POLICY BRIEF

From Code to Compliance: The EU's Bid to Regulate General-Purpose AI

Lisa Soder, Ema Prović

July 17, 2025

Tech analysis and policy ideas for Europe

interface

Table of Contents

The recently published Code of Practice, a tool to comply with the EU AI Act's GPAI rules, is the world's first major effort to translate high-level AI safety principles into enforceable measures. A crucial test case for frontier AI regulation, the Code's success will ultimately depend on the political will and institutional capacity to enforce it.

Introduction

The past years have seen no shortage of AI governance initiatives. International summits in Bletchley, Seoul, and Paris have convened industry, governments, and civil society. A number of governments, such as the UK, Canada and Singapore, established AI Safety (or Security) Institutes that now coordinate through an international Network. Meanwhile, the G7 developed governance principles for advanced AI systems through its Hiroshima Process, while industry leaders signed voluntary safety commitments in Seoul. The most recent addition to this evolving governance landscape is the EU's Code of Practice, a compliance tool for the EU AI Act's rules on GPAI.

This comes just in time: In a few weeks, on August 2, 2025, the AI Act's GPAI rules enter into force, establishing the world's first binding requirements for model documentation, safety evaluations, and systemic risk assessments. Providers seeking to deploy models on the European market will have to meet these obligations, or else they could face fines and market restrictions. The Code of Practice, developed through a multi-stakeholder process and facilitated by the newly established EU AI Office, translates broad legal requirements into specific technical measures. Unlike other frameworks—such as the Seoul Frontier AI Safety commitments or the G7 Hiroshima process—the Code derives its significance through its connection to enforceable legislation. While providers can demonstrate compliance through alternative means, these may face greater regulatory scrutiny, most likely making the Code the practical pathway for most providers. Indeed, major AI providers, including OpenAI and Mistral, have already signed the Code, signaling early industry adoption of this compliance pathway.

As the world's first attempt to create a binding legal framework for GPAI, the Code could have an impact beyond Brussels. If implemented successfully, it could provide much-needed visibility into the practices of frontier AI companies and generate crucial evidence for future AI governance efforts. However, its success is not guaranteed and now hinges on the AI Office's ability to drive implementation. These range from building up the AI Office's regulatory capacity, creating updating mechanisms for the code, to navigating a new geopolitical environment, not least by managing its relationship with the US' Center for AI Standards and Innovation, whose explicit mission is to '*guard against burdensome and unnecessary regulation of*

American technologies'.

What is the code, and why was it developed?

The EU's Code of Practice was borne out of a classic challenge regulators face when governing rapidly evolving technologies: write rules too specific, and they become obsolete before implementation; keep them too vague, and companies are left with uncertainty about what compliance looks like.

The AI Act addressed the first challenge by being deliberately broad, requiring developers to use 'state-of-the-art' measures and 'assess systemic risks' to future-proof it against new developments. The second challenge, providing concrete guidance, would traditionally be resolved through European standardisation bodies like CEN-CENELEC, which develop detailed technical standards. However, this process can take years, far too long, given that the GPAI provisions will become enforceable on August 2, 2025.

The Code's role is to bridge this gap. Developed in eleven months, it serves as a compliance tool designed to guide providers of GPAI models in meeting their obligations under the AI Act. While technically voluntary, as providers can pursue alternative means of compliance and draw up their own documentation framework, the Code offers more or less a manual for compliance and, crucially, reveals how the Commission will likely interpret the Act's requirements.

Noteworthy about the Code of Practice is not only its function as an expedited pathway to implementing the AI Act, but also the process by which it was developed. The EU AI Office facilitated an eleven-month multi-stakeholder process, led by 13 leading scientists who were nominated to develop the text. Over 1,000 participants from industry, academia, civil society, as well as international and EU governments, contributed through workshops and written feedback across four iterations.

What and who is covered by the code?

