The European Union's AI Act: A Primer

The EU AI Act is a comprehensive legal framework for AI, with broad extraterritorial effect and, hence, a global impact. Officially published on July 12, 2024, it became effective on August 2, 2024. Its primary aim is to regulate AI systems placed on the EU market, balancing innovation with ethical considerations and safety, with the EU aiming to position itself at the forefront of trustworthy AI development.

The Act is being implemented in several key stages:

- Prohibitions on unacceptable risk AI and AI literacy obligations took effect on February 2, 2025,
- Governance rules for General Purpose AI (**GPAI**) apply from August 2, 2025.
- The majority of requirements become enforceable from August 2, 2026.
- Final implementation steps, particularly for the public sector, are slated for 2030.

The EU AI Act possesses broad extraterritorial reach, meaning it applies to providers and deployers regardless of their establishment location if the AI system is placed on the EU market, put into service in the EU, or if its output is used within the EU.

Enforcement mechanisms are robust, with severe sanctions for non-compliance. Fines can reach up to:

- €35 million or 7% of global annual turnover for severe violations (e.g., engaging in prohibited practices),
- €15 million or 3% for moderate violations (e.g., non-compliance with provider obligations), and
- €7.5 million or 1.5% for minor violations (e.g., providing inaccurate information).

Enforcement is carried out by National Competent Authorities (NCAs) within each Member State and the newly established EU AI Office, particularly for GPAI models.

A key feature of the EU AI Act is its detailed hierarchical risk classification system.

Risk Level	Description and Examples
Unacceptable Risk	Practices deemed a clear threat to safety and fundamental rights are **banned**. These include government-led social scoring, harmful manipulative techniques, exploitation of vulnerabilities, untargeted scraping of facial images from the internet or CCTV, and emotion recognition in workplaces and educational institutions.
High-Risk	This is the most regulated category. A system is high-risk if it is a safety component of a regulated product or falls into one of eight critical areas listed in the AI Act's Annex III: ● biometrics; ● critical infrastructure; ● education and vocational training; ● employment and worker management; ● access to essential public and private services (e.g., credit scoring); ● law enforcement; ● migration and border control; and ● administration of justice. Providers face extensive obligations, including risk management, data governance, technical documentation, logging, human oversight, and mandatory conformity assessments.
Limited Risk	Systems like chatbots or those generating deepfakes are subject to **transparency obligations**. Users must be informed they are interacting with an AI or that content is artificially generated.
Minimal Risk	The vast majority of AI applications (e.g., AI-enabled video games, spam filters) are largely **unregulated**, with voluntary codes of conduct encouraged.
GPAI system	**Definition:** An AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently

	performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development, or prototyping activities before they are placed on the market (**Art. 3(63)**). Further, GPAI systems are defined as having "systemic risk" where such models: a. Have "**high impact capabilities**" evaluated on the basis of appropriate technical tools and methodologies; b. Are decided as such by the European Commission that it has capabilities or an impact equivalent to those set out in point (a). Models are presumed to have **high impact capabilities** when the cumulative amount of computation used for its training measured in floating point operations is greater than 10(^25). **Obligations on providers of GPAI models include (see Arts 53 and 54):** a. Maintaining technical document on GPAI model, including its training and testing process and the results of its evaluation; b. Maintaining and making available information and documentation to providers of AI systems who intend to integrate the GPAI model into their AI systems; c. Putting in place a policy to comply with EU law on copyright and related rights; d. Creating and making available a summary about the content used for training the GPAI model according to a template from the AI Office; and e. For providers of GPAI models established in third countries, appointing an authorised representative established in the EU. **Obligations on providers of GPAI models with systemic risk (in addition to obligations listed above) (see Art 55) include:** a. Performing model evaluation in accordance with standardised protocols and tools, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risks; b. Assessing and mitigating possible systemic risks at the EU-level;

	c. Tracking, documenting and reporting (without undue delay) to the AI Office and national competent authorities relevant information about serious incidents and possible corrective measures; and d. Ensuring an adequate level of cybersecurity protection for the relevant model and its physical infrastructure.

Beyond risk categorization, the Act mandates extensive transparency and accountability obligations:

- Providers must ensure users are informed when interacting with an AI system, unless this is obvious or the AI is used for legal purposes.
- AI systems generating synthetic content must mark their outputs in a machine-readable format as artificially generated or manipulated.
- Deployers of emotion recognition or biometric categorization systems must inform affected individuals of the system's operation.
- This information must be provided clearly, distinguishably, and at the latest at the time of the first interaction or exposure.

Accountability is reinforced through requirements for comprehensive technical documentation (see the FPF - OneTrust Guide to Conformity Assessments under the EU AI Act), robust risk and quality management systems throughout the AI system's lifecycle, and automatic event recording for traceability and monitoring.