# Position Paper on v2 of the draft GPAI Code

As providers of general-purpose AI (GPAI) models, we appreciate the complexity of the Code of Practice (the Code) negotiations which is subject to many different views, while needing to ensure it is fit for purpose to enable GPAI providers' compliance with the AI Act. The companies endorsing this letter present their shared feedback with the understanding that individual companies' views may not be exhaustively expressed in these comments. We appreciate the consideration given by the working group chairs and vice chairs to our concerns expressed after the publication of the first draft of the Code and recognize some clarifications in the second draft of the Code.

At the same time, a considerable number of commitments, measures and KPIs continue to be of significant concern, as they risk restricting the development of emerging best practices while creating an overly complex and burdensome framework for compliance. First, there are still provisions that extend beyond the scope of the AI Act, in particular on copyright, transparency, audits, and the mandatory involvement of third parties in risk assessments. Second, the scale, cadence, and redundancy of documentation and reporting requirements are concerning and in several cases do not support AI safety, especially given the rapid pace of industry development and the increasing number of models that will meet the 10^25 threshold. Third, the introduction of measures that are not technically feasible or risk disclosing sensitive data poses significant challenges. These issues are particularly evident in areas such as model parameters, reporting frequency and public transparency. We call for the Code to more adequately consider legitimate concerns regarding intellectual property, trade secrets and privacy.

Greater clarity is also needed regarding the distinction between measures clearly falling within the scope of the AI Act that count towards demonstrating legal compliance under the Code, and additional measure that signatories would be encouraged to take but that go beyond what's required under the AI Act. The draft moreover creates uncertainty by expecting signatories to commit to following future AI Office (AIO) guidance and templates that have yet to be developed and preambles that would integrate non-binding recitals and other explanatory text in signatories' commitment.

The Code should serve as a proportionate and workable instrument aimed at facilitating effective compliance with what the co-legislators have enshrined within the text of the AI Act, thereby promoting trust and safeguarding AI innovation in Europe

We are at a crucial juncture in the development of the Code, as we understand the upcoming third draft is expected to closely resemble the final version. We strongly urge the chairs, vicechairs, and the AIO to consider our shared recommendations below, ensuring the principles of proportionality and feasibility are better reflected in the third version, allowing the Code to be broadly adopted by GPAI model providers. We look forward to collaborating on this effort to ensure its success.

# Annex

## Transparency

Current obligations set out in Commitment 1.1 related to the disclosure of model parameters, data usage, computational resources, energy reporting, model release dates and distribution methods,

and "model reach" documentation are overly prescriptive as they go far beyond the legal text and risk revealing sensitive information that could undermine innovation and safety.

Overly detailed technical documentation expectations raise significant concerns around confidentiality, trade secret protection, information hazards, and impact on innovation. We recommend removing and aligning with the AI Act's legal text, disclosure of information that:

- Constitutes trade secrets, e.g., model parameters; description of how the model architecture departs from standard practices; information about "each step or stage" of model training. Instead of requiring detailed parameter counts, the use of predefined model categorization bands could provide a more practical approach. Reporting on data provenance should avoid prescriptive requirements on data collection, processing, and licensing, focusing instead on high-level compliance mechanisms rather than exhaustive details about harmful content mitigation methods. Testing processes and test results are likely to include confidential information including trade secrets. Rather than a description of all tests and test results, providers could instead produce a general statement explaining any testing carried out on the model.

- Clearly goes beyond the AI Act and poses risks of misinterpretation: Given a lack of standardised methods for tracking and reporting (e.g., computational resources for model inference; energy use), relevant reporting should rely on industry-supported estimate bands rather than specifying the number and type of hardware units, as this adds an unnecessary burden without improving transparency. Similarly, several of the energy reporting obligations, in addition to going beyond the AI Act, exceed reasonable standards and lack agreed-upon methodologies such as specifying hardware ownership, location, energy costs and emissions.

- Raises cybersecurity and privacy risks: The proposed category on how human created training data is sourced, would impose new and prescriptive reporting requirements that are likely not practically feasible. Signatories should not be required to provide location in this context, especially with regard to data centers, as disclosure of such information would present serious cyber/physical security implications and create data privacy risks.

- Risk confusing downstream AI system providers, therefore disrupting innovation (e.g., specifying a list of allowed types of high-risk systems or "restricted tasks" at the model layer).

Moreover, the new "model reach" documentation requirement, particularly for GPAI models with systemic risk, conflates user volume with model-specific risk and should be removed to prevent inappropriate assessments.