In line with the AI Act's risk-based approach, the Code distinguishes between two categories of GPAI models. All providers (with some open source exceptions) must meet baseline obligations related to transparency and copyright, laid out in the first two chapters. Additional requirements apply to providers whose models are

classified as posing “systemic risk”—typically the most advanced systems on the market, currently identified by the scale of computational resources used in training. These models are subject to further safety and security obligations outlined in the Code’s third chapter.

Tier 1: Baseline Requirements for All GPAI Providers

The first chapter on Transparency applies to all providers of general-purpose AI models, except those releasing models under a free and open-source license—unless those open-source models are found to pose a "systemic risk." Providers are required to implement three transparency measures to ensure clear and up-to-date documentation throughout the AI development process. To facilitate compliance, the chapter introduces a user-friendly "Model Documentation Form," outlining the necessary information for providers to fulfill the transparency obligations set by the AI Act.

The second chapter on copyright applies to all providers of GPAI models, irrespective of whether they are open-source. The chapter outlines five specific measures for creating and implementing copyright policies, establishing web crawling practices, and devising strategies to mitigate copyright infringement risks. Importantly, while following this chapter allows providers to demonstrate compliance with the AI Act’s copyright provisions, it does not automatically ensure compliance with existing EU Copyright law - this is still up for courts to decide.

Tier 2: Safety & Security Requirements for Models with ‘Systemic Risk’

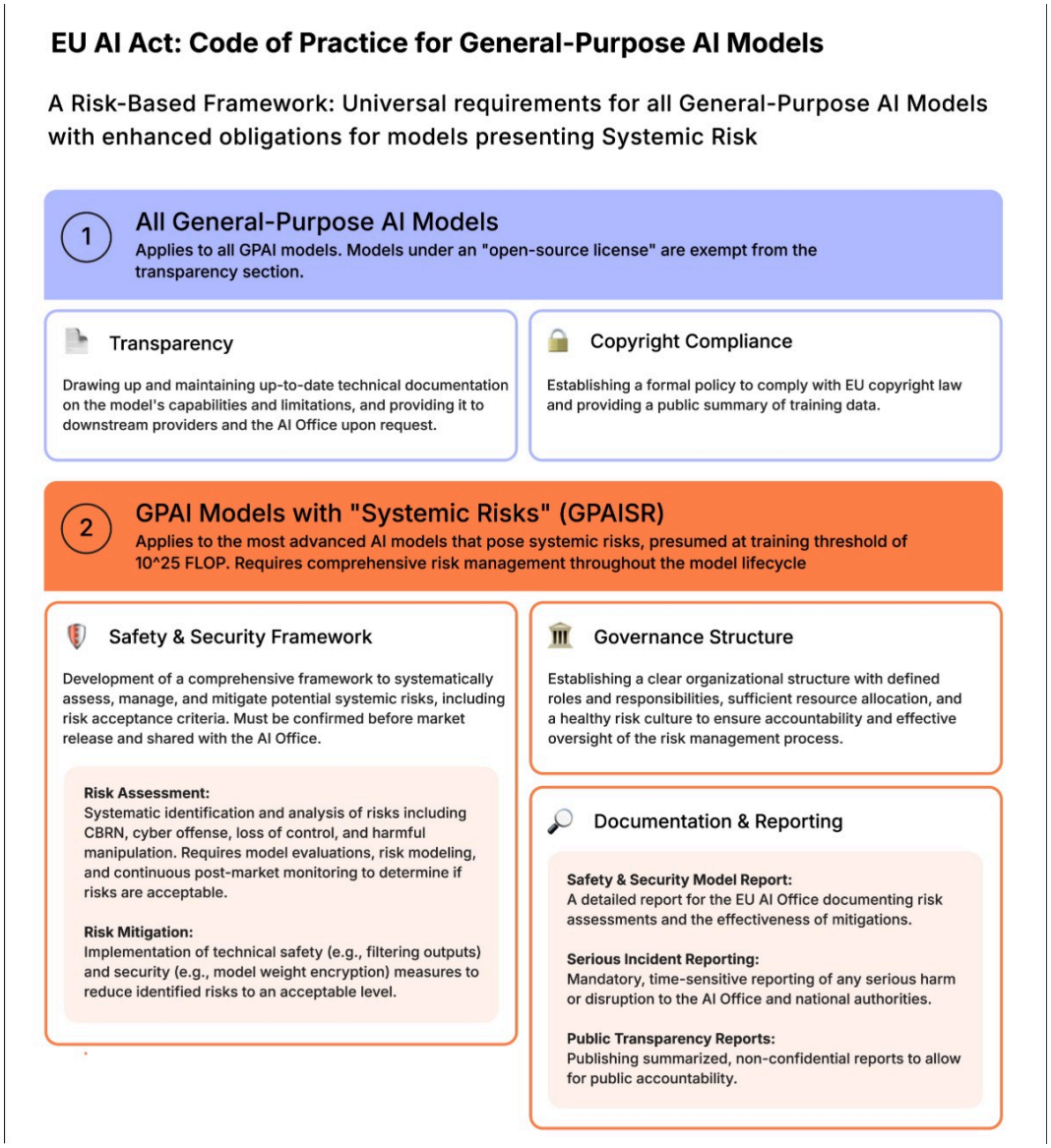
The third chapter on ‘Safety and Security’ applies to ‘General Purpose AI Models with Systemic Risk’ (GPAISR). According to the authors of the code, this should capture 5-15 providers at any given time, ensuring it only covers the most advanced models. However, the initial designation mechanism, a training compute threshold of 10^{25} FLOP (which essentially measures the computational resources invested in training a model), would currently capture approximately 33 models. This creates a notable discrepancy between the intended and actual scope of the Code, which can, however, be updated by the Commission through a delegated act.

