Need for clarification of the distinction between the classification as a high-risk AI system and AI practices that are prohibited under Article 5 AI Act (and further specified in the Commission's guidelines on prohibited AI practices[3] from 3 February 2025) and interplay of the classification with other Union legislation.

Section 3. General questions for high-risk classification. This category includes questions related to:
- The notion of intended purpose, including its interplay with general purpose AI systems.
- Cases of potential overlaps within the AI Act classification system under Annex I and III.

Section 4. Questions in relation to requirements and obligations for high-risk AI systems and value chain obligations. This category includes questions related to:
- the requirements for high-risk AI systems and obligations of providers.
- the obligations of deployers of high-risk AI systems.
- the concept of substantial modification and the value chain obligations in Article 25 AI Act.

Section 5. Questions in relation to the need for amendment of the list of high-risk use cases in Annex III and of prohibited AI practices laid down in Article 5.
- Input for the mandatory annual assessment of the need for amendment of the list of high-risk use-cases set out in Annex III
- Input for the mandatory annual assessment of the list of prohibited AI practices laid down in Article 5

All contributions to this consultation may be made publicly available. Therefore, please do not share any confidential information in your contribution. Individuals can request to have their contribution anonymised. Personal data will be anonymised.

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.

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689).
[2] Article 1(1) AI Act.
[3]https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

Information about the respondent

*First name

Alfonso

*Surname

De Gregorio

*Email address

[redacted]

*Do you represent an organisation (e.g., think tank or civil society/consumer organisation) or act in your personal capacity (e.g., independent expert or from a downstream provider)?

- ○ Organisation
- ◉ In a personal capacity

*Select your country of residence

Italy

*Which stakeholder category would you consider yourself in? If more than one category is applicable, please select the category that is best applicable in your situation / from the capacity you are responding.

Other independent expert or organisation with relevant expertise

*Sector(s) of activity

- ☑ Information technology
- ☑ Public administration
- ☑ Law enforcement
- ☐ Justice sector
- ☐ Legal services sector
- ☑ Cultural and creative sector, including media
- ☐ Healthcare
- ☐ Employment
- ☐ Education and training
- ☐ Consumer services
- ☐ Business services
- ☐ Banking and finances
- ☐ Manufacturing
- ☐ Energy
- ☐ Transport
- ☐ Telecommunications
- ☐ Retail
- ☐ E-commerce
- ☐ Advertising
- ☐ Consumer protection
- ☑ Others

*Please, specify

30 character(s) maximum

Research services

*Describe the activities of your organisation or yourself

1,300 character(s) maximum

As a researcher and cybersecurity technologist, my work focuses on the intersection of artificial intelligence and cybersecurity. I analyse the offensive and defensive capabilities of advanced AI models, particularly large language models (LLMs). My research evaluates the policy and regulatory implications of new AI release paradigms, such as open-weight models, identifying potential security risks, policy gaps in existing frameworks like the EU AI Act, and proposing technical and policy mitigations to ensure responsible innovation.

---

*All contributions to this consultation may be made publicly available. Therefore, please do not share any confidential information in your contribution. Your e-mail address will never be published. Should your contribution be anonymised in the instance that all contributions are made publicly available?

If you act in your personal capacity: *All contributions to this consultation may be made publicly available. You can choose whether you would like your details to be made public or to remain anonymous. The type of respondent that you responded to this consultation as, your answer regarding residence, and your contribution may be published as received. Your name will not be published. Please do not include any personal data in the contribution itself.*

If you represent one or more organisations: *All contributions to this consultation may be made publicly available. You can choose whether you would like respondent details to be made public or to remain anonymous. Only organisation details may be published: The type of respondent that you responded to this consultation as, the name of the organisation on whose behalf you reply as well as its size, its presence in or outside the EU and your contribution may be published as received. Your name will not be published. Please do not include any personal data in the contribution itself if you want to remain anonymous.*

- ○ Yes, please anonymise my contribution.
- ◉ No

*Do you agree that we may contact you in the event of follow-up questions or if we want to learn more about your responses?