Overall, transparency measures should be streamlined to focus on what is directly necessary for evaluating model safety and compliance with the AI Act as listed in Annexes XI and XII, ensuring obligations remain proportionate and consider the distinct roles of the AIO, NCAs, and downstream providers in the value chain, thereby preventing disproportionate burdens on model providers that could stifle innovation.

## Copyright

The copyright section continues to raise serious concerns. First, the draft exceeds the scope of the AI Act by granting the AIO the authority to interpret EU copyright law (which should only lie with the

judiciary) and confers the AIO status of copyright regulator. Second, ambiguities in the territorial scope of certain provisions (e.g. lawful access, upstream compliance) may create the perception that signatories agree to the Code having an extraterritorial effect and the AIO applying EU copyright law to situations otherwise not subject to EU law. Third, copyright provisions grant the AI Office the authority of a copyright regulator able to interpret copyright law (a power reserved for the judiciary) beyond the mandate of the AI Act and create problematic ambiguity as to potential extraterritorial effects.

The preamble presents significant challenges, as it combines existing legal requirements and interpretations in a way that seeks to bind signatories to obligations that are not clearly grounded in legislative provisions. We recommend removing a preamble with no legal value and no real utility which creates uncertainty.

Measure 2.3, 2.4 and KPI 2.4.1, which would require signatories to proactively demonstrate proof of compliance, should be removed. This requirement would be inconsistent with other regulatory frameworks, which generally rely on policies being in place and compliance being enforced as necessary, rather than imposing an upfront burden of proof. Furthermore, it reverses the burden of proof in relation to copyright law, potentially placing developers in a position where they must prove their systems do not infringe copyright law, rather than requiring any allegations of infringement to be substantiated. Such a reversal would undermine fundamental legal principles and stifle innovation.

Additionally, Measure 2.3 introduces strict upstream compliance requirements that exceed both the AIA and copyright law, imposing legal uncertainty with unclear review standards and burdensome obligations for private and public datasets, and should therefore be removed from the Code. The provision fails to recognize that information made available in relation to publicly available datasets may be limited and/or there may be no mechanism to request information or seek assurances from dataset providers. Similarly, the lawful access requirement (Measure 2.4) introduces ambiguity regarding 'lawful access' and should also be eliminated or simply encouraged. Also, both measures 2.3 and 2.4 deviate from the copyright regulatory framework by reversing the burden of proof.

Measure 2.5, which would prohibit training on known piracy websites, should be narrowed down and simplified to ensure legal alignment, as it exceeds the Act's intended scope. While we agree that AI providers should take reasonable and proportionate measures to not illegitimately train on pirated content, the measure doesn't allow for cases where training on websites making available copyright-infringing content pursues legitimate aims (e.g. to facilitate future detection of piracy websites). Additionally, the broad reference to (potentially unreliable) exclusion lists is particularly unclear

Measures related to third-party opt-outs, (Measure 2.7) remain extremely problematic and should be removed. It is unclear which 'widely used standards' providers will be required to comply with. In particular, "work-level" reservations are impractical due to the lack of an authoritative rights ownership source. Again, instead of “best efforts”, the Code should require “reasonable efforts”.

Similarly, the requirement to prevent model "overfitting", in order to prevent outputs too similar or identical to training data (Measure 2.9), is based on a vague concept and incorrectly links overfitting to copyright infringement going beyond the scope of the current legislative acquis (i.e. Copyright Directive and AI Act). Overfitting is sometimes desirable, and potential risks are better mitigated downstream.

Finally, we question the value of a dedicated "point of contact" (measure 2.11) for copyright complaints, given the number of rightsholders, the existence of contact points for other EU regulations, and the lack of a similar AIA requirement for copyright. In addition, the complaints process resembles a "take-down request" for training data. Reservations should be expressed in machine-readable form at the point of access – complaints cannot substitute for these reservations, as there is no way to remove data from already-trained models.

## Risk assessment and risk mitigation

### Risk taxonomy

While the risk taxonomy (commitment 3) has improved by replacing "Persuasion and Manipulation" with "Large-scale, harmful manipulation", and removing "misinformation" and "homogenization of knowledge" in Measure 3.2, we remain concerned by the inclusion of "large-scale discrimination" as a systemic risk, since this is not specific to a model's high-impact capabilities and is contextual and/or heavily influenced by system-level deployment decisions—and therefore especially difficult to measure at the model layer. As suggested in the open questions raised by the Chairs at the end of section 3 of the feedback form, we believe "large scale discrimination" should be removed from Measure 3.2.

