Centrum für Europäische Politik
FREIBURG | BERLIN

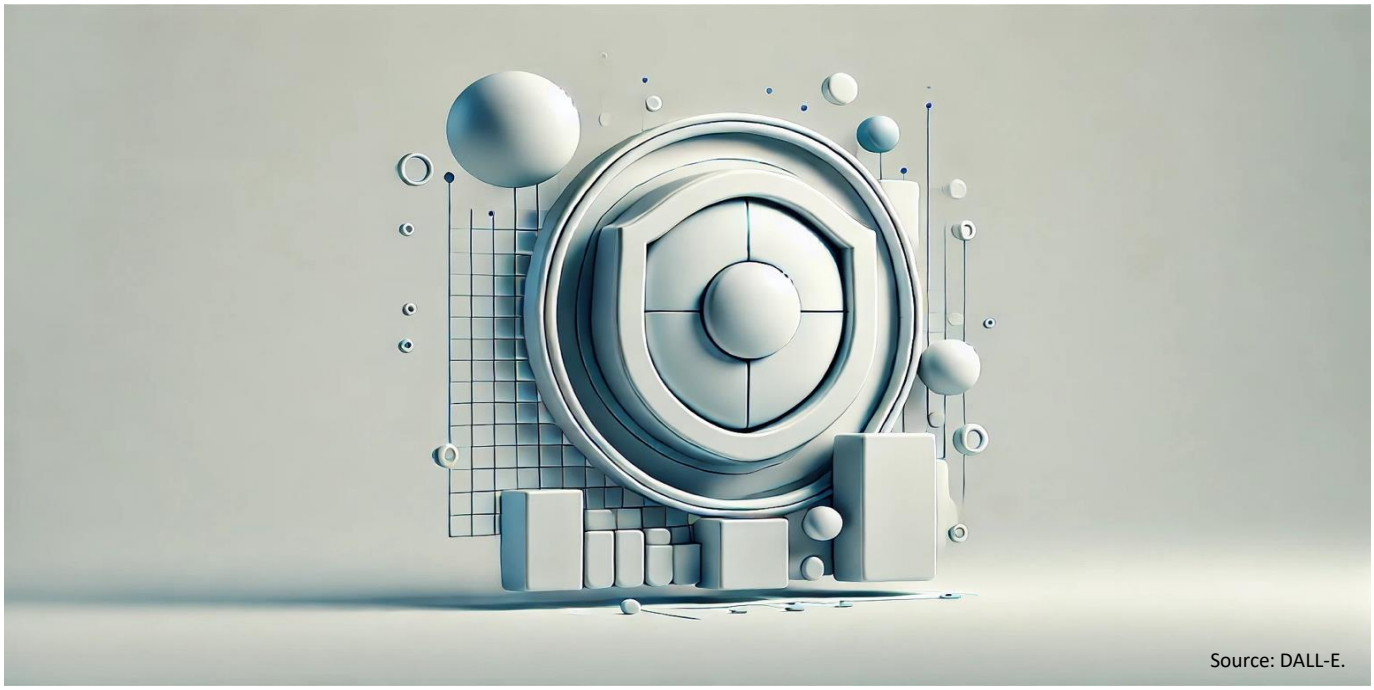
cepAdhoc

No 8 | 2025 31 March 2025

The Third Time Is the Charm?

An Ordoliberal Perspective on the Third Draft of the GPAI Code of Practice

Anselm Küsters



Source: DALL-E.

The third draft of the General Purpose AI Code of Practice promises a more targeted approach focused on large upstream AI providers, but has significantly weakened its stance on systemic risks. The forthcoming final draft must address this gap to ensure a level playing field for all market participants. In particular, notable uncertainties remain around the obligations of smaller downstream actors, the scope of compliance obligations and external assessments, as well as whistleblower protections.