- ◉ Yes
- ○ No

☑ I acknowledge the attached privacy statement.
Privacy_statement_high_risks.pdf (/eusurvey/files/f7c81fe9-bf9b-412b-9a2b-335acc4f56eb)

***On which part(s) of the public consultation are you interested to contribute to?** *Multiple answers are possible. Please note that selecting a particular answer will direct you to a set of questions specifically related to subject specified.*

- ☐ Questions in relation to **Annex I of the AI Act.** (Section 1)
- ☐ Questions in relation to **Annex III of the AI Act.** (Section 2)
- ☐ Questions on **horizontal aspects** of the high-risk classification. (Section 3)

- [x] Questions in relation to **requirements and obligations for high-risk AI systems and value chain obligations**. (Section 4)
- [ ] Questions in relation to the **need for possible amendments of high-risk use cases in Annex III and of prohibited practices in Article 5**. (Section 5)

Section 4 – Questions in relation to requirements and obligations for high-risk AI systems and value chain obligations

A. Requirements for high-risk AI systems

The AI Act sets mandatory requirements for high-risk AI systems as regards risk management (Article 9), data and data governance (Article 10), technical documentation (Article 11) and record-keeping (Article 12), transparency and the provision of information to deployers (Article 13), human oversight (Article 14), and robustness, accuracy and cybersecurity (Article 15).

Providers are obliged to ensure that their high-risk AI system is compliant with those requirements before it is placed on the market. Harmonised standards will play a key role to provide technical solutions to providers that can voluntarily rely on them to ensure compliance and rely on a presumption of conformity. The Commission has requested the European standardisation organisations CEN and CENELEC to develop standards in support of the AI Act. This work is currently under preparation.

Question 35. Beyond the technical standards under preparation by the European Standardisation Organisations, are there further aspects related to the AI Act's requirements for high-risk AI systems in Articles 9-15 for which you would seek clarification, for example through guidelines?

If so, please elaborate on which specific questions you would seek further clarification.
3,000 character(s) maximum

Yes, please. Clarification is urgently needed on how Articles 9-15 apply to high-risk systems derived from open-weight General-Purpose AI (GPAI) models. The current requirements appear to be designed for closed, API-gated systems where the provider maintains control. This assumption breaks down for open-weight models. Nota bene: I have written a research paper detailing the policy gaps and the technical realities of this scenario. Please find its preprint at https://arxiv.org/abs/2505.17109 – my comments below elaborate on the arguments being raised in the paper itself.

Specific clarifications needed:
1/ Cybersecurity (Article 15): The paper highlights that open-weight models can be fine-tuned by third parties to remove safety alignments or specialise them for malicious purposes (e.g., enhanced malware generation). How can the original provider of the open-weight model be held responsible for the cybersecurity of a model they no longer control? Guidelines should clarify that the responsibility for robustness against misuse in a modified system shifts to the actor performing the modification.

2/ Record-keeping (Article 12) & Transparency (Article 13): Once an open-weight model is downloaded, the original provider has no ability to log its use. Furthermore, technical transparency measures like watermarking can often be disabled or removed by the downstream user. Guidelines should acknowledge this technical reality and focus on obligations that can be met pre-release (e.g., comprehensive documentation of capabilities, known risks, and data provenance) rather than unenforceable post-release monitoring.

3/ Risk Management System (Article 9): A provider of an open-weight model cannot foresee all possible downstream modifications and applications. The guidelines should specify that the provider's risk management obligation is focused on the model as released, and that any actor who substantially modifies the model or deploys it in a high-risk context inherits the obligation to conduct a new risk assessment for that specific use case.

In essence, the guidelines must address the "mitigation gap" that arises from the loss of control inherent in open-weight distribution.

Question 36. Are there aspects related to the requirements for high-risk AI systems in Articles 9-15 which require clarification regarding their interplay with other Union legislation?