We believe that measures 3.3 and 3.4 identify several sociotechnical factors that are difficult to evaluate at the model level as they are typically associated with functionality and usability enhancements that emerge once a model is integrated into a system. We suggest they should be clearly labelled as voluntary compared to section 3.2, in line with the language of commitment that references this section.

### Post-deployment measures

While we support references to privacy rights in downstream monitoring (Measure 10.12), the draft still lacks clarity regarding the interplay between Privacy-Enhancing Technologies (PETs) and data access expectations in enterprise contexts where model developers cannot access customer data directly. The current reference to downstream logging should therefore be removed, as it poses privacy risks and is impractical to implement. Providers should instead have the flexibility to offer mitigation measures, such as customer-side scanning tools, without intrusive monitoring obligations.

It is critical to clarify that the Code does not introduce general monitoring requirements, in line with Article 8 of the Digital Services Act (DSA). The introduction of such obligations in the Code would create unnecessary burdens that were not envisioned by the Act. A balanced framework is necessary—one that supports effective risk mitigation while avoiding overly prescriptive measures that could undermine innovation and create legal uncertainty.

The commitment to re-run evaluations every six months (Measure 6.4 and related KPIs) is overly burdensome and should be replaced with evaluations tied to material changes rather than arbitrary timelines, preventing the diversion of internal safety resources from more impactful measures.

After retirement (Measure 6.5), providers that do not deploy their own models have no way of engaging in ongoing monitoring, and requiring them to do so raises privacy, confidentiality and trade secrets concerns. This is even more challenging for open source.

**Additional safety and security measures**

During training (Measure 6.2), parts of this measure seem overly prescriptive, such as the obligations related to potential model capability. The Code should be more flexible to allow providers to evaluate and mitigate risk, generally, taking into consideration factors, such as potential model capability, rather than having a prescriptive requirement for evidence and pre-defined milestones that may or may not be relevant in a particular situation.

Before deployment (Measure 6.3), this measure does not seem necessary to place on providers if they already are addressing systemic risk in the pre-training and training phase. It is not clear what is different between the pre-deployment phase and the pre-training and fine-tuning phase.

We have a significant concern regarding Measure 10.5, which exceeds the obligations outlined in the AI Act for providers and infringes upon contractual freedoms. Liabilities for downstream activities were extensively discussed during the legislative process and were ultimately rejected by the co-legislators. Providers should not be held responsible for enforcing licensing terms to ensure third parties evaluate AI systems incorporating the provider's model because enforcing licensing terms in B2B contexts is both burdensome and impractical. Additionally, in the context of open-source, such a licensing requirement contradicts open-source principles and would be impossible to comply with. It would be disproportionate to impose rules on GPAI providers deploying the model internally, as all deployers are already subject to their own obligations under the AI Act. The Code should remain focused on providers rather than addressing deployment-related issues. This measure therefore should be either removed or made voluntary.

The mandatory sharing of tools and best practices for risk assessment (Measure 10.8 and related KPIs) also goes beyond the AIA's scope and raises concerns over IP and competition, necessitating its removal.

Commitment 12, which addresses technical risk mitigations, introduces an overly prescriptive approach to cybersecurity and must be revised significantly. It advocates for untested controls that could place a disproportionate burden on organizations and conflicts with existing European AI security standards. This commitment is not well-suited for the open-source ecosystem. Specifically, Measures 12.3 (protection of stored model weights and related assets) and 12.4 (interfaces and access control to model weights) should be removed entirely, as existing regulatory regimes already address intellectual property protections, including model weights. Where these regimes are insufficient, they should be updated rather than compensated for with duplicative obligations. Disclosing the locations of where model weights are stored can also present serious cyber/physical security implications and create data privacy risks. This section also includes unclear concepts, such as "algorithmic insights" and "non-controlled devices," and overly prescriptive requirements, such as mandating that model weights be stored on dedicated devices rather than focusing on outcome-based security goals. Future requirements, like the use of confidential computing "once available and practical," create uncertainty and should be avoided unless concrete guidelines are established.

## Governance

The scope of commitments related to adherence assessments, third-party risk evaluations, documentation, and public transparency introduces overly prescriptive and burdensome obligations that exceed the intent of the AI Act and undermine safety efforts and operational efficiency.

Commitment 14 and related KPIs on the allocation of responsibility for systemic risk is overly prescriptive and cumbersome. Companies should have flexibility to decide how they allocate responsibilities internally. The Code should not mandate internal risk management approaches including reporting lines and corporate audit functions.