For GPAISR models, the AI Act mandates roughly four categories of measures: a safety and security framework, risk assessment, risk mitigation, and governance measures. The first commitment is a Safety and Security Framework, essentially a risk management protocol that AI companies must maintain, documenting how

they will identify, assess, and mitigate potential harms throughout development and deployment. This mirrors existing 'frontier AI safety policies' that developers like OpenAI, and Anthropic have already drawn up, but adds specificity, such as pre-defining risk thresholds.

Following this framework, companies must assess risks along the model lifecycle and check whether the model exceeds specified risk thresholds. The commitments on risk assessment require AI companies to systematically identify, analyse, and evaluate potential harms at predetermined milestones throughout development, using methods like red-teaming, capability evaluations, and safety margin calculations. Companies must then compare these assessed risks against pre-defined thresholds to make explicit decisions about whether to proceed with deployment, implement additional safeguards, or halt development entirely. This goes hand in hand with the third section, risk mitigation measures, requiring companies to implement technical 'state of the art' measures to ensure safety mitigation (though it doesn't describe which ones) and keep cybersecurity standards to prevent unauthorised access to unreleased model weights.

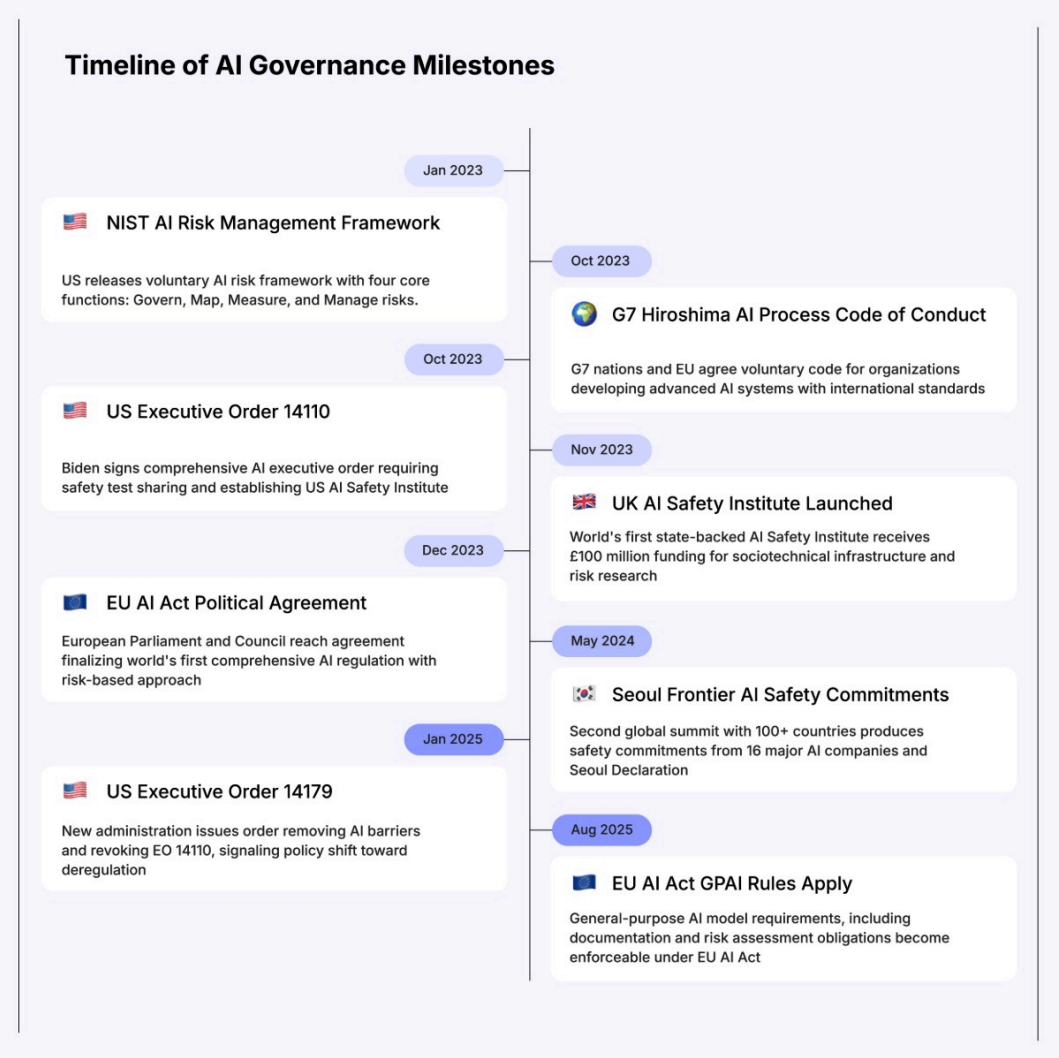