If so, please elaborate which specific aspects require clarification regarding their interplay with other Union legislation and point to concrete provisions of specific other Union law.

3,000 character(s) maximum

B. Obligations for providers of high-risk AI systems

Beyond ensuring that a high-risk AI system is compliant with the requirements in Articles 9-15, providers of high-risk AI systems have several other obligations as listed in Article 16 and further specified in other corresponding provisions of the AI Act. These include:

- *Indicate on the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, as applicable, their name, registered trade name or registered trademark, the address at which they can be contacted;*
- *Have a quality management system in place which complies with Article 17;*
- *Keep the documentation referred to in Article 18;*
- *When under their control, keep the logs automatically generated by their high-risk AI systems as referred to in Article 19;*
- *Ensure that the high-risk AI system undergoes the relevant conformity assessment procedure as referred to in Article 43;*
- *Draw up an EU declaration of conformity in accordance with Article 47;*
- *Affix the CE marking to the high-risk AI system, in accordance with Article 48;*
- *Comply with the registration obligations referred to in Article 49(1);*
- *Take the necessary corrective actions and provide information as required in Article 20;*
- *Cooperate with national competent authorities as required in Article 21;*
- *Ensure that the high-risk AI system complies with accessibility requirements in accordance with Directives (EU) 2016/2102 and (EU) 2019/882.*

Question 37. Are there aspects related to the AI Act's obligations for providers of high-risk AI systems for which you would seek clarification, for example through guidelines?

If so, please elaborate on which specific questions you would seek further clarification.
3,000 character(s) maximum

Yes, particularly concerning post-market obligations for providers of open-weight models that may be used to build high-risk systems.

1/ Corrective Actions (Article 20): This obligation is technically infeasible for a provider of a publicly released open-weight model. Once the model weights are distributed, the provider cannot recall or unilaterally apply corrective actions to copies running on third-party infrastructure. The guidelines should clarify how a provider can comply, perhaps by publicly issuing advisories and releasing patched versions, while acknowledging they cannot force an update.

2/ Quality Management System (Article 17): The requirement for a "post-market monitoring system" is fundamentally challenged. A provider cannot monitor the performance or use of a model they do not control. The guidelines should define what "post-market monitoring" means in this context. It could be redefined as monitoring the public ecosystem for reports of misuse, vulnerabilities discovered in the model, and malicious fine-tuned variants, rather than monitoring specific deployments.

3/ Registration Obligations (Article 49(1)): It should be clarified whether the original provider of a foundational open-weight GPAI model must register it if it could be adapted into a high-risk system, or if the obligation falls solely on the downstream actor who actually creates and deploys the high-risk system. The latter approach seems more practical and targeted.

The guidelines must differentiate between obligations that can be met pre-release and those that become impossible post-release for open-weight models, to avoid creating a regulatory environment that implicitly favours closed systems.

Question 38. Are there aspects related to the obligations for providers of high-risk AI systems which require clarification regarding their interplay with other Union legislation?

If so, please elaborate which specific aspects require clarification regarding their interplay with other Union legislation and point to concrete provisions of specific other Union law.

3,000 character(s) maximum

C. Obligations for deployers of high-risk AI systems

Article 3(4) defines a deployer as a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity.

Deployers of high-risk AI systems have specific responsibilities under the AI Act. Transversally, Article 26 obliges all deployers of high-risk AI systems to:

- *Take appropriate technical and organisational measures to ensure that AI systems are used in accordance with the instructions accompanying the AI systems;*
- *Assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support;*
- *Ensure that input data is relevant and sufficiently representative in view of the intended purpose of the high-risk AI system;*
- *Monitor the operation of the high-risk AI system on the basis of the instructions for use and, where relevant, inform providers in accordance with Article 72;*
- *Keep the logs automatically generated by that high-risk AI system to the extent such logs are under their control, for a period appropriate to the intended purpose of the high-risk AI system of at least six months.*

