Third Draft of the General-Purpose AI Code of Practice

Opening statement by the Chairs and Vice-Chairs

As the Chairs and Vice-Chairs of the four Working Groups, we hereby present the third draft of the General-Purpose AI Code of Practice under the AI Act (the "Code"). Participants in the Working Groups and observers of the Code of Practice Plenary are welcome to submit written feedback on this draft by Sunday, 30 March 2025, via a dedicated survey shared with them.

We encourage all readers – whether they have engaged with previous drafts or not – to visit this website. It contains the text of this third draft as well as two FAQs and an Explainer on parts of the Code and is aimed at making the Code more accessible to all observers and working group participants.

The third draft significantly advances the content compared to the second draft. In the upcoming final drafting round, it will be further improved based on stakeholder feedback. For this third draft, we have focused primarily on streamlining the structure of the Code, providing clarifications, adding essential details, and simplifying the Code.

This third draft of the Code addresses key considerations for providers of general-purpose AI models and providers of general-purpose AI models with systemic risk when complying with Chapter V of the AI Act, through four Working Groups working in close collaboration:

- Working Group 1: Transparency and copyright-related rules
- Working Group 2: Risk assessment for systemic risk
- Working Group 3: Technical risk mitigation for systemic risk
- Working Group 4: Governance risk mitigation for systemic risk

Working Group 1 Transparency applies to all general-purpose AI models, except for those that are released under a free and open-source licence satisfying the conditions specified in Article 53(2) AI Act and not classified as general-purpose AI models with systemic risk. Working Group 1 Copyright applies to all general-purpose AI models. Working Groups 2, 3, and 4 (Safety & Security Section) only apply to providers of general-purpose AI models classified as general-purpose AI models with systemic risk based on Article 51 AI Act.

Following a thorough review of the feedback received from stakeholders on the second draft, we have refined Commitments and Measures while maintaining the Code's Objectives. We present this third draft as the basis for the final drafting round, in which we will again draw on your feedback provided via the EU survey, in provider workshops, and in Working Group meetings. Like in previous drafting rounds, we have found your feedback extremely helpful, resulting in substantial changes. **We therefore encourage stakeholders to continue providing comprehensive feedback on all aspects of the Code, including both**

new and unchanged elements. Your feedback will help shape the final version of the Code, which will play a crucial role in guiding the future of general-purpose AI model development and deployment.

We have once again included a high-level drafting plan which outlines our guiding principles for the Code, and the assumptions it is based on.

The AI Act came into force on 1 August 2024, stating that the final version of the Code should be ready by 2 May 2025. The third draft builds upon previous work while aiming to provide a "future-proof" Code, appropriate for the next generation of models which will be developed and released in 2025 and thereafter.

In formulating this third draft, we have been principally guided by the provisions in the AI Act as to matters within the scope of the Code. Accordingly, unless the context and definition contained within the Code indicates otherwise, the terms used in the Code refer to identical terms from the AI Act.

Like the first and second drafts, this document is the result of a collaborative effort involving hundreds of participants from across industry, academia, and civil society. It has been informed by three rounds of feedback, including on the previous two drafts, which has been insightful and instructive in our drafting process. We continue to be informed by the evolving literature on AI governance, international approaches (as specified in Article 56(1) AI Act), industry best practice, and the expertise and experience of providers and Working Group members.

Key features of the development process of the Code include:

- Drafted by Chairs and Vice-Chairs who were selected by the AI Office for their expertise, experience, independence (including absence of financial interests), and to ensure gender and geographic diversity.
- A multi-stakeholder consultation which closed in September and received 427 submissions
- A multi-stakeholder survey on the first draft of the Code which received 354 submissions, and on the second draft which received 336 submissions
- Provider workshops led by Chairs and Vice-Chairs
- Four specialised working groups led by Chairs and Vice-Chairs
- Meetings with representatives from EU Member States in the AI Board and from the European Parliament

Additional time for consultation and deliberation – both externally and internally – will be needed to further improve the Code. As a group of independent Chairs and Vice-Chairs, we strive to make this process as transparent and accessible to stakeholders as possible, aiming to share our work and our thinking as early as possible, while taking sufficient time to coordinate and discuss key questions within Working Groups. We count on your continued engaged collaboration and constructive criticism.

We welcome written feedback by the Code of Practice Plenary participants and observers by Sunday, 30 March 2025, via a dedicated survey shared with them.

Thank you for your support!

