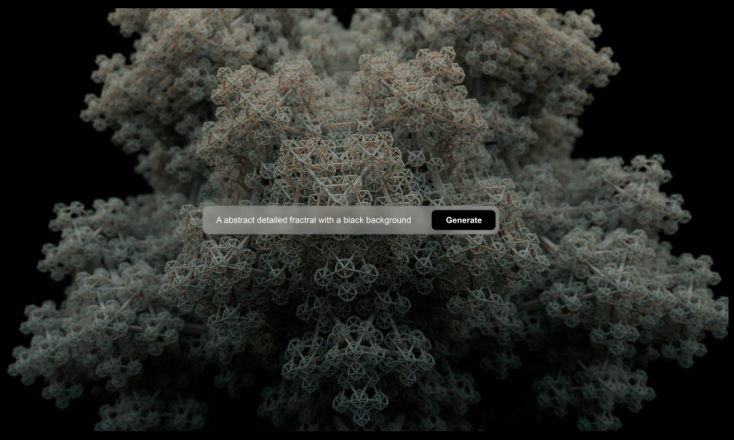
DIGI CON



EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE PUBLISHED FIRST DRAFT OF THE GENERAL-PURPOSE AI CODE OF PRACTICE

The Generative AI (GAI) models have gained significant global attention with the release of OpenAI's ChatGPT in late 2022. This development has marked a turning point, positioning GAI models as a distinct category within the AI discourse. As a result, AI models that had previously been seen as cutting-edge are now often referred to as

‘traditional’ in legal literature, despite being their recent innovation. This situation calls for a clear understanding of the differences between ‘traditional’ AI models and generative AI (GAI) models. The primary distinction lies in GAI’s capacity to perform a wide array of tasks by generating human-like language, rather than being solely limited to producing predictions or scores. Traditional AI models are in principle designed for specialized tasks with specific outputs, while GAI models can create nuanced, contextually appropriate language and responses across diverse scenarios, resembling human communication. In Recitals 99, the AI-Act considers GAI models as a type of general-purpose AI models and identifies two key characteristics, namely (i) *generality*, and (ii) *capability*, which set apart from ‘traditional’ AI systems:

“THE DEFINITION SHOULD BE BASED ON THE KEY FUNCTIONAL CHARACTERISTICS OF A GENERAL-PURPOSE AI MODEL, IN PARTICULAR THE GENERALITY AND THE CAPABILITY TO COMPETENTLY PERFORM A WIDE RANGE OF DISTINCT TASKS. THESE MODELS ARE TYPICALLY TRAINED ON LARGE AMOUNTS OF DATA, THROUGH

VARIOUS METHODS, SUCH AS SELF-SUPERVISED, UNSUPERVISED OR REINFORCEMENT LEARNING" (RECITAL 97).

This distinction has resulted a publication of a Draft Code of Practice by the European AI Office, titled as the "*First Draft General-Purpose AI Code of Practice*" (the "Code"), which has the purpose of providing guidance on compliance with the obligations set forth in Articles 53 and 55 of the AI-Act for providers of general-purpose AI models and general-purpose AI models with systemic risks. The Code is a guiding, dynamic, and a bridge document until the adoption of harmonised EU standards for these types of AI models.

In alignment with the requirements of the AI-Act, the final version of this Code is scheduled for release on May 1, 2025, marking a significant step in shaping the governance framework for general-purpose AI. Created by a diverse group of independent experts from industry, academia, and civil society, the Code embodies a holistic approach to shaping AI governance. It also incorporates insights from international frameworks, highlighting its

commitment to global collaboration and alignment. In general, the Code sets forth guiding principles and objectives, offering stakeholders a clear vision of the potential structure and content of the final version. This marks a pivotal step by the European AI Office, as the advanced capabilities and broad applicability of GAI models bring with them an inherent unpredictability in their outcomes, posing unique challenges in understanding and managing their behavior. Furthermore, the legal concerns surrounding GAI models span a wide spectrum, including issues related to privacy and data protection, liability for outcomes, intellectual property rights, and cybersecurity issues.

