EUROPEAN LAW & GOVERNANCE SCHOOL
An International University School by EPLO

Navigating the Regulatory and Ethical Challenges of Artificial Intelligence: The European Commission's AI Act and new proposals

Author: Salvatore Gabriele*****

Introduction

Artificial Intelligence (AI) is fundamentally reshaping contemporary society, transforming modes of social interaction, educational practices, economic structures, and systems of governance. The emergence of advanced AI technologies, particularly foundation models and generative AI systems, has introduced unprecedented opportunities for innovation. On the other side, it raises complex ethical, legal, and societal challenges. Within this evolving technological landscape, the European Commission (EC) has positioned itself as a central regulatory authority, seeking to strike a balance between promoting innovation and safeguarding fundamental rights through the development of human-centric, trustworthy AI frameworks. The EU Artificial Intelligence Act[1] (hereinafter "AI Act"), proposed in April 2021 and entered into force in August 2024, represents the first comprehensive attempt globally to regulate AI through a harmonized, risk-based framework. This legislation is further complemented by the Commission's recent proposals, including the establishment of an AI Office and the AI Continent Action Plan, which aim to institutionalize AI governance and solidify Europe's leadership in this transformative domain.

This article aims to analyze and describe the major developments in a concise manner, focusing on the characteristics of digital regulation and its possible legal developments. It offers a critical examination of these developments, focusing on the classification of AI systems by risk, transparency and data governance obligations for foundation models, developers' duties to disclose training data and uphold copyright, and the emphasis on ethical, human-centric AI principles such as explainability, fairness, and human oversight together with its geopolitical implications of a ruled-based AI. It explores the regulatory innovation tension from a legal perspective, scrutinizing intellectual property (IP) rights, liability frameworks and compliance burdens. In conclusion, the article assesses the European Union's strategic vision for AI governance as articulated through its most recent proposals.

REGULATORY CLARITY THROUGH RISK-BASED CLASSIFICATION OF AI SYSTEMS

The AI Act introduces a risk-tiered regulatory regime as a foundational principle to manage AI's multifaceted impacts. This stratification categorizes AI systems into four risk levels: minimal, limited,

---

* EU Law LLM graduated and Contractual Agent of the Legal Service of the European Commission. The views expressed are the author's views and cannot be attributed to the European Commission or its Legal Service.

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

high, and unacceptable. This approach reflects a proportionality principle, ensuring that regulatory requirements correspond with the potential harm associated with AI applications[2].

Minimal and Limited Risk

AI systems deemed minimal risk, such as AI-enabled video games or spam filters, are exempted from stringent regulation, reflecting the negligible potential for societal harm. Limited risk applications, including AI-driven chatbots or emotion recognition tools, must satisfy transparency obligations, such as informing users when they interact with AI-generated content, thus enabling informed consent and mitigating deceptive practices.

High Risk

The core of the AI Act's regulatory architecture pertains to high-risk AI systems. These systems operate in sectors with critical societal implications, including healthcare, transportation, law enforcement, and education. For high-risk AI, the Act imposes comprehensive compliance requirements encompassing data governance, documentation, technical robustness, accuracy, transparency, human oversight, and cybersecurity. The intention is to mitigate risks such as discriminatory outcomes, operational failures, and privacy infringements. Notably, the high-risk classification is not static; it accommodates new AI applications by virtue of delegated acts, thereby providing flexibility to address evolving technological contexts.

Unacceptable Risk

AI systems that pose an unacceptable risk (including manipulative subliminal techniques, social scoring by public authorities, and exploitative uses of vulnerable groups) are categorically prohibited. For AI technologies incompatible with fundamental rights and democratic values of the European Union it represents a normative boundary and an important novelty about digital compliance.

Critical Assessment and guidelines

While the tailored risk-based approach contributes regulatory clarity and proportionality, academic commentary has highlighted inherent challenges. Scholars observe the subjective nature of risk assessment and the difficulty in classifying emerging AI systems with dual-use potential[3]. Moreover, exponential technological evolution and worldwide developing actors to get into the market may outpace regulatory updates, necessitating agile governance mechanism and short timeframes. In February 2025, the European Commission published two sets of guidelines to clarify crucial aspects of the AI Act (Guidelines on the definition of an AI system[4] and Guidelines on prohibited AI practices[5]). They provide

[2] European Commission, Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act), COM (2021) 206 final.
[3] S. Wachter / B. Mittelstadt / L. Floridi, Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation, *International Data Privacy Law* 7(2), 76–99 (2017)
[4] Annex to the Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), Brussels, 6.2.2025 C(2025) 924 final.

guidance on the set of AI Act obligations that started to apply on February 2, 2025 – which includes the definitions section of the AI Act, obligations relating to AI literacy, and prohibitions on certain AI practices.