Nuria Oliver *Working Group 1* *Co-Chair*	**Alexander Peukert** *Working Group 1* *Co-Chair*	**Matthias Samwald** *Working Group 2* *Chair*	**Yoshua Bengio** *Working Group 3* *Chair*	**Marietje Schaake** *Working Group 4* *Chair*
Rishi Bommasani *Working Group 1* *Vice-Chair*	**Céline Castets-Renard** *Working Group 1* *Vice-Chair*	**Marta Ziosi** *Working Group 2* *Vice-Chair*	**Daniel Privitera** *Working Group 3* *Vice-Chair*	**Anka Reuel** *Working Group 4* *Vice-Chair*
		Alexander Zacherl *Working Group 2* *Vice-Chair*	**Nitarshan Rajkumar** *Working Group 3* *Vice-Chair*	**Markus Anderljung** *Working Group 4* *Vice-Chair*

Drafting plan, principles, and assumptions

This third draft provides a more streamlined structure with more nuanced Commitments and Measures. In the upcoming final drafting round, it will be further improved based on stakeholder feedback. At this stage, it still does not contain the level of clarity and coherence that we expect in the final adopted version of the Code.

The Code first outlines the Commitments. Concretely, these are 2 Commitments for providers of general-purpose AI models and further 16 Commitments only for providers of general-purpose AI models classified as general-purpose AI models with systemic risk. Next, in separate documents, the Commitments are detailed out with respective Measures. The draft does not include KPIs and instead sharpened the reporting Commitments. Stakeholders should not expect the final adopted version of the Code to contain KPIs.

Related to transparency, Chairs have included a user-friendly Model Documentation Form which allows Signatories to easily document the necessary information in a single place. With regards to the review and adaptation of the Code, this draft includes an Appendix to the Safety and Security section with recommendations to the AI Office.

Below are some high-level principles we follow when drafting the Code:

1. **Alignment with EU Principles and Values** – Commitments and Measures will be in line with general principles and values of the Union, as enshrined in EU law, including the Charter of Fundamental Rights of the European Union, the Treaty on European Union and Treaty on the Functioning of the European Union.
2. **Alignment with the AI Act and International Approaches** – Commitments and Measures will contribute to a proper application of the AI Act. This includes taking into account international approaches (including standards or metrics developed by AI Safety Institutes, or standard-setting organisations), in accordance with Article 56(1) AI Act.
3. **Proportionality to Risks** – Commitments and Measures should be proportionate to risks, meaning they should be (i) suitable to achieve the desired end, (ii) necessary to achieve the desired end, and (iii) should not impose a burden that is excessive in relation to the end sought to be achieved. Some concrete applications of proportionality include:
    a. Commitments and Measures should be more stringent for higher risk tiers or uncertain risks of severe harm.
    b. Measures should be specific. While Commitments may be articulated at a higher level of generality, general-purpose AI model providers should have a clear understanding of how to meet Measures. Measures should be designed to be effective and robust against misspecification or any attempts of circumvention. The Code strives to accomplish this by, for example, avoiding unnecessary use of proxy terms or metrics. The AI Office will monitor and review Measures that may be susceptible to circumvention and other forms of misspecification.
    c. Commitments and Measures should differentiate, where applicable, between different types of risks, distribution strategies and deployment contexts of the concerned general-

purpose AI model, and other factors that may influence the tiers of risk, and how risks need to be assessed and mitigated. For example, Commitments and Measures assessing and mitigating systemic risks might need to differentiate between intentional and unintentional risks, including instances of misalignment. Additionally, Commitments may need to be adapted to take into account the different tools providers have available to assess and mitigate systemic risk where model weights are freely released.

4. **Future-Proof** – AI technology is changing rapidly. Measures should maintain the AI Office's ability to improve its assessment of compliance based on new information. Therefore, the Code shall strive to facilitate its rapid updating, as appropriate. It is important to find a balance between specific commitments on one hand, and the flexibility to update rapidly in light of technological and industry developments on the other. The Code can accomplish this by, for example, referencing dynamic sources of information that providers can be expected to monitor and consider in their risk assessment and mitigation. Examples of such sources could include incident databases, consensus standards, up-to-date risk registers, state-of-the-art risk management frameworks, and AI Office guidance. As technology evolves, it may also be necessary to articulate an additional set of Measures for specific general-purpose AI models, for example, certain models used in agentic AI systems.
5. **Proportionality to the size of the general-purpose AI model provider** – Measures related to the obligations applicable to providers of general-purpose AI models should take due account of the size of the general-purpose AI model provider and allow simplified ways of compliance for small and medium enterprises (SMEs) and start-ups with fewer financial resources than those at the frontier of AI development, where appropriate.
6. **Support and growth of the ecosystem for safe, human centric and trustworthy AI** – We recognise that the development, adoption, and governance of general-purpose AI models are global issues. Many Commitments in this draft are intended to enable and support cooperation between different stakeholders, for example by sharing general-purpose AI safety infrastructure and best practices amongst model providers, or by encouraging the participation of civil society, academia, third parties, and government organisations in evidence collection. We promote further transparency between stakeholders and increased efforts to share knowledge and cooperate in building a collective and robust evidence base for safe, human centric and trustworthy AI in line with Article 56(1) and (3), Recital 1, and Recital 116 AI Act. We also acknowledge the positive impact that open-source models have had on the development of safe, human centric and trustworthy AI.
7. **Innovation in AI governance and risk management** – We recognise that determining the most effective methods for understanding and ensuring the safety of general-purpose AI models remains an evolving challenge. The Code should encourage providers to compete in and advance the state-of-the-art in AI safety governance and related evidence collection methods and practices. When providers can demonstrate equal or superior safety outcomes through alternative approaches that are less burdensome, these innovations should be recognised as improving the state of the art of AI governance and evidence and we should support their wider adoption.