- Streamlined requirements, clearer exemptions for SMEs, and a cohesive structure suggest the Code is moving toward a framework that facilitates responsible AI deployment without imposing disproportionate burdens on smaller businesses. However, a reduced risk taxonomy and ambiguities concerning downstream actors could inadvertently expose startups and smaller firms to unpredictable liabilities.
- The draft's weakening of whistleblower protections potentially limits regulators' access to essential information and might undermine consumer confidence. Moreover, the removal of mandatory six-month model re-evaluations and post-withdrawal monitoring, replaced by a subjective trigger-based approach, introduces governance gaps that could delay detection of emerging AI risks related to time-test compute.
- Intensified US lobbying against European digital regulations and the recent global policy shift toward more permissive AI governance following the Paris AI Summit may have influenced the current draft's approach of merely codifying existing industry practices among major technology firms. However, regulatory frameworks should proactively safeguard fundamental rights and anticipate emerging threats. In its current form, the Code risks missing the opportunity to meaningfully elevate AI governance toward higher standards.

Content

1 Introduction

This feedback document offers an overarching assessment of the third draft of the General-Purpose AI (GPAI) Code of Practice (CoP).[1] In line with the Chairs' and Vice-Chairs' request for structured, concise submissions, it is limited to five pages. The analysis highlights elements of the Code from an ordoliberal perspective,[2] emphasizing implications for small and medium-sized enterprises (SMEs). By examining how the Code addresses risk assessments for systemic risk (working group 2), this feedback aims to determine whether the proposed measures maintain a balance between safeguarding the public interest and fostering an environment conducive to competition and innovation. The main goal is to ensure that the final iteration of the Code not only supports responsible AI development but also promotes equitable market conditions for all actors, including smaller and less-resourced organizations.

2 Key observations on the third draft

2.1 Global context and general structure

The third draft of the GPAI Code of Practice emerges against a backdrop of heightened international scrutiny and evolving regulatory priorities. In particular, intensified US lobbying against European digital regulation, often framed as imposing "tariff-equivalent" barriers, has coincided with a global policy shift toward more permissive AI governance following the Paris AI Action Summit. These pressures might help explain why the current draft largely codifies existing industry practices among Big Tech firms. As highlighted in the explanatory notes themselves, the Code's Safety and Security Section translates the AI Act's obligations into a framework structure already in widespread use by leading AI firms, including Anthropic, OpenAI, Google DeepMind, Meta, Cohere, Microsoft, Amazon, xAI, and NVIDIA. However, the underlying purpose of regulatory frameworks is not merely to codify current industry norms but also to proactively safeguard fundamental rights and anticipate emerging threats. While minimizing compliance burdens, the current structure might thus undermine the Code's potential to meaningfully shape the trajectory of AI governance toward higher standards and greater protection.

Compared to the previous draft, the current CoP is a more streamlined document with enhanced clarity around the distribution of obligations, notably targeting larger upstream providers of high-impact AI models. The division of the Code's 18 commitments – two applying universally to all GPAI providers, and the remaining sixteen concerning only those models classified as posing systemic risks – aims to minimize undue burdens on the broader European AI ecosystem, especially SMEs. Concentrating compliance requirements at the source (i.e., large AI developers) aligns with the ordoliberal principle of maintaining fair market structures without placing excessive burdens on smaller entities. By ensuring that significant obligations remain "upstream," the Code helps to safeguard the competitive potential of SMEs that depend on accessing and integrating advanced AI models. However, some recent changes in the third draft, discussed below, together with the softened commitments on copyright and whistleblowing, mean that the current iteration has notably moved in favor of large AI companies.

---

[1] The current draft can be downloaded here: https://digital-strategy.ec.europa.eu/en/library/third-draft-general-purpose-ai-code-practice-published-written-independent-experts. An Explainer document and an FAQ about the Safety and the Security Section can be accessed on this webpage: https://code-of-practice.ai/?section=safety-security.

[2] Ordoliberalism advocates for a robust legal and institutional framework that ensures competition and social responsibility. Rather than imposing overly prescriptive controls, ordoliberal principles seek to establish fair and predictable conditions in which markets can thrive. See: Nils Goldschmidt and Michael Wohlgemuth, eds., *Grundtexte zur Freiburger Tradition der Ordnungsökonomik*, Untersuchungen zur Ordnungstheorie und Ordnungspolitik 50 (Tübingen: Mohr Siebeck, 2008).