TRANSPARENCY, DATA GOVERNANCE, CYBERSECURITY AND SAFEGUARDS FOR FOUNDATION MODELS

The foundation models are massive pre-trained AI systems capable of performing a wide array of downstream tasks. Their increase, in terms of number and use, introduced profound regulatory and ethical considerations[6], mostly concerning fundamental rights and extraterritoriality dimension challenges. These models, exemplified by GPT (Generative Pre-trained Transformer) architectures and other large-scale language or vision models, are trained on enormous datasets encompassing diverse, often unvetted sources[7].

Transparency Obligations

The European Commission's attempt to discipline also foundation model[8] emphasize transparency requirements, compelling developers to disclose capabilities, limitations, intended and foreseeable uses, and potential risks. This aligns with the principle of meaningful human control, allowing users and regulators to assess AI-generated outputs.

Data Governance

Data governance is paramount given the scale and opacity of datasets underpinning foundation models. The Commission proposes rigorous requirements to ensure datasets respect fundamental rights, including data protection norms under the GDPR, copyright laws[9], and ethical standards. The risk of copyright infringement[10] - through training on protected works without consent - raises thorny legal questions regarding the legality and ethics of large-scale data scraping[11] and creates legal uncertainties that may inhibit innovation or foster liability risks[12]. Quality control measures must also address data representativeness to mitigate algorithmic bias and ensure fairness, a concern echoed widely in AI ethics literature[13].

Cybersecurity

---

[5] Communication to the Commission Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), Brussels, 4.2.2025 C(2025) 884 final.
[6] S. Much / M.C. Borrelli / K. Charles, The EU AI Act as global artificial intelligence regulation (2023).
[7] R. Bommasani et al., On the opportunities and risks of foundation models, *arXiv preprint* arXiv:2108.07258 (2021).
[8] R. Bommasani / A. Hau / K. Klyman / P. Liang, Foundation models under the EU AI Act, Center for Research on Foundation Models – People, Report, Research, Policy, Blog (2024).
[9] EUIPO, The development of generative artificial intelligence from a copyright perspective, 2025.
[10] J.P. Quintais, Generative AI, copyright and the AI Act, *Computer Law & Security Review* 56 (2025).
[11] P. Samuelson, Copyright and AI: Charting a course for innovation and creativity, *Journal of Intellectual Property Law* 27(1), 1-30 (2020).
[12] D. Gervais, The machine as author, *Iowa Law Review* 107(1), 2053–2106 (2021).
[13] S. Barocas / M. Hardt / A. Narayanan, *Fairness and Machine Learning*, MIT Press (2023).

Foundation models' susceptibility to adversarial attacks, data poisoning, and other cyber threats necessitates robust cybersecurity safeguards. The Commission advocates for continuous monitoring, incident reporting, and resilience strategies to prevent and mitigate manipulation risks, reflecting a growing consensus that AI cybersecurity is integral to trustworthy AI deployment[14].

Safeguards Implementation

Beyond legal compliance, developers need to institute technical and organizational safeguards to prevent misuse, including detection of harmful or biased outputs, content moderation, and mechanisms for human intervention[15]. These obligations resonate with broader trends in AI ethics, emphasizing risk mitigation and harm prevention. However, the feasibility of auditing and enforcing these safeguards on opaque, self-evolving models remains a significant challenge, inviting further research and regulatory innovation.

ETHICAL AND HUMAN-CENTRIC AI: EXPLAINABILITY, FAIRNESS, NON-DISCRIMINATION, AND HUMAN OVERSIGHT

The ethical governance of AI, deeply embedded in European policy discourse, is foregrounded in the AI Act and related proposals, which strive for explainable, fair, non-discriminatory and subject to human oversight AI systems[16].

Explainability

Explainability, or the ability of an AI system to provide understandable and meaningful accounts of its decisions, is a precondition for trust and accountabilit[17]. However, researchers emphasize the technical and epistemological difficulties of achieving genuine explainability, especially in deep learning systems with millions of parameters[18]. Regulatory mandates must therefore be calibrated to the capabilities and limitations of current technology.

Fairness and Non-Discrimination

AI systems have repeatedly been shown to replicate or exacerbate social biases, particularly along lines of race or gender[19]. The AI Act mandates rigorous testing and mitigation strategies to address these risks, reflecting established ethical frameworks such as the AI's guidelines.

Human Oversight

---

[14] M. Brundage et al., Toward trustworthy AI development: Mechanisms for supporting verifiable claims, *arXiv preprint* arXiv:2004.07213 (2020).
[15] L. Floridi / J. Cowls, A unified framework of five principles for AI in society, *Harvard Data Science Review* (2019).
[16] European Commission, Ethics Guidelines for Trustworthy AI, 2019.
[17] F. Doshi-Velez / B. Kim, Towards a rigorous science of interpretable machine learning, *arXiv preprint* arXiv:1702.08608 (2017).
[18] C. Rudin, Stop explaining black box models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence* 1(5), 206-215 (2019).
[19] J. Buolamwini / T. Gebru, Gender Shades: Intersectional accuracy disparities in commercial gender classification, *Proceedings of Machine Learning Research* (2018).

