Feedback on the Draft Implementing Regulation on Evaluations and Enforcement Proceedings under the AI Act

April 2026

Future of Life Institute
Risto Uuk risto@futureoflife.org | Head of European Policy and Research

Introduction

The Future of Life Institute (FLI) is an independent nonprofit organisation with the goal of reducing large-scale risks and steering transformative technologies to benefit humanity, with a particular focus on artificial intelligence (AI).

This feedback responds to the Commission's draft implementing regulation establishing detailed arrangements for evaluations of general-purpose AI (GPAI) models and enforcement proceedings under Articles 92 and 101 of the AI Act.[1] [2] We welcome the Commission's initiative in putting forward a procedural framework that provides much-needed legal certainty and helps operationalise the Act in several important respects. The provisions discussed below represent meaningful clarifications, supporting the EU's capacity to scrutinise GPAI models and hold their providers accountable.

Our feedback is structured in two parts. We first identify the provisions we believe merit support, then present seven targeted recommendations for improvement. Throughout, our concern is to ensure that the procedural architecture established by this implementing regulation is robust enough to serve its purpose: enabling rigorous, independent evaluations of GPAI models and effective enforcement when providers fail to meet their obligations.

Provisions we support

Broad access powers for model evaluations

Article 2(1) of the draft clarifies the scope of the Commission's access powers set out in Article 92(3) of the AI Act. The AI Act itself refers only to access via APIs and "further appropriate technical means and tools, including source code".[3] The draft specifies that these technical means and tools include access to model weights, hosting infrastructure, system state inspection, and all levels of access granted to a provider's own employees. This is a meaningful step towards ensuring that evaluations of GPAI models are technically rigorous. It also provides legal certainty and limits the risk of protracted legal battles over the Commission's access powers.

Access to model weights and infrastructure is particularly important for detecting systemic risks that cannot be identified through API-level testing alone. This concern is borne out by the published experience of leading third-party evaluator organisations. For example, METR's preliminary evaluation of OpenAI's o3 model expressly noted that the evaluation setup was "not robust to sandbagging by the model" and that the team had "limited access to information about the model that is important for interpreting the results".[4] Consequently, deliberate underperformance by the model could not have been detected without access to its internal reasoning. Recent academic work mapping evaluator access confirms more generally that external evaluators are typically restricted to black-box, API-level interactions, and cannot examine model internals such as activations, gradients or weights. This has significant consequences for both the rigour of evaluations and the confidence stakeholders can place in their findings.[5]

1 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (OJ L, 2024/1689, 12.7.2024).

2 Draft Commission Implementing Regulation on detailed arrangements for the conduct of certain proceedings pursuant to Regulation (EU) 2024/1689, Ref. Ares(2026)2709234, 12 March 2026.

3 Article 92(3) of Regulation (EU) 2024/1689.

4 METR, "Details about METR's preliminary evaluation of OpenAI's o3 and o4-mini" (April 2025), https://evaluations.metr.org/openai-o3-report/.

5 J. Charnock et al., "Expanding External Access to Frontier AI Models for Dangerous Capability Evaluations", arXiv:2601.11916, (January 2026), https://arxiv.org/abs/2601.11916.

Accordingly, anchoring the access modalities listed above in the implementing regulation directly underpins the Commission's ability to detect the very categories of systemic risk that the AI Act is designed to address.

Prohibition on tracking of Commission evaluations

The provision in Article 2(1) empowering the Commission to require model providers to disable logging measures that could track the evaluation process is a welcome and proportionate addition. It addresses the risk that providers could monitor the Commission's testing activity and adapt model behaviour accordingly, undermining the integrity of the evaluation.

Interim measures

We strongly support the inclusion of Article 5(3), which enables the Commission to order interim measures. The provision clarifies that the Commission may act swiftly in response to serious risk, which is in line with the precautionary principle, firmly enshrined in EU law. It is also indispensable to credible product safety legislation and central to the Commission's ability to exercise its enforcement powers effectively. The importance of this power is underlined by the particular characteristics of AI-related harms, as failures from non-compliant GPAI models can propagate rapidly, affecting downstream providers and end users at scale.

Downstream provider complainant rights

Article 6(1) requires the Commission to give complainants under Article 89(2) of the AI Act the opportunity to express their views before proceedings are closed. This is consistent with established practice in EU competition law and gives greater procedural weight to downstream providers affected by GPAI model provider conduct. In the AI value chain, downstream providers bear significant costs from non-compliant or unsafe GPAI models but have limited commercial leverage to challenge upstream providers directly. This provision helps ensure that their experience informs the Commission's decisions.

Limitation periods and interruption provisions

The five-year limitation period for fining and enforcement under Articles 10 and 11 is consistent with the approach taken under the DMA, DSA, and EU competition law. We consider this framework appropriate. We also welcome the introduction of the interruption provisions - they have an important role in supporting fairness and deterrence, and enable the Commission to undertake complex regulatory investigations.