2.2 Ambiguities in downstream obligations

While the third draft attempts to clarify downstream obligations for modified models, significant ambiguities persist regarding the circumstances under which SMEs that fine-tune or modify existing GPAI models become classified as providers with full compliance responsibilities. This issue remains critical, given the considerable economic implications for smaller enterprises, as explicitly acknowledged in the Commission's updated online Q&A. The current explanatory notes by the Vice Chairs of Working Groups 2-4 suggest the drafters anticipate that fine-tuning of existing GPAI systems generally falls outside the immediate scope of GPAI-specific obligations under the AI Act. Alternatively, even if such fine-tuned models fall within the scope, the recommendation is not to enforce full provider obligations unless evidence emerges that fine-tuning introduces novel and unacceptable risks. Although this provides apparent relief for SMEs, it introduces further legal uncertainty until official guidance from the European AI Office becomes available. From an ordoliberal perspective, it would be prudent for the European AI Office's guidance to define concrete thresholds, such as the extent of dataset alteration or degree of model parameter changes, that would trigger provider-level obligations.

A similar unresolved area pertains to large and significant GPAI models made available before finalization of the Code. At the moment, it seems that providers of these legacy models are not retrospectively bound by new transparency requirements. Consequently, developers who fine-tune or otherwise adapt pre-Code models may find themselves without essential documentation of training data or copyright policies. From an ordoliberal perspective, this presents two challenges. First, it creates informational asymmetries that may impede downstream market entry and competition. Second, it potentially distorts competition if certain firms can use "grandfathered" models without incurring the same compliance obligations as newly regulated entities. In this context, Measure II.4.2. on "safely derived models" will be helpful[3] but must be linked with forthcoming AI Office guidance on fine-tuning models.

2.3 Loopholes in external assessment requirements

Recital 114 AI Act states that providers of models with systemic risks should conduct model evaluations before market placement "as appropriate, through internal or independent external testing". The third draft seeks to clarify the conditions under which independent third-party assessments are mandated, notably incentivizing Signatories to obtaining independent external systemic risk assessments throughout the model lifecycle (Measures II.11.1 and II.11.2). However, several loopholes remain that allow companies to circumvent external scrutiny. In particular, Measure II.11.1 provides multiple pathways whereby a developer of GPAI with systemic risk may avoid external evaluation, notably through classification as a "safely derived model" or by establishing "similar safety" relative to previously assessed models. These exemptions rely heavily on companies' *internal* determinations of similarity and absence of new risks. Further complicating matters, the draft permits companies unable to identify qualified external assessors – whether due to immaturity of the external assessment ecosystem, or the refusal of assessors to adequately protect commercial confidentiality – to proceed without external evaluations. Although such providers must formally factor increased uncertainty and potential risk into their internal assessments, the absence of external oversight may incentivize under-compliance. From an ordoliberal perspective, the broad conditions under which external evaluations can be circumvented create vulnerabilities that could distort competition by favoring incumbents with a lot of internal expertise. External assessments serve as a crucial market-correcting mechanism, counteracting

[3] Safely derived models are distilled, quantized, fine-tuned, or post-trained from a safe originator model.

both information asymmetries and dominant market positions. To strengthen compliance, the final version of the Code should minimize exceptions to external assessments.

2.4 "Selected systemic risks" vs. "other risks"

The third draft presents an abbreviated risk taxonomy based on a distinction between "selected systemic risks" and "other risks". While the AI Act lists about 40 different examples of potential systemic risks, the drafters of the Safety and Security Section interpret the AI Act's definition of "systemic risk" as unequivocally applying only to four categories: (1) cyber offence risk; (2) chemical, biological, radiological, and nuclear (CBRN) risk; (3) harmful manipulation risk; and (4) loss of control risk. These four categories are listed in Appendix 1.1 and are treated as selected types of systemic risk for the purpose of the systemic risk selection in Measure II.3.1. Consequently, companies adhering to the Safety and Security Section of the Code are mandated to systematically assess and potentially mitigate these selected systemic risks. For other reasonably foreseeable systemic risks outside these four categories, which are listed in Appendix 1.2 under the vaguely termed header "Other types of risks for potential consideration in the selection of systemic risks", companies are merely encouraged to consider them within their systemic risk selection processes. Actually addressing these risks seems optional, given the wording ("The following are treated as other types of risks, from which systemic risks may arise").[4] Note, the list in Appendix 1.2 includes risks to public health, risks to fundamental rights, as well as risk from child sexual abuse material (CSAM) and non-consensual intimate images (NCII).