Commitment 15 on framework adherence assessments should be removed due to the unclear benefit relative to its burden. The production of the SSR will already require extensive documentation and reporting in accordance with the SSF and should be sufficient for the AI Office and the AI Board to assess compliance, pursuant to article 56.6. The assessments outlined in Measure 15- which resemble compliance audits- divert critical resources from actual risk mitigation efforts. Additionally, the six-month assessment cadence (KPI 15.1) should be reduced to an annual frequency, and KPIs 15.4 and 15.5 should be removed entirely.

This second draft requires signatories to provide grey-box and white-box access (Measure 10.9,) as well as non-restrictive access (Measure 16.2) to third-party evaluators. We are concerned that these requirements could introduce significant security vulnerabilities and potentially conflict with the security provisions outlined in Measure 12, without effectively advancing the safety objectives of the Code. These measures remain a red line for us due to their potential impact on security and intellectual property protection. External access should be limited to API-based methods to mitigate security risks, and public reporting of results should remain at the discretion of the provider.

We recommend that Measure 16 is aligned with the legal scope of the AI Act. This means it should reflect that in measure 16.1 external assessment is only optional and should be encouraged when both criteria are met: "the Signatory has insufficient relevant internal expertise or information to effectively conduct the risk assessment (such as those assessments requiring sensitive national security-relevant information)" and "the provider cannot provide sufficient evidence that the model does not pose additional risk beyond that of general-purpose AI models with systemic risk already on the EU market, for example due to novel model capabilities or differences in implemented mitigations". Additionally, we suggest qualifying 'additional risk' with a materiality threshold, such as 'substantial', to ensure that external assessment is warranted. Finally, the measure should clarify that a qualified third-party evaluator must have demonstrated relevant expertise, maintain robust security and confidentiality measures, implement proper safeguards, show operational feasibility, demonstrate genuine impartiality and objectivity, and respect the need to protect trade secrets and proprietary information. Clarifications are also needed to limit model access to API-based methods to reduce safety risks, and external assessors should not have the ability to publicly disclose results, as this could disincentivize necessary cooperation.

GPAI model providers should commit to considering external reports but not be required to report disagreements. We do not think it should be necessary to document a justification of external assessors in the Model Report, but GPAI model providers should provide such a justification upon request from the AI Office. This maintains the same incentives, while minimizing documentation compliance burden.

Measure 16.2 is not a requirement of the AIA and does not take into consideration the significant burden of reporting updates to external evaluations over the lifespan of a model. There are cases where GPAI model providers may work directly with third party evaluators on post-deployment testing, but these are not necessarily grouped in batches or on a regular cadence. . We believe GPAI model providers should be required to do ongoing evaluations appropriate for model capability level, however there should not be onerous reporting requirements. Measure 16.2 also refers to 'open weight' and 'open source' when neither are defined in the AI Act, nor is there any generally agreed definition of what open source means in the context of AI which creates significant legal uncertainty. Additionally, this measure conflicts with Measure 19.4 that states providers only need to provide a model report upon market entry.

Commitment 17 should be modified to clarify that signatories "commit to considering future guidance and templates" and are not bound to "following" guidance and templates from the AI Office that do not yet exist and cannot be reviewed. Further, Commitment 17 should be modified to clarify that the commitment requires allocating sufficient resources "to investigate a reasonable suspicion of their model's involvement in a serious incident".

Commitment 20, which mandates documentation of adherence to the Code, should be removed entirely due to its duplication of existing documentation requirements under Commitment 1 and its reference to undefined future templates. It is also problematic that signatories could only respond to requests for documentation by the AI Office within a "reasonable period of time", which is not defined. Measure 20.1, which ties documentation to the size of a provider's user base, should also be removed, as it uses "model reach" as a risk proxy rather than the actual risk profile of the AI model, potentially disadvantaging smaller providers working with high-risk models.

Finally, aspects of Commitment 21 on public transparency remain outside the legal scope of the AI Act and presents confidentiality and security concerns. While some flexibility has been introduced, such as redacting sensitive information, the practical benefits of some of the transparency requirements remain unclear. Aspects without tangible safety benefits should be removed.

In summary, these commitments introduce unnecessary complexity and obligations downstream conflict with existing regulations and best practices, diverting focus from meaningful risk management. Revisions should prioritize efficiency, legal alignment, and targeted risk mitigation rather than broad compliance burdens.