Finally, the last section addresses governance measures, including reporting requirements and the AI Office, such as submitting the 'Safety and Security Model Report' when or shortly after a model is released to the market, establishing processes for incident reporting, and, under specific circumstances, conducting external evaluations. However, beyond model-specific measures, it also prescribes wider measures for the organisation that is developing, such as defining who is responsible for risk oversight, allocating adequate resources to safety teams, and promoting a healthy risk culture throughout the organisation. Notably, while earlier drafts included whistleblower provisions, the final text only references the existing EU whistleblower directive, which likely wouldn't protect employees outside the EU, such as those in the US.



How does the Code fit within the international context?

The EU's Code of Practice is emerging in a global governance landscape where an almost remarkable consensus has formed around the core principles for governing the most capable AI models. From international agreements, such as the G7 Hiroshima Process, to national guidance like the UK's emerging process and the US's risk management framework, and frontier AI safety commitments made at the Seoul Summit, all converge on similar ideas: setting up a risk management process, requiring model evaluations, cybersecurity provisions, and accountability measures such as external scrutiny and sharing relevant information with governments. In line with a key drafting principle of international alignment, the Code of Practice builds directly upon this global consensus, with the notable exception of its copyright

measures, which reflect specific, existing EU law. Consequently, the framework is designed less to export new rules via the 'Brussels Effect' and more to formalise and add legal weight to emerging international norms.



The key differences, therefore, lie not in *the what*, but in *the how*. The European approach stands out in three ways.