Additionally, Article 26 foresees the following obligations in specific cases:

- *For high-risk AI system at the workplace, deployers who are employers shall inform workers' representatives and the affected workers that they will be subject to the use of the high-risk AI system;*
- *Specific authorization requirements and restrictions apply to the deployer of a high-risk AI system for post-remote biometric identification for law enforcement purposes;*
- *Deployers of high-risk AI systems referred to in Annex III that make decisions or assist in making decisions related to natural persons shall inform the natural persons that they are subject to the use of the high-risk AI system.*

Question 39. Are there aspects related to the AI Act's obligations for deployers of high-risk AI systems listed in Article 26 for which you would seek clarification, for example through guidelines?

If so, please elaborate on which specific questions you would seek further clarification.

3,000 character(s) maximum

Question 40. Are there aspects related to the obligations for deployers of high-risk AI systems listed in Article 26 which require clarification regarding their interplay with other Union legislation?

If so, please elaborate which specific aspects require clarification regarding their interplay with other Union legislation and point to concrete provisions of specific other Union law.

3,000 character(s) maximum

Moreover, according to Article 27, deployers of high-risk AI systems that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an ***assessment of the impact on fundamental rights*** *that the use of such system may produce. The AI Office is currently preparing a template that should facilitate compliance with this obligation.*

Article 27 specifies that where any of its obligations are already met through the data protection impact assessment conducted pursuant to Article 35 of Regulation (EU) 2016/679 or Article 27 of Directive (EU) 2016/680, the fundamental rights impact assessment referred to in paragraph 1 of this Article shall complement that data protection impact assessment.

Question 41. Are there aspects related to the AI Act's obligations for deployers of high-risk AI systems for the fundamental rights impact assessment for which you would seek clarification in the template?

3,000 character(s) maximum

Question 42. In your view, how can complementarity of the fundamental rights impact assessment and the data protection impact assessment be ensured, while avoiding overlaps?

3,000 character(s) maximum

Finally, deployers of high-risk AI systems may have to provide an explanation to an affected person upon their request. This right is granted by Article 86 AI Act to affected persons which are subject to a decision, which is taken on the basis of the output from a high-risk AI system listed in Annex III and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights.

Question 43. Are there aspects related to the AI Act's right to request an explanation in Article 86 for which you would seek clarification, for example through guidelines?

If so, please elaborate on which specific questions you would seek further clarification.

3,000 character(s) maximum

D. Substantial modification (Article 25 (1) AI Act)

Article 3 (23) defines a substantial modification as a change to an AI system after its placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment carried out by the provider. As a result of such a change, the compliance of the AI system with the requirements for high-risk AI systems is either affected or results in a modification to the intended purpose for which the AI system has been assessed.

The concept of ‘substantial modification’ is central to the understanding of the requirement for the system to undergo a new conformity assessment. Pursuant to Article 43(4), the high-risk AI system should be considered a new AI system which should undergo a new conformity assessment in the event of a substantial modification.

This concept is also central for the understanding of the scope of obligations between a provider of a high-risk AI system and other actors operating in the value chain (distributor, importer or deployer of a high-risk AI system). Pursuant to Article 25, any distributor, importer, deployer or other third-party shall be considered to be a provider of a high-risk AI system and shall be subject to the obligations of the provider, in any of the following circumstances:

(a), they put their name or trademark on a high-risk AI system already placed on the market or put into service, without prejudice to contractual arrangements stipulating that the obligations are otherwise allocated;

(b), they make a substantial modification to a high-risk AI system that has already been placed on the market or has already been put into service in such a way that it remains a high-risk AI system;

(c), they modify the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk and has already been placed on the market or put into service in such a way that the AI system concerned becomes a high-risk AI system.

Question 44. Do you have any feedback on issues that need clarification as well as practical examples on the application of the concept of 'substantial modification' to a high-risk AI system.
3,000 character(s) maximum

Yes, the concept of 'substantial modification' is critical for the open-weight model ecosystem and requires explicit clarification.