The current draft is written with the **assumption that there will only be a small number of both general-purpose AI models with systemic risk and providers thereof**. That assumption seems to be confirmed by the information provided from the AI Office accompanying the publication of this draft. The AI Office

plans to publish guidance in due time to clarify the scope of the respective AI Act rules in proximity to the publication of the final Code of Practice, including topics addressed in the dedicated Q&A such as downstream modifiers to which obligations should only apply in clearly specified cases. In particular, we want to highlight that even if modifications of general-purpose AI models increase the number of providers in scope, the modifiers' obligations under Articles 53 and 55 AI Act should be limited to the extent of their respective modifications, as appropriate. We expect more clarifications from the AI Office on these points on an ongoing basis, as stated in the Q&A.

Preamble

a) The Signatories of this Code of Practice (hereafter, "Code") recognise the importance of improving the functioning of the internal market and creating a level playing field for the development, placing on the market, and use of human-centric and trustworthy artificial intelligence (hereafter, "AI"), while ensuring a high level of protection of health, safety, and the fundamental rights enshrined in the Charter, including democracy, the rule of law, and environmental protection, against harmful effects of AI in the Union and supporting innovation as emphasised in Article 1(1) AI Act. The Code shall be interpreted in this context.

b) The Signatories recognise that this Code is to be interpreted in conjunction and in accordance with any European AI Office (hereafter, "AI Office") guidance on the AI Act and with applicable Union laws.

c) Whenever the Code refers to providers of general-purpose AI models it shall encompass providers of general-purpose AI models with systemic risk (hereafter "GPAISRs" or "GPAISR"), too. Whenever the Code refers to providers of GPAISRs it shall not encompass providers of other general-purpose AI models. This shall only include general-purpose AI models that are within the scope of the AI Act.

d) The Signatories recognise that the Code serves as a guiding document for demonstrating compliance with the AI Act, while recognising that adherence to the Code does not constitute conclusive evidence of compliance with the AI Act.

e) The Signatories recognise the importance of regularly reporting to the AI Office on their implementation of the Code and its outcomes (Article 56(5) AI Act), including to facilitate the regular monitoring and evaluation of the Code's adequacy by the AI Office and the Board (Article 56(6) AI Act).

f) The Signatories recognise that the Code shall be subject to regular review by the AI Office and the Board (Article 56(6) AI Act) and that the AI Office may encourage and facilitate updates of the Code to reflect advances in AI technology, emerging standards, societal changes, and emerging systemic risks (Article 56(8) AI Act), without prejudice to the need for Signatories to sign such updates.

g) The Signatories recognise that the Code may serve as a bridge until the adoption of a harmonised standard. Updates may be needed to facilitate a gradual transition towards future standards.

h) The Signatories recognise that the absence of specific Commitments or Measures within this Code does not absolve providers of GPAISRs from their responsibility to assess and mitigate systemic risks.

i) The Signatories recognise the importance of working in partnership with the AI Office to foster collaboration between providers of general-purpose AI models, researchers, and regulatory bodies to address emerging challenges and opportunities in the AI landscape.

The Objectives of the Code are as follows:

I. Assisting providers of general-purpose AI models to effectively comply with their obligations under the AI Act - if assessed as adequate by the AI Office and the Board (Article 56(6) AI Act). The Code should also enable the AI Office to assess compliance of providers who choose to rely on the Code to demonstrate compliance with their obligations under the AI Act. This can involve, e.g., allowing sufficient visibility into trends in the development, making available, and use of general-purpose AI models, particularly of the most advanced models.