Recommendations for improvement

1. Strengthening expert independence and competence criteria

The independence criteria under Article 3(1) focus on shared ownership, governance, management, personnel, resources, and contractual relationships over the preceding 12 months. While these criteria are relevant, they are narrower and weaker than comparable frameworks used by other EU agencies that evaluate products affecting public health and safety.

The European Medicines Agency (EMA) restricts scientific committee members and experts from having employment relationships with the pharmaceutical industry within the preceding three years.[6] EMA's cooling-off provisions also cover consultancy and strategic advisory roles, regardless of whether they involve a formal contract or remuneration, including lectures organised by companies, participation in scientific

6 EMA, "European Medicines Agency policy on the handling of competing interests of scientific committees' members and experts" (EMA/54457/2024), https://www.ema.europa.eu/en/documents/other/policy-44-european-medicines-agency-policy-handling-competing-interests-scientific-committees-members-experts_en.pdf.

advisory boards, and membership of steering committees. The European Food Safety Authority (EFSA) requires experts to declare interests in financial investments, managerial roles, employment, consultancy, membership of advisory bodies, research funding, intellectual property rights, and other relevant interests over the preceding five years.[7] Both agencies publish their experts' declarations of interest.[8]

We recognise that AI is a newer and smaller field than pharmaceuticals or food safety, and that overly restrictive criteria could limit the pool of qualified evaluators. Technical experience gained at a GPAI provider can be a genuine asset for an evaluator, and the framework should be designed to welcome such experience rather than deter it. At the same time, the current criteria leave several gaps. The focus on formal "contractual relationships" omits other significant forms of involvement with the GPAI industry, including research funding, computing access arrangements, advisory board membership, and institutional dependencies. The criteria need only be "taken into account" rather than applied within a clear procedural framework, and there is no requirement for experts to declare conflicts of interest, despite the Commission's own parallel implementing regulation on the scientific panel requiring such declaration.[9]

Building on this, we make the following recommendations.

- **First, appointed evaluators should not have current ties to any GPAI model providers throughout their appointment.** This should cover the broader range of relationships described above, going beyond formal contractual or employment ties. It should also extend to the current interests of close family members, in line with established practice in comparable EU bodies.[10]

- **Second, we recommend that the framework provide for a clearly defined transitional period regarding past relationships.** During an initial phase running until 2 August 2028, individuals coming directly from GPAI model providers, or with recent ties to such providers, should be able to take up the evaluator role without restrictions tied to those past relationships. The rationale is practical: at a time when the pool of qualified candidates remains very limited, this arrangement would allow technical experts currently employed in industry to move into the evaluator community. After 2 August 2028, more stringent requirements on past relationships should apply, covering the broader scope of interests described above. By that point, the field will have had time to mature and a more independent evaluator base will be needed on a sustainable basis. Specifically, any relationship leading to a conflict of interest with any GPAI model provider in the 12 months preceding appointment should automatically exclude an expert from the evaluation; in the period between 36 and 12 months prior to appointment, only a relationship with the GPAI model provider whose model is under evaluation should be taken into account in assessing the expert's independence (i.e. not necessarily disqualify). Past relationships of close family members should not be taken into account.

- **Third, mandatory declarations of current and past interests over the past three years should be required of all appointed experts,** covering the broader range of relationships described above. Such declarations should be made publicly available and also cover the interests of close

7 EFSA, "EFSA's policy on independence" (2024), https://www.efsa.europa.eu/sites/default/files/corporate_publications/files/independence-policy-2024.pdf.

8 See EMA's policy on the handling of competing interests of scientific committees' members and experts (EMA/54457/2024); EFSA's policy on independence (2024).

9 Article 10 of Implementing Regulation (EU) 2025/454 of 7 March 2025 laying down the rules for the application of Regulation (EU) 2024/1689 of the European Parliament and of the Council as regards the establishment of a scientific panel of independent experts in the field of artificial intelligence.

10 EMA's policy on its experts' independence includes current relationships of close family members.

family members, in line with established practice in comparable EU bodies[11] and the implementing regulation on the scientific panel.[12]

- **Fourth, and irrespective of the transitional arrangements on pre-appointment relationships described above, a 12-month post-appointment restriction on taking up employment or establishing other ties with GPAI providers should apply throughout.** This would address the risk that experts produce favourable assessments to secure future employment.