First, the Code is more prescriptive than the above-mentioned frameworks, drawing up templates, specifying timelines, and formal structures that leave less room for interpretation. For instance, while in the voluntary commitments companies agreed to setting up a risk management framework, the EU's approach has an explicitly demarcated Safety and Security Framework that lays out concrete risks to be tested against, and specifies what needs to be tested, when in the lifecycle this must occur, and how it should be documented.

Second, the Code diverges on the means of enforcement. Most voluntary frameworks rely on 'soft power' methods such as optional adoption and public

pressure. The UK, for example, delayed its planned 'Frontier AI bill' but also increasingly positions its AI Security Institute as a technical partner to industry, rather than a regulator, given its lack of enforcement power. Although it remains to be seen how the AI Office will approach enforcement, it has various tools available. These range from requesting information and conducting interviews with industry to levying substantial fines of up to 3% of a company's annual total worldwide turnover for breaches of the AI Act. In extreme cases, it can even call for a model to be taken off the market. In other words, the EU has some deterrent mechanisms to encourage compliance.

Third, and perhaps most distinctively, the Code's development process itself offers a model that differs from traditional standard-setting approaches. While surveys show broad public support for AI regulation (71% globally), they reveal mistrust in single-actor governance: 82% of U.S. voters say tech executives cannot self-regulate, and 63% believe government regulators lack adequate understanding of emerging technologies. Instead, a majority wants universities, ethicists, and civil society involved in setting AI standards. After all, these standards will address a lot of normative questions like defining what a 'systemic risk' is or what meaningful accountability should look like.

Traditional standardisation bodies usually allow for limited public input and participation fees can exclude smaller actors. The Code's approach, which tasked independent expert chairs with mediating input from a broad and diverse forum, could offer a more accessible and legitimate alternative. While this multi-stakeholder model carries risks, including potential tokenism where stakeholder inclusion becomes performative, it represents an important experiment in democratic tech governance that, if genuinely implemented, could create standards that are both technically sound and socially trusted.

Ultimately, the global landscape shaped by a logic of 'convergent goals, divergent methods' is not necessarily a weakness. While the risk of regulatory fragmentation exists, it also creates an opportunity to test out different approaches and learn from them. By requiring documentation, Safety and Security Frameworks, Model Reports, incident logs, the Code might generate a first empirical record of how AI governance works in frontier AI companies. This data, if used wisely, could enable evidence-based refinement across all jurisdictions, revealing which risk assessments prove predictive and which governance structures enable effective responses.

The challenge ahead for the international community is to ensure that they can learn from one another. This will require deliberate coordination through shared technical standards, mutual recognition of assessments, and continuous engagement through emerging venues like the Network of AI Safety Institutes. If the international community can successfully learn from diverging approaches, the

current diversity may prove a strength rather than a weakness in building out our AI governance toolkit.

The Frontier AI Safety Commitments

Announced at the 2024 AI Seoul Summit, the Frontier AI Safety Commitments represent a voluntary industry initiative endorsed by leading AI developers. Signatories pledge to conduct pre-deployment testing, invest in safety research, and cease development if unable to guarantee safety. However, the commitments remain high-level, lacking specific implementation guidance or oversight mechanisms.

The G7 Hiroshima Process

The Hiroshima Process Code of Conduct, endorsed by G7 nations and over 50 additional countries, calls for transparency reports, risk assessments, and appropriate safeguards for advanced AI systems. The OECD is developing monitoring mechanisms, but participation remains voluntary and implementation details are still emerging.

The US Innovation-Led Model

The United States favours voluntary standards over prescriptive rules; after revoking the 2023 AI executive order in January 2025, it placed the NIST-based Center for AI Standards and Innovation (CAISI) at the helm to test frontier models for national-security risks and to diffuse best practices that industry adopts through market incentives.

The UK's "Governance-as-a-Service" Strategy

The United Kingdom has carved out a unique role for itself as a global hub for AI safety expertise. Its AI Safety Institute acts as a specialized, state-led technical auditor, publishing its own guidelines and seeking to build influence through its technical authority and its capacity to evaluate the most advanced AI models, rather than through broad market regulation.