On a general level, this change follows our previous cep recommendation to focus only on those risks that can be, at least to some extent, traced and evaluated.[5] Still, the current division risks leaving some significant threats unaddressed, particularly those that may not yet be fully recognized but could emerge as technology evolves. In the current climate, where calls for deregulation are growing and digital regulations risk being perceived as barriers to trade, there is little incentive for companies to address optional risk categories as the resulting reputational gain is minimal. From an ordoliberal perspective, transparent risk categorization must avoid creating a two-tier system in which certain hazards remain effectively unregulated. This also clashes with the Precautionary Principle as laid down in Article 191 TFEU. The current distinction between selected systemic risks in Appendix 1.1 and other risks in Appendix 1.2 is probably intended as a practical compromise to streamline compliance by focusing primarily on risks that can be clearly identified and empirically evaluated. However, prioritizing only easily measurable risks inadvertently creates a hierarchy that may not align well with a holistic fundamental rights assessment. A risk is not less severe simply because it defies straightforward measurement. To address this potential blind spot, it is crucial to introduce a formalized, transparent process outlining precisely when an additional risk crosses the threshold to become "systemic". Furthermore, explicitly linking each identified risk category to specific fundamental rights through concrete use cases would significantly enhance clarity and the adaptability of the Code. Finally, such a structured approach would help ensure that compliance remains robust against future developments in EU case law.

2.5 Elimination of oversight mechanisms

The third draft has eliminated two critical oversight mechanisms: the requirement to re-evaluate models every six months and the obligation to monitor risks even after market withdrawal. Instead of

[4] Measure II.1.2 makes systemic risk tiers mandatory only for selected systemic risks (Appendix 1.1) but optional for other systemic risks. Measure II.8.3 calls for documenting the systemic risk selection in the Model Report.

[5] Küsters (2025), Next Steps in Addressing Systemic AI Risk, cepAdhoc No 1/2025.

mandatory periodic reviews, the new draft only requires updates to Model Reports when providers "have reason to believe that there has been a material change in the systemic risk landscape that materially undermines the reasons for concluding that the model posed acceptable systemic risk". However, the rapid iteration of GPAI technology means that systemic risks can emerge in unexpected ways through novel applications, unforeseen interactions with other systems, or simply through the discovery of new capabilities in existing models. The shift to a trigger-based approach also introduces subjectivity into when reassessments should occur, potentially allowing providers to delay necessary reviews by claiming they have no "reason to believe" that material changes have occurred. This is especially problematic for released models whose applications and uses in various contexts cannot be fully anticipated by their providers. Similarly, post-withdrawal monitoring acknowledges that LLMs can continue to impact markets and societies beyond their commercial lifecycle.

While reducing administrative burden is important, the elimination of periodic and ex-post re-evaluation requirements appears to prioritize business convenience over prudent risk management. The removal of these provisions creates governance gaps for AI systems such as reasoning models based on "test-time compute" or autonomous AI agents that may develop unforeseen behaviors or continue to operate in various contexts long after formal market withdrawal. Current industry developments show that the computational resources available at the time of inference have a significant impact on model capabilities and hence implicit risk thresholds, but this variable can often change over time.[6] From an ordoliberal perspective, the described changes substantially weaken the temporal dimension of risk governance. Regular re-evaluation serves as a crucial safeguard against emergent risks that may develop as GPAI models interact with changing environments and real-world applications over time. Similarly, an ordoliberal approach would typically favor transparency and accountability mechanisms to ensure proper market functioning and to prevent the externalization of long-term costs to society. Without these mechanisms, there is a risk that harmful effects may go undetected until they have caused substantial damage, at which point intervention becomes more costly and less effective.

2.6 References to the "state of the art"

The third draft incorporates multiple references to the "state of the art", supported by input from the AI Office, as a means of ensuring that the implementation of the Safety and Security Section remains adaptive to evolving scientific, technological, and practical standards. For instance, systemic risk management by the Signatories must build on state of the art measures and, after implementation, they must "continuously" assess their robustness and adequacy following state of the art processes such as red-teaming. The Code defines "state of the art" as the most up-to-date stage of development in methods that is either (a) objectively and reasonably recognized by relevant industry consensus (such as industry standards), or (b) explicitly confirmed by the AI Office in formal guidance as reflecting the forefront of relevant research, technology, and practical experience. Conversely, any methods explicitly identified by the AI Office as insufficient are deemed to fall outside of the state of the art standard.