- **Finally, the Commission shall revoke or terminate an expert's appointment if their independence is compromised.** Precedents exist in comparable EU bodies, such as the de-registration process with the EMA expert management system.[13] We propose including the following two provisions. First, where the Commission becomes aware that an appointed expert's independence is compromised, or is reasonably perceived to be compromised, or where the expert has failed to disclose a material conflict of interest, the Commission shall revoke the appointment and may appoint a replacement expert in accordance with Article 4. Second, an appointed expert who becomes aware of circumstances that may compromise their independence shall immediately inform the Commission and withdraw from the evaluation pending a determination by the Commission.

More broadly, the credibility of the evaluation framework set out in the implementing regulation will depend on the existence of a robust and growing community of independent evaluators. We therefore encourage the Commission to take an active role in fostering the emergence of an evaluator profession in the EU. Relevant measures could include capacity building and training programmes, dedicated funding for independent evaluation organisations, organisation of regular meetings and exchanges between evaluators to share methodologies and good practices, and the establishment of a public register of evaluators who have worked with the AI Office. Such measures would help build the talent pipeline that the AI Act's enforcement architecture requires, and reinforce the legitimacy of evaluation outcomes by anchoring them in a recognised and visible professional community.

Beyond independence, the framework would also benefit from clearer expectations on the technical profile of appointed experts. Relevant criteria should include demonstrated expertise in the evaluation of GPAI models, familiarity with red-teaming and adversarial testing methodologies, and a track record of independent technical assessment work, whether in academia, civil society, or dedicated evaluator organisations.

2. Clarifying the scope of interim measures

For interim measures to function as an effective tool protecting the public interest, the Commission must be able to act in a broad range of circumstances. Article 5(3) refers to measures "including" the prevention of a model from being made available on the market. While the use of "including" suggests that this list is non-exhaustive, the provision would benefit from explicit confirmation that interim measures apply equally to models already deployed on the market to support legal certainty.

We further recommend that the power to impose interim measures not be confined to the period before the proceedings are opened. Limiting interim measures to the pre-proceedings phase is an arbitrary constraint

11 Both EFSA and EMA publish the declarations of interest of their external experts.

12 Article 10 of Implementing Regulation (EU) 2025/454 of 7 March 2025 laying down the rules for the application of Regulation (EU) 2024/1689 of the European Parliament and of the Council as regards the establishment of a scientific panel of independent experts in the field of artificial intelligence.

13 EMA, "Expert management" (EMA/182440/2024), https://www.ema.europa.eu/system/files/documents/other/bpd-h-002-expert-management_en.pdf.

that could leave the public exposed to harm. Investigations into complex AI models are likely to be time-consuming, and the Commission must be able to act swiftly where it identifies serious risks.

3. Protecting the identity of complainants

The draft contains detailed provisions on the confidentiality of business secrets and access-to-file procedures but does not explicitly address the confidentiality of the complainant's identity.

Downstream providers complaining about a GPAI model provider under Article 89(2) are often deeply commercially dependent on that provider. They build applications, services, and products on top of a foundation model, and their business viability may depend on continued access to it. The power asymmetry is acute. A downstream provider that files a complaint risks serious prejudice to its commercial interests, including degraded model access, unfavourable pricing, or termination of service. Fear of retaliation could have chilling effect, undermining the entire information-gathering mechanism in Article 89(2) of the AI Act.

We recommend an explicit provision treating the complainant's identity as confidential information under Article 9, subject to the same protections as business secrets. Even where full anonymity may not be sustainable throughout proceedings, the default should be confidentiality, with disclosure requiring a reasoned decision balancing the interests at stake.

4. Clarifying the legal basis for periodic penalty payments

Both Article 10 of the draft implementing regulation and Article 101 of the AI Act refer to periodic penalty payments, indicating that such penalties are an available enforcement tool. However, neither instrument sets out the grounds or the procedure for imposing them, or the maximum penalty levels. This contrasts with the DSA and the DMA, both of which contain express provisions on periodic penalty payments.[14] One-off fines may be insufficient to incentivise prompt compliance, particularly where non-compliance is ongoing. Periodic penalty payments are an important tool for encouraging providers to remedy identified shortcomings without delay. We note that the Commission rationalises the newly introduced limitation periods on grounds of legal certainty. The same rationale applies to periodic penalty payments. We recommend that the Commission address this gap in the implementing regulation.

5. Transparency of evaluation outcomes and civil society participation

The draft regulation contains no provisions on publication of evaluation outcomes, summary enforcement decisions, or public reporting on the status of proceedings. The confidentiality framework under Article 9 is oriented toward protecting providers' interests and *inter partes* disclosure, not toward informing the public about how GPAI models are being scrutinised.