Issue for Clarification: The act of fine-tuning an open-weight model to remove or bypass its safety features or alignment training must be explicitly defined as a "substantial modification."

Practical Example: As my research paper discusses, a general-purpose model like Llama 3 or DeepSeek-R1 is released with safety guardrails to prevent it from generating harmful content. A third party then downloads the weights and, using techniques like fine-tuning or "abliteration," creates a new, uncensored version specialised for generating sophisticated phishing emails or exploit code.

This act should unequivocally be considered a substantial modification. Consequently, under Article 25(b), the third party making this change becomes the new provider of this now-high-risk AI system and is subject to all the corresponding obligations.

Without this clarification, a significant loophole exists where malicious actors could modify open models to create dangerous tools while the responsibility remains ambiguously with the original provider, who has no control over the modification.

Article 43(4) second sentence describes the circumstances under which the change does not qualify as a substantial modification: 'For high-risk AI systems that continue to learn after being placed on the market or put into service, changes to the high-risk AI system and its performance that have been pre-determined by the provider at the moment of the initial conformity assessment and are part of the information contained in the technical documentation referred to in point 2(f) of Annex IV, shall not constitute a substantial modification.'

Question 45. Do you have any feedback on issues that need clarification as well as practical example of pre-determined changes which should not be considered as a substantial modification within the meaning the Article 43(4) of the AI Act.

3,000 character(s) maximum

E. Questions related to the value chain roles and obligations

Throughout the AI value chain, multiple parties contribute to the development of AI systems by supplying tools, services, components, or processes. These parties play a crucial role in ensuring the provider of the high-risk AI system can comply with regulatory obligations. To facilitate compliance with

regulatory obligations, Article 25(4) require these parties to provide the high-risk AI system provider with necessary information, capabilities, technical access and other assistance through written agreements, enabling them to fully meet the requirements outlined in the AI Act.

However, third parties making tools, services, or AI components available under free and open-source licenses are exempt from complying with value chain obligations. Instead, providers of free and open-source AI solutions are encouraged to adopt widely accepted documentation practices, such as model cards and datasheets, to facilitate information sharing and promote trustworthy AI.

To support cooperation along the value chain, the Commission may develop and recommend voluntary model contractual terms between providers of high-risk AI systems and third-party suppliers.

Question 46. From your organisation's perspective, can you describe the current distribution of roles in the AI value chain, including the relationships between providers, suppliers, developers, and other stakeholders that your organisation interacts with?

3,000 character(s) maximum

From a research perspective focused on open-weight models, the AI value chain has become significantly more complex and decentralized than a traditional linear model. The key roles are:

1/ Foundational Model Provider: An entity (e.g., Meta, Mistral AI, DeepSeek) that trains a large, general-purpose model and releases its weights publicly. Their direct relationship with the end-user is often non-existent after the initial download.

2/ Downstream Actor (Modifier/Fine-tuner): A diverse group ranging from academic researchers and individual developers to SMEs and startups. They take the open-weight model and adapt it for specific purposes. This actor effectively becomes a "secondary provider" if they make a substantially modified model available to others, or a "deployer" if they integrate it into a final product.

3/ Platform/Community Host: Entities like Hugging Face that host and distribute these models, creating a hub for sharing and collaboration.

The critical feature of this value chain is the one-way flow of the artifact (the model weights) without a corresponding flow of control or data back to the foundational provider.

Question 47 Do you have any feedback on potential dependencies and relationships throughout the AI value chain that should be taken into consideration when implementing the AI Act's obligations, including any upstream or downstream dependencies between providers, suppliers, developers, and other stakeholders, which might impact the allocation of obligations and responsibilities between various actors under the AI Act? In particular, indicate how these dependencies affect SMEs, including start-ups.

3,000 character(s) maximum

The key dependency is that downstream actors (including SMEs and start-ups) are entirely reliant on the safety, security, and documentation of the upstream open-weight model. However, the upstream provider has no control over how their model is used or modified downstream.

Impact on Allocation of Obligations:
This one-way dependency means that liability and obligations must be clearly allocated based on who has control at each stage.
* The foundational provider should be responsible for rigorous pre-release testing, transparency about capabilities (including dual-use potential, as seen with DeepSeek-R1's high OCCULT scores), and robust documentation.
* The downstream actor who modifies the model or deploys it in a high-risk context must assume the obligations of a "provider" or "deployer" under the Act. They are the only ones who can perform the context-specific risk assessment and ensure compliance for the final application.

Impact on SMEs and Start-ups:
This dynamic is a double-edged sword.
* Pro: SMEs can innovate rapidly and cost-effectively by building on powerful open-weight models.
* Con (Risk): An SME could inadvertently become the provider of a high-risk AI system simply by fine-tuning an open model. They may lack the resources and expertise to handle the extensive compliance obligations, which could stifle the very innovation the open approach fosters. Guidelines must be clear and pragmatic to ensure compliance is not prohibitively burdensome for smaller actors who are crucial to the ecosystem.

Question 48. What information, capabilities, technical access and other assistance do you think are necessary for providers of high-risk AI systems to comply with the obligations under the AI Act, and how should these be further specified through written agreements?

3,000 character(s) maximum

For a downstream actor to build a compliant high-risk system on top of an open-weight model, the foundational provider must supply comprehensive information at the point of release. Since written agreements are not feasible for public releases, this should take the form of standardised, public documentation.

Necessary information includes:
1/ Detailed Model Cards: Outlining the model's intended use, limitations, and ethical considerations.
2/ Dataset Transparency: Summaries of the training data, particularly highlighting the inclusion of any sensitive or specialised data (e.g., cybersecurity literature, code repositories) that could grant the model high-risk capabilities.
3/ Capability & Safety Evaluations: Results from rigorous, standardised pre-release testing. This should include evaluations for dual-use potential, such as benchmarks like MITRE's OCCULT for offensive cyber capabilities. This provides a transparent "Responsible Labelling" of known risks.
4/ Information on Safety Mechanisms: Clear documentation of any embedded safety features (e.g., via RLHF), their robustness, and known methods for bypassing them.
5/ Technical Specifications: Information on model architecture, computational requirements, and other data needed for downstream actors to ensure robustness and accuracy in their specific deployment environment.

This information is not "assistance" in an ongoing sense but a necessary "dowry" provided with the model to enable downstream responsibility.

Question 49. Please specify the challenges in the application of the value chain obligations in your organisation for compliance with the AI Act’s obligations for high-risk AI systems and the issues for which you need further clarification; please provide practical examples.

1,500 character(s) maximum

From a research and policy analysis perspective, the single greatest challenge is applying a value chain model based on control and contractual agreements to the open-weight ecosystem, which is defined by a deliberate loss of control.

The Challenge: Article 25(4) requires parties to provide assistance through "written agreements." This is impossible for a foundational model provider who releases weights publicly to millions of anonymous users. There is no agreement and no mechanism for enforcement.

Issue for Clarification: The "free and open-source" exemption is the central point of ambiguity.
* If interpreted strictly (requiring training data), most open-weight models like Llama or Mistral would not qualify, potentially burdening their providers with unenforceable downstream obligations and stifling innovation.
* If interpreted loosely ("open-weight"), it could exempt highly capable and potentially risky models from necessary pre-release diligence.

Practical Example: A developer downloads DeepSeek-R1. The original provider cannot force the developer to sign an agreement or provide them with "technical access" to their deployment. The entire framework of reciprocal obligation in the value chain collapses.

The guidelines must provide a pragmatic interpretation of the open-source exemption that focuses obligations on the party with actual control at each stage of the AI lifecycle.

Contact

Contact Form (/eusurvey/runner/contactform/AIhighrisk2025)