This flexible definition contributes positively to future-proofing the Code by embedding responsiveness to technological advancement and changing scientific understanding. However, given the very rapid pace of research in this area, it may inadvertently impose significant burdens on SMEs and new start-ups, especially those with fewer resources to continuously track evolving industry standards and

[6] Yixin Ji et al., 'Test-Time Compute: From System-1 Thinking to System-2 Thinking' (arXiv, 2025), https://doi.org/10.48550/ARXIV.2501.02497.

scientific literature, or to regularly adapt to updated guidance from the AI Office. Although the drafters indicate in their public explanation note that the Safety and Security Section will realistically impact only a small number of companies (estimated at between 5 to 15 entities at any given time), typically comprising US Big Tech firms, there remains the possibility of newcomers falling within the scope of these obligations. Such newcomers may thus face disproportionate compliance costs or barriers to market entry due to their limited capacity to always implement state of the art practices. Thus, from an ordoliberal perspective, it would be advisable to develop tailored support measures or clear guidance specifically aimed at reducing the "state of the art" compliance burden for SMEs.

2.7 Erosion of whistleblower protections

Finally, explicit whistleblower protections have been minimized (see Commitment II.13 on non-retaliation protections). What was initially a comprehensive framework is now limited to a non-retaliation commitment, offering weaker safeguards for individuals who come forward with critical information about potential misconduct. Historical experiences with whistleblower protection as well as ordoliberal theory point to the role of information in maintaining long-term security and fair competition. Whistleblower protections are vital for unearthing malpractices, especially in sectors such as AI that are characterized by complex technology and strong informational asymmetries between businesses and regulators. Lessons from EU competition law demonstrate how well-structured incentives and protections for informants can effectively dismantle cartels and uncover abuses of market power.[7] Weakening these provisions could reduce the likelihood that employees, contractors, or other stakeholders will report systemic risks or unethical practices, thereby undermining the Code's objectives.

3 Conclusion

The third draft of the EU General-Purpose AI Code of Practice makes notable progress by aligning obligations more closely with the greatest sources of potential harm, particularly large upstream AI providers. From an ordoliberal standpoint – with its emphasis on fair competition, transparency, and proportional regulation – this targeted approach will help foster an environment conducive to competition, innovation, and market entry for SMEs. In general, streamlined commitments, clearer exemptions for SMEs and open-source models, and a more cohesive structure all point toward a framework that can support responsible AI deployment without imposing undue burdens on smaller European businesses. However, several gaps remain unaddressed. Ambiguities over when downstream/fine-tuning entities become new "providers" could lead to unpredictable liabilities for startups, while legacy AI systems risk circumventing transparency requirements. Potential loopholes in external assessment obligations, the relegation of many important risks to an "optional" category, and the reduction in whistleblower protections likewise present challenges to the Code's credibility. Finally, the current draft eliminates mandatory six-month model re-evaluations and post-withdrawal monitoring requirements, replacing them with a subjective trigger-based approach. This shift creates significant governance gaps in a rapidly evolving technological landscape where AI risks emerge unexpectedly and based on computing resources available at inference time, potentially allowing harmful effects to go undetected until they cause damage. A competitive and SME-friendly market structure remains essential for safeguarding Europe's economic interests but also for demonstrating leadership in responsible AI.

[7] See: https://globalcompetitionreview.com/article/number-and-quality-of-eu-whistleblower-submissions-increasing-jaspers-says.

Author:

Dr. Anselm Küsters, LL.M., Head of Department Digitalisation and New Technologies, Berlin
kuesters@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg
Schiffbauerdamm 40 Raum 4205 | D-10117 Berlin
Tel. + 49 761 38693-0

The **Centrum für Europäische Politik** FREIBURG | BERLIN, the **Centre de Politique Européenne** PARIS, and the **Centro Politiche Europee** ROMA form the **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA.