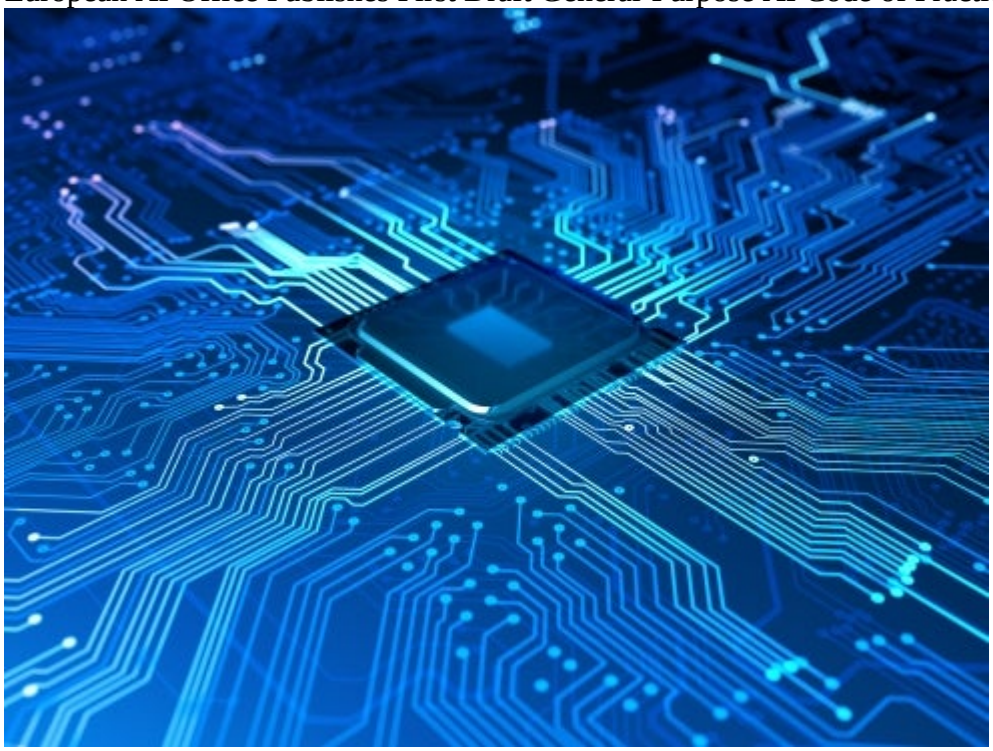
Updates
December 02, 2024
European AI Office Publishes First Draft General-Purpose AI Code of Practice



In an effort led by industry experts and nearly 1,000 stakeholders, the European Union introduced the first draft of the General-Purpose AI (GPAI) Code of Practice (the Code), an important addition to its regulatory framework for artificial intelligence (AI). Supplementing the initial EU AI Act that went into effect August 1, 2024, this draft Code aims to address the nuanced challenges of GPAI systems. For purposes of the EU AI Act, GPAI refers to foundational AI models that power a variety of applications, such as large language models and generative AI systems. These models provide flexible frameworks that businesses and developers can adapt for specialized tasks. The EU AI Act requires the AI Office, which was created by the European Commission to provide a comprehensive governance system for AI in the EU, to establish codes of practice by May 2, 2025. If covered parties comply with the codes of practice, they will be treated as presumptively compliant with the EU's final AI rules when they come into effect on August 2, 2025. Thus, compliance with the Code gives GPAI providers a safe harbor from claims of potential violation of the EU AI Act, at least until EU standard-setting bodies release final regulations. For more background on the EU AI Act, see our previous Update.

Scope

The Code aims to guide GPAI providers in meeting obligations under Articles 53 and 55 of the EU AI Act, particularly with respect to those GPAI models posing higher risks, such as misuse or unintended consequences. It addresses critical areas like:

- Transparency in how these models are developed and used.
- Managing risks associated with powerful AI systems.
- Ethical and legal compliance.

Core Principles Guiding the Code

The Code is informed by certain core principles aimed at aligning AI developments with EU values and legal standards. These principles include ensuring:

- Alignment with EU principles and values.
- Alignment with EU AI Act and international approaches.
- Proportionality to risks.
- Proportionality to size and capacity of providers.
- Support and growth of the AI safety ecosystem.
- Future-proofing of regulatory measures.

These principles are designed to provide a regulatory framework that is adaptable and reflective of a range of technological landscapes.

Navigating the Rules for GPAI Models—Including GPAI With Systemic Risk

1. Rules for Providers of GPAI Models

The Code specifies requirements for GPAI model providers, focusing on:

- **Transparency.** Providers should maintain clear and detailed documentation about their models, including their design, intended uses, and risks.
- **Acceptable use policies.** Providers should establish clear guidelines for users, outlining permitted and prohibited uses.
- **Public trust.** Providers are encouraged to disclose relevant information publicly where feasible.