Ensuring human-in-the-loop or human-on-the-loop supervision is crucial for accountability, particularly in high-risk settings such as healthcare or law enforcement[20]. The Act's requirements for human oversight aim to prevent autonomous, unaccountable AI decision-making, preserving fundamental rights and democratic control.

INNOVATION VERSUS REGULATION: LEGAL CHALLENGES IN INTELLECTUAL PROPERTY, LIABILITY, AND COMPLIANCE

The tension between promoting AI innovation and imposing regulatory constraints remains acute, with significant legal challenges implicated.

Intellectual Property Ownership

The ownership and protectability of AI-generated outputs are unsettled areas of law. Questions persist about whether AI systems can be considered "authors" or inventors under copyright or patent law, and about the implications for rights in AI-created works[21]. The training process itself, involving vast copyrighted inputs, complicates this landscape, requiring a re-examination of copyright doctrines such as fair use and data mining exceptions[22].

Liability Frameworks

Existing liability regimes, based primarily on human agency, face challenges when applied to autonomous or semi-autonomous AI systems[23]. The Commission acknowledges the need for adapting liability laws to clarify responsibility for harms caused by AI, whether attributable to developers, deployers, or users. Debates continue over strict versus fault-based liability models and the role of mandatory insurance schemes[24].

Compliance Burden

The AI Act's extensive documentation, auditing, and transparency obligations impose significant compliance costs, potentially disproportionately impacting startups and SMEs. Critics warn that excessive regulatory burden could stifle innovation and entrench incumbents. Balancing protection and dynamism is therefore a core policy challenge, following a (necessary) interdisciplinary approach[25].

NEW PROPOSALS: THE AI OFFICE[26] AND THE AI CONTINENT ACTION PLAN[27]

[20] J. Whittlestone / R. Nyrup / A. Alexandrova / K. Dihal / S. Cave, The role and limits of principles in AI ethics: Towards a focus on tensions, *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2019).
[21] R. Abbott, *The Reasonable Robot: Artificial Intelligence and the Law*, Cambridge University Press (2020).
[22] P. Samuelson, *ibidem*.
[23] U. Pagallo, *The Laws of Robots: Regulating Autonomous Artificial Agents*, Springer (2018).
[24] J. P. Quintais, *ibidem*.
[25] H. Zhong, Implementation of the EU AI act calls for interdisciplinary governance, *AI Magazine* 45, 333–337 (2024).
[26] The AI Office is established within the European Commission as the centre of AI expertise and forms the foundation for a single European AI governance system.

Recognizing the complexity of AI governance, the European Commission has established a dedicated AI Office, envisioned as a centralized body tasked with overseeing compliance and enforcement of the AI Act across member states, facilitating cooperation among national authorities and stakeholders, providing guidance on regulatory standards and best practices and supporting innovation ecosystems through coordination and funding. Parallelly, the AI Continent Action Plan aims to consolidate Europe's strategic investments in AI research, skills development, infrastructure, and ethical standards. It seeks to balance technological sovereignty with global cooperation, positioning Europe as a leader in responsible AI innovation and governance.

These proposals reflect a forward-looking vision to manage AI's omnipresence comprehensively[28], yet they also raise questions regarding institutional capacity, democratic legitimacy, and adaptability to fast-paced innovation.

CONCLUSION

The European Commission's AI Act and its associated proposals represent an ambitious and pioneering regulatory blueprint designed to ensure that AI technologies develop in harmony with European values, fundamental rights, and societal welfare. By adopting a risk-based classification, establishing rigorous transparency and cybersecurity standards for foundation models, and mandating developer accountability, the EU sets not only a European but a global benchmark for responsible AI governance. However, the persistent tension between regulatory oversight and the need to foster innovation[29] remains a formidable challenge, particularly with respect to intellectual rights, liability rules, and a new potential Brussels effect around the world, not only limited to our Continent.

The creation of an AI Office and the articulation of the AI Continent Action Plan indicate a strategic recognition that regulatory frameworks must be accompanied by institutional and policy infrastructures to steer AI's future and to not lag behind the USA and China. Ultimately, the success of these initiatives will depend on both legal innovation and agile governance mechanisms that can adapt to AI's rapidly evolving landscape, ensuring that AI serves humanity's best interests without compromising innovation and competitiveness in Europe.

---

[27] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (AI Continent Action Plan), Brussels, 9.4.2025 COM(2025) 165 final.

[28] Another important initiative of the European Commission is the Code of Practice for General-Purpose AI (forthcoming).

[29] P. Csernatoni, The EU's AI Power Play: Between Deregulation and Innovation, *Carnegie Europe* (2025).