II. Assisting providers of general-purpose AI models to effectively keep up-to-date technical documentation of their models and to effectively ensure a good understanding of general-purpose AI models along the entire AI value chain, both to enable the integration of such models into downstream products and to fulfil subsequent obligations under the AI Act or other regulations (see Articles 53(1)(a) and (b) and Recital 101 AI Act).

III. Assisting providers of general-purpose AI models to effectively comply with Union law on copyright and related rights and increase transparency on the data that is used in the pre-training and training of general-purpose AI models (see Articles 53(1)(c) and (d) and Recitals 106 and 107 AI Act).

IV. Assisting providers of GPAISRs to effectively and continuously assess and mitigate systemic risks, including their sources, that may stem from the development, the placing on the market, or the use of GPAISRs (see Article 55(1) and Recital 114 AI Act).

COMMITMENTS

I. COMMITMENTS BY PROVIDERS OF GENERAL-PURPOSE AI MODELS

TRANSPARENCY SECTION

Commitment I.1. Documentation

In order to fulfil the obligations in Article 53(1), points (a) and (b) AI Act, Signatories commit to drawing up and keeping up-to-date model documentation in accordance with Measure I.1.1, providing relevant information to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems (downstream providers hereafter), and to the AI Office upon request (possibly on behalf of national competent authorities when this is strictly necessary for the exercise of their supervisory tasks under the AI Act, in particular to assess the compliance of high-risk AI systems built on general-purpose AI models where the provider of the system is different from the provider of the model[1]), in accordance with Measure I.1.2, and ensuring quality, security, and integrity of the documented information in accordance with Measure I.1.3. These Measures do not apply to providers of open-source AI models satisfying the conditions specified in Article 53(2) AI Act, unless the models are general-purpose AI models with systemic risk.

COPYRIGHT SECTION

Commitment I.2. Copyright policy

In order to fulfil the obligation to put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790 pursuant to Article 53(1), point (c) AI Act, Signatories commit to drawing up, keeping up-to-date, and implementing a copyright policy in accordance with Measure I.2.1, as well as adopting Measures I.2.2—I.2.6 for their general-purpose AI models placed on the EU market.

II. COMMITMENTS BY PROVIDERS OF GENERAL-PURPOSE AI MODELS WITH SYSTEMIC RISK

SAFETY AND SECURITY SECTION

Commitment II.1. Safety and Security Framework

Signatories commit to adopting and implementing a Safety and Security Framework (hereafter, "the Framework") that will: (1) apply to the Signatories' GPAISRs; and (2) detail the systemic risk assessment, systemic risk mitigation, and governance risk mitigation measures and procedures that Signatories intend to adopt to keep systemic risks stemming from their GPAISRs within acceptable levels.

[1] See Article 75(1) and (3) and Article 88(2) AI Act.

Commitment II.2. Systemic risk assessment and mitigation along the entire model lifecycle, including during model development

Signatories commit to conducting systemic risk assessment systematically at appropriate points along the entire model lifecycle, in particular before making the model available on the market. Specifically, Signatories commit to starting to assess and mitigate systemic risks during the development of a GPAISR, as specified in the Measures for this Commitment.

Commitment II.3. Systemic risk identification

Signatories commit to selecting and further characterising systemic risks stemming from their GPAISRs that are significant enough to warrant further assessment and mitigation, as specified in the Measures for this Commitment.

Commitment II.4. Systemic risk analysis

As part of systemic risk assessment, Signatories commit to carrying out a rigorous analysis of the systemic risks identified pursuant to Commitment II.3 in order to understand the severity and probability of the systemic risks. Signatories commit to carrying out systemic risk analysis with varying degrees of depth and intensity, as appropriate to the systemic risk stemming from the relevant GPAISR and as specified in the Measures for this Commitment. Whenever systemic risk mitigations are implemented, Signatories commit to considering their effectiveness and robustness as part of systemic risk analysis.

As further specified in the Measures for this Commitment, Signatories commit to making use of a range of information and methods in their systemic risk analysis including model-independent information and state-of-the-art model evaluations, taking into account model affordances, safe originator models, and the context in which the model may be made available on the market and/or used and its effects.

Commitment II.5. Systemic risk acceptance determination

Signatories commit to determining the acceptability of the systemic risks stemming from their GPAISRs by comparing the results of their systemic risk analysis (pursuant to Commitment II.4) to their pre-defined systemic risk acceptance criteria (pursuant to Measure II.1.2), in order to ensure proportionality between the systemic risks of the GPAISR and their mitigations. Signatories commit to using this comparison to inform the decision of whether or not to proceed with the development, the making available on the market, and/or the use of their GPAISR, as specified in the Measures for this Commitment.