The EU Regulatory Framework

The EU's model, embodied in the Code of Practice, is designed as a comprehensive and enforceable market regulation. Its objective is to create a legally certain environment by translating shared principles into detailed, quasi-binding procedures for any company wishing to access the EU market.

What challenges lie ahead?

With the publication of the Code of Practice, the EU's regulatory ambitions enter a new and no less challenging phase.

If implemented effectively, the Code, along with the AI Office's enforcement powers, provides strategic benefits for policy-makers and providers that go beyond existing voluntary commitments. For the first time, the AI office will have insight into the risk management processes of the world's most advanced AI models. Additionally, the documentation provided through the Code may generate a more detailed empirical record of critical information, including incident logs, risk assessments, and the effectiveness of risk mitigation measures. This information could serve as feedback mechanism to enhance future AI policies. For GPAI providers, the Code should ensure legal certainty of their compliance with AI Act. In light of internationally harmonised standards, this may set a global precedent and become the preferred choice for providers looking to deploy their models in the European market.

Realising these opportunities, however, is entirely contingent on the nascent AI Office confronting several formidable challenges.

The first challenge is whether the EU AI Office can build the institutional capacity and technical expertise to make use of its enforcement powers, be it requesting information, accessing models for evaluations, or asking them to implement risk mitigations. This requires more than just legal authority; it demands a deep bench of talent capable of scrutinizing the complex systems of well-resourced companies, especially given providers' anti-regulatory threats to delay European deployments. By doing so, it could take inspiration from the UK's AI Safety Institute, which has built a world-class team by hiring experts directly from major tech firms and top research institutions. Secondly, this technical capacity must be backed by political will to resist pressure from the US administration. Promising signals have emerged on this front, with Henna Virkkunen, the EU's Commissioner on Tech, reportedly making it clear that the EU's digital rulebook is non-negotiable in trade talks with the US.

Finally, beyond building its own capacity, the AI Office faces the challenge of ensuring the Code itself remains a precise and relevant tool. This requires demonstrating regulatory agility on two critical fronts. First, it must clarify the scope of the rules relating to ‘systemic risk’. The Code’s drafters intended its strictest rules to apply narrowly to only a handful of frontier models, but the current compute threshold (10^{25} FLOP) is a crude proxy that risks applying these obligations

far more broadly. Second, to prevent the Code from becoming obsolete, the Office must establish a formal mechanism for updating its technical provisions on a regular, predictable cadence, ensuring the framework can evolve alongside the rapid pace of AI development.

Conclusion

Viewed against the existing AI governance landscape, where voluntary commitments have so far set the tone, the EU's Code of Practice stands apart as the first framework developed under, and linked to, binding legislation. With leading frontier AI companies—including OpenAI and Mistral—already committed, its impact may extend beyond Europe, potentially reshaping how major global developers document, assess, and mitigate AI risks. However, the real test is yet to come. Whether the Code becomes an influential global template or remains merely a regional regulatory experiment will depend entirely on effective implementation. For the EU AI Office, that task will require not only significant technical expertise but sustained political courage and a readiness to enforce the rules rigorously. Either way, the insights and lessons generated by this effort will shape AI governance for years to come.

Disclosure: The views and opinions expressed are solely those of the authors and do not reflect or represent any official position of the European Commission. Prović is a policy officer with the European AI Office, and the European Commission has the right to review publications by its staff for accuracy and sensitive information.

Authors

Lisa Soder
Senior Policy Researcher / Acting Head Technical AI Governance
lsoder@interface-eu.org

Ema Vrović
Officer at the European AI Office of the European Commission
Ema.PROVIC@ext.ec.europa.eu

Imprint

interface – Tech analysis and policy ideas for Europe
(formerly Stiftung Neue Verantwortung)

W www.interface-eu.org
E info@interface-eu.org
T +49 (0) 30 81 45 03 78 80
F +49 (0) 30 81 45 03 78 97

interface – Tech analysis and policy ideas for Europe e.V.
Ebertstraße 2
D-10117 Berlin

Design by Make Studio
www.make.studio
Code by Convoy
www.convoyinteractive.com