Public trust in AI governance depends on the availability of credible information about whether and how the most capable models have been evaluated and what was found. Information asymmetries between providers and regulators, and between regulators and the public, are significant. Without a transparency mechanism, it will be difficult for civil society, the media, or the European Parliament to exercise meaningful oversight of the Commission's enforcement activity under the AI Act. When regulators and safety researchers become dependent on the same concentrated set of firms for data, compute, and technical expertise, the independence necessary for effective oversight is further compromised.[15]

14 See Article 76 of Regulation (EU) 2022/2065 (Digital Services Act) and Article 31 of Regulation (EU) 2022/1925 (Digital Markets Act), both of which contain express provisions on periodic penalty payments.

15 M. Whittaker, "The Steep Cost of Capture" (November 2021), https://dl.acm.org/doi/10.1145/3488666.

We recommend two measures. First, the Commission should be required to publish summary evaluation findings upon conclusion of evaluations, appropriately redacted to protect legitimate trade secrets but sufficiently detailed to inform the public about the nature and outcome of the assessment. Second, a formal mechanism should be established enabling civil society organisations, independent researchers, and affected downstream actors to submit evidence or observations during enforcement proceedings. Comparable mechanisms exist in EU competition law, where interested third parties may submit observations during proceedings.[16] Introducing an equivalent in the AI Act enforcement context would improve the quality of the Commission's decision-making and strengthen the legitimacy of its enforcement activity.

6. Ensuring adequate evaluation duration

The draft regulation does not address the duration of the evaluation period, leaving this entirely to the Commission's discretion in individual access decisions. We recommend that the implementing regulation provide for evaluation periods that are commensurate with the complexity and risk profile of the model under examination. Longer duration should in particular be envisaged when the model is classified as presenting systemic risk or when the independent experts appointed under Article 3 indicate that additional time is required. The Commission should also retain the ability to extend the evaluation period at any stage where this is necessary to ensure the thoroughness and integrity of the assessment.

This recommendation reflects the broadly shared view among third-party evaluator organisations that the thoroughness of an evaluation is closely correlated with the duration of access.[17] The challenge is compounded by the emerging capabilities of GPAI models, which can require evaluators to design and execute more sophisticated techniques, including techniques that account for the possibility that models may perform differently when they detect that they are being tested. Current industry practice falls short of what meaningful pre-deployment evaluation of the most capable models requires, and the implementing regulation provides an opportunity to address this gap.

7. Deeper access for model evaluations

While the draft's enumeration of access modalities is welcome, meaningful evaluation of GPAI models for systemic risk also requires access to the model in configurations beyond the version prepared for deployment. The established pre-deployment safety evaluation practice distinguishes between assessments aimed at understanding the upper bound of a model's underlying capabilities before safety mitigations are applied, and assessments of the model's residual risks with its safeguards in place.[18] Both perspectives are necessary to form a credible view of the risks associated with the most capable models, and an evaluation limited to the deployed configuration alone may mask underlying capabilities directly relevant to systemic risk assessment.

We therefore recommend that the implementing regulation make clear that the Commission's access powers extend both to versions of the model that allow evaluators to assess underlying capabilities before the application of safety mitigations, and to the version of the model with guardrails in place. The specific configurations to be made available will depend on the architecture and development pipeline of the model under examination and could usefully be further developed in technical guidance.

16 See, by analogy, Article 27(3) of Council Regulation (EC) No 1/2003 on the implementation of rules on competition, which allows interested third parties to submit observations during proceedings.

17 See METR, "Details about METR's preliminary evaluation of OpenAI's o3 and o4-mini" (April 2025), https://evaluations.metr.org/openai-o3-report/; NIST and UK AISI, "US AISI and UK AISI Joint Pre-Deployment Test" (December 2024), https://www.nist.gov/system/files/documents/2024/12/18/US_UK_AI%20Safety%20Institute_%20December_Publication-OpenAIo1.pdf.

18 Frontier Model Forum, "Issue Brief: Preliminary Taxonomy of Pre-Deployment Frontier AI Safety Evaluations" (December 2024), https://www.frontiermodelforum.org/updates/issue-brief-preliminary-taxonomy-of-pre-deployment-frontier-ai-safety-evaluations/.

Conclusion

The draft implementing regulation represents a welcome and, in several respects, ambitious step in operationalising the AI Act's enforcement framework. We strongly support the proposals regarding access powers, evaluation integrity, and interim measures. Our recommendations are intended to further strengthen the framework, in particular by ensuring that expert independence is credibly guaranteed, that enforcement tools are effective in practice, and that the public has meaningful visibility into how the most consequential AI systems are governed.

Fostering a robust and transparent enforcement framework enhances the credibility of the AI Act and supports the broader goal of ensuring that advanced AI models are developed and deployed safely and beneficially. We remain available to share our perspective as these discussions progress.

For further information, please contact:

Risto Uuk
Head of European Policy and Research
risto@futureoflife.org

Transparency register number: 787064543128-10

Feedback on the Draft Implementing Regulation on Evaluations and Enforcement Proceedings under the AI Act

Future of Life Institute