Commitment II.6. Safety mitigations

Signatories commit, as specified in the Measures for this Commitment, to: (1) implementing technical safety mitigations along the entire model lifecycle that are proportionate to the systemic risks arising from the development, the making available on the market, and/or the use of GPAISRs, in order to reduce the systemic risks of such models to acceptable levels, and further reduce systemic risk as appropriate, in accordance with this Code; and (2) ensuring that safety mitigations are proportionate and state-of-the-art.

Commitment II.7. Security mitigations

Signatories commit to mitigating systemic risks that could arise from unauthorised access to unreleased model weights of their GPAISRs and/or unreleased associated assets. Associated assets encompass any information critical to the training of the model, such as algorithmic insights, training data, or training code.

Consequently, Signatories commit to implementing state-of-the-art security mitigations designed to thwart such unauthorised access by well-resourced and motivated non-state-level adversaries, including insider threats from humans or AI systems, so as to meet at least the RAND SL3 security goal or equivalent, and achieve higher security goals (e.g. RAND SL4 or SL5), as specified in the Measures for this Commitment.

Commitment II.8. Safety and Security Model Reports

Signatories commit to reporting to the AI Office about their implementation of the Code, and especially the application of their Framework to the development, making available on the market, and/or use of their GPAISRs, by creating a Safety and Security Model Report (hereafter, a "Model Report") for each GPAISR which they make available on the market, which will document, as specified in the Measures for this Commitment: (1) the results of systemic risk assessment and mitigation for the model in question; and (2) justifications of decisions to make the model in question available on the market.

Commitment II.9. Adequacy assessments

Signatories commit to assessing the adequacy of their Framework, the adoption and implementation of which they have committed to under Commitment II.1, and to updating it based on the findings as specified in the Measures for this Commitment.

Commitment II.10. Systemic risk responsibility allocation

For activities concerning systemic risk assessment and mitigation for their GPAISRs, Signatories commit, as specified in the Measures for this Commitment, to: (1) clearly defining and allocating responsibilities for managing systemic risk from their GPAISRs across all levels of the organisation; (2) allocating appropriate resources to actors who have been assigned responsibilities for managing systemic risk; and (3) promoting a healthy risk culture.

Signatories commit to allocating appropriate levels of responsibility and resources proportionately to, at least, the Signatory's organisational complexity and governance structure, and the systemic risks stemming from their GPAISRs.

Commitment II.11. Independent external assessors

Before placing a GPAISR on the market, Signatories commit to obtaining independent external systemic risk assessments, including model evaluations, unless the model can be deemed sufficiently safe, as specified in Measure II.11.1. After placing the GPAISR on the market, Signatories commit to facilitating exploratory independent external assessments, including model evaluations, as specified in Measure II.11.2.

Commitment II.12. Serious incident reporting

Signatories commit, to the extent and under the conditions specified in Measures II.12.1 to II.12.4, to setting up processes for keeping track of, documenting, and reporting to the AI Office and, as appropriate, to national competent authorities without undue delay relevant information about serious incidents throughout the entire model lifecycle and possible corrective measures to address them, with adequate resourcing of such processes relative to the severity of the serious incident and the degree of involvement of their model.

Commitment II.13. Non-retaliation protections

Signatories commit to not retaliating against any worker providing information about systemic risks stemming from the Signatories' GPAISRs to the AI Office or, as appropriate, to national competent authorities, and to at least annually informing workers of an AI Office mailbox designated for receiving such information, if such a mailbox exists.

Commitment II.14. Notifications

Signatories commit, as specified in the Measures for this Commitment, to: (1) notifying the AI Office of relevant information regarding their general-purpose AI models meeting the condition for classification as GPAISRs; and (2) regularly notifying the AI Office of the implementation of the Commitments and Measures of this Code. For the purpose of assessing the implementation of this Code through the AI Office, Signatories commit to offering clarifications, including via further documentation or interviews, where requested by the AI Office.

Commitment II.15. Documentation

Signatories commit to documenting relevant information under the AI Act and the Code, as specified in Measure II.15.1.

Commitment II.16. Public transparency

Signatories commit to publishing information relevant to the public understanding of systemic risks stemming from their GPAISRs, where necessary to effectively enable assessment and mitigation of systemic risks, to the extent and under the conditions specified in Measure II.16.1.

The foregoing Commitments are supplemented by Measures found in the relevant Transparency, Copyright or Safety and Security section in the separate accompanying documents.