The Code conisders these legal concerns and is organized into four key areas: (i) transparency and copyright rules, (ii) risk identification and assessment for systemic risk, (iii) technical risk mitigation for systemic risk, and (iv) governance risk mitigation for systemic risk. These elements work collectively to promote innovation while addressing the ethical and safety challenges associated with AI.

Six high-level guiding principles

underpin the Code. First, it ensures alignment with European Union laws and values, including the Charter of Fundamental Rights. Second, the Code integrates international standards and global best practices, fostering a collaborative approach to governance (Article 56(1) of the AI-Act). Third, it emphasizes proportionality, requiring that measures align appropriately with the level of risk. This principle involves two key tests: (i) suitability and (ii) necessity for the maeasures, both of which leave significant room for interpretation. Fourth, it strives to be future-proof, enabling flexibility and adaptation as AI technologies evolve. Fifth, it provides support for small and medium enterprises (SMEs) by simplifying compliance requirements. Finally, it aims to foster a robust AI safety ecosystem, encouraging transparency and cooperation among various stakeholders.

Transparency and copyright compliance are central to the draft. Providers are required to maintain detailed documentation of their AI models, ensuring accessibility for AI Office and the national competent authorities and downstream providers. Copyright compliance is also a priority,

with providers expected to align with the EU's text and data mining exceptions and to establish clear policies for managing copyright issues. Transparent reporting mechanisms, including provisions for handling copyright claims, are integral to this process.

The draft also outlines a taxonomy of systemic risks, categorizing them by *type*, *nature*, and *source*, which is subject to interpretation in light of the severity and probability of each risk. As a reminder, the concept of risk is defined as "the combination of the probablity of an occurance of harm and the severity of that harm" under the AI-Act (Article 3(2)). The systemic risk, on the other hand, defined as a "risk that is specific to high-impact capabilities of general-puporse AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole" (Article 3(65)). These risks include large-scale discrimination, misuse of AI for disinformation or manipulation, cyber vulnerabilities, loss of control, and chemical, biological, rediological, and nuclear risks. To

address these risks, the draft proposes a comprehensive framework for mitigation strategies, proportionate to the severity and likelihood of each risk.

For general-purpose AI models with systemic risks, providers are also expected to adopt a robust Safety and Security Framework (SSF). This framework will outline risk mitigation policies across the model's lifecycle. Continuous risk assessment is mandated, supported by evidence collection using rigorous methodologies. Providers must also ensure transparent reporting of incidents and systemic risks to the AI Office and, where appropriate, the public. High-severity risks require independent expert assessments both before and after deployment, allowing for necessary adjustments informed by lifecycle monitoring.

Governance measures include assigning responsibility for risk management at the executive and board levels, conducting periodic assessments of adherence to the Code, and facilitating independent evaluations to validate systemic risk mitigation efforts. These steps are designed to enhance accountability

and ensure that organizations prioritize AI safety.

Stakeholders are encouraged to provide feedback on the draft by November 28, 2024. Although the document remains high-level, future iterations will include more detailed measures, key performance indicators (KPIs), and compliance mechanisms. This draft represents a significant milestone in harmonizing AI governance within the EU, balancing the need for safety and transparency with the promotion of innovation.

Sümeyye Elif Biber

Postdoctoral Researcher at the University Luxembourg | Head of Digital Rights of DigiCon at University Luxembourg

Dr. Sümeyye Elif Biber is a legal scholar in European public law and digitalisation at the University of Luxembourg Faculty of Law. She holds a PhD degree (cum laude, highest distinction) from Sant'Anna School of Advanced Studies in Pisa, with her dissertation, "A Rights-Based Inter-Legal Approach to the Fundamental and Human Rights Challenges posed by Artificial Intelligence Systems", supervised by Prof. Gianluigi Palombella. Currently, she is writing a book, "A Rights-Based Inter-Legal Approach to Artificial Intelligence", which is under contract with Hart Publishing. Since 2023 February, she is the Head of Digital Rights of the Digital Constitutionalist in Florence (EUI).