2. Taxonomy of Systemic Risks

The Code introduces a taxonomy of "systemic risks," designed to address the wide-ranging impacts that GPAI models can pose, "in which GPAI models may cause large-scale negative effects on public health, safety, democratic processes, public and economic security, critical infrastructure, fundamental rights, environmental resources, economic stability, human agency, or society as a whole." This framework categorizes risks by their type, nature, and source, enabling providers to identify and mitigate potential harms effectively. The Code provides the following examples as an initial list of categories:

- **Cyber offence**. Risk related to offensive cyber capabilities, such as security vulnerability exploitation.
- **Chemical, biological, radiological, and nuclear risks**. Risk stemming from dual-use science enabling chemical, biological, radiological, and nuclear weapons development, design, acquisition, and use offensively.
- **Loss of control**. Risk related to powerful, autonomous GPAI models that are not able to be contained.
- **Automated use of models for AI research and development**. Risk from unpredictable developments of GPAI models due to automated uses, resulting in fast-paced growth.
- **Persuasion and manipulation**. Risk to democratic values and human rights from large-scale disinformation or misinformation, loss of trust in the media, and oversimplification of knowledge.
- **Large-scale discrimination**. Risk of illegal discrimination against individuals and communities.

Providers should align their risk assessments with these categories and incorporate updates as the taxonomy evolves to reflect emerging challenges.

3. Rules for Providers of GPAI Models With Systemic Risk

Providers of GPAI models identified as having systemic risks would have additional standards to comply with the Code. To satisfy the Code, they must adopt robust safety and security frameworks to proactively manage risks throughout the lifecycle of their models. Further evidence may require literature reviews, open-source project analysis, forecasting of general trends, and involving academia and other relevant stakeholders.

- **Risk identification**. One measure of the Code would be continuous evaluations, including adversarial testing, to identify vulnerabilities and prevent potential harms. This includes mapping model capabilities, propensities, and other sources of risk to respective systemic risk indicators.
- **Risk assessment**. Transparency is a core requirement of the Code, with providers needing to document their risk assessments comprehensively and submit to independent expert evaluations. This would require provider assessments throughout the training, deployment, and post-deployment processes.
- **Risk mitigation**. Mitigation strategies will detail safety mitigations and security mitigations proportional to the severity of the risks and tailored to the provider's size and resources, ensuring flexibility for smaller entities and open-source models.
    - Safety mitigations may include behavioral modifications to a model, safeguards for deployment in a system, and safety tools made available to others to reduce systemic risk.
    - On the other hand, security mitigations would detail weights and assets, such as protection of weights and assets at rest, in motion, and in use; access control, monitoring, and hardened interfaces; ongoing security red-teaming and accredited security reviews; and threat screening.

4. Technical and Governance Measures

As part of risk mitigation and assessment, compliance with the Code would entail comprehensive technical and governance measures. This approach embeds systemic risk considerations into the organizational decision-making process toward greater accountability and compliance.

For further proof of compliance, providers would integrate risk management practices into their organization's structure by:

- **Preparing safety and security reports**. These reports document risk assessments, mitigation strategies, and deployment decisions.
- **Embedding accountability**. Responsibility for managing systemic risks extends to the executive and board levels.

Next Steps

As the next step in developing the Code, the Plenary, a group of nearly 1,000 stakeholders that includes EU Member State representatives, industry experts, and international observers, will convene for discussions in four working groups to refine the draft. The chairs of the working groups will gather insights and facilitate interactive feedback sessions, which will be synthesized and presented to the full Plenary to guide three additional drafting rounds. This input will shape adjustments to the Code, ensuring its measures, sub-measures, and KPIs are proportionate to risks, accommodating smaller providers like small and midsize enterprises and open-source developers. The finalized Code, expected to be completed by April 2025, will aim to balance clear compliance requirements with flexibility for technological advancements.

Takeaway

While the Code is still in draft form, it signals the EU's commitment to regulating foundational AI technologies in a detailed and proactive manner. Businesses operating in or interacting with the European market and that may be within scope of the Code should begin assessing how these proposed measures might apply to their GPAI models or downstream applications. Key areas to consider include transparency obligations, such as documentation and acceptable use policies, and readiness for systemic risk management, including monitoring and mitigating potential harms across a model's lifecycle.

Although the Code is not yet final, taking steps now to identify potential gaps in compliance will position covered parties to adapt to the evolving regulatory landscape more readily. Early engagement with these principles may not only mitigate future compliance risks but also demonstrate leadership in responsible AI practices, potentially enhancing both regulatory readiness and market trust.

Authors

Explore more in

Technology Transactions & Privacy Law Privacy Litigation Artificial Intelligence & Machine Learning Communications Advertising, Marketing & Promotions

Related insights

Update

EU Reaches Political Agreement on AI Act, But Questions Remain

Update

States Begin To Regulate AI in Absence of Federal Legislation