Regulating Generative AI towards the future¹

Giovanna De Minico* - Michela Tuozzo*

Abstract

This intervention is focused on two issues. The first one aims to consider the current regulatory framework of Generative Artificial Intelligence systems, with specific attention to the obligations of providers and deployers and system governance as dictated in the AI Act.

The second issue is dedicated to exploring points of intersection with other regulations applicable to AI systems within the European digital ecosystem.

Keywords

Generative AI systems, AI Act, Governance

1. Introduction

Generative intelligence, a cutting-edge development in the realm of artificial intelligence, has significantly influenced the legislative trajectory of the European Regulation on Artificial Intelligence - 2021/0106(COD).

Al systems utilizing Large Language Models have the capacity to generate a wide range of outputs, including texts, translations, images, sounds, videos, and more. The prospect of these systems harmoniously integrating with other AI systems amplifies their usefulness for users, both professionals and non-professionals, as well as for public and/or judicial authorities. The latter can leverage them for forecasting, adopting recommendations, or making informed decisions.

The unique characteristics of generative AI have raised questions about the applicability of the traditional risk management approach, which forms the foundation of European technological Regulation, and how to effectively categorize this new form of AI.

The proposed intervention aims to highlight the status of generative AI under the AI Act and its governance. Consequently, critical reflections will be elaborated upon regarding these elements and their potential implications on a crucial issue: hate speech and online misinformation.

2. GPAIs' classification

The critical characteristics of general-purpose AI (GPAI) models include their large size, opacity, and potential to develop unexpected capabilities beyond those intended by their creators. According to article 3 (63), a general-purpose AI model means an AI model trained with a large amount of data using self-supervision at a 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.

On December 6, 2022, the General Secretariat of the Council classified GPAIs as high-risk systems. This classification enforces specific compliance obligations (Articles 10-15), requires a risk impact assessment on fundamental rights (Article 27), mandates auditing prior to market entry (Article 43), and necessitates registration in the EU database, along with postmarket surveillance obligations.

¹University of Naples Federico II, via Marina 33, Naples, 80125, Italy

² University of Naples Federico II, via Marina 33, Naples, 80125, Italy

¹ The article reflects collective thoughts; however, the paragraphs can be attributed as follows: paragraphs 1, 2, 2.1, and 3 to Dr. Michela Tuozzo; paragraphs 2.2 and 4 to Prof. Giovanna De Minico.

The high-risk classification of GPAIs has sparked two types of criticism: the failure to adhere to a precautionary approach and the imposition of obligations that are perceived as difficult to achieve.

Through its amendments on June 14, 2023, these critical aspects led the European Parliament to develop an autonomous classification of GPAIs as foundation models. Article 28 *ter* outlined three categories of obligations for the provider: risk identification and mitigation, testing and evaluation, and documentation.

However, it's important to note that the regulatory model was significantly altered in the final version of the adopted AI Act, a result of the compromise reached among the European institutions during the trilogue phase. This change was influenced by lobbying efforts (Bareis), underscoring the political dynamics at play in the regulatory process.

2.1. Tiered approach

The Regulation of generative intelligence became its category in the final version of the AI Act approved on March 13, 2024. Within this category, generative AI is classified into three types: general-purpose AI model, general-purpose AI model with systemic risk, and open-source general-purpose AI model (articles 51-55). As a result, we have different rules for different GPAIs.

In essence, Spanish Presidency of the EU Council has aimed to strike a balance between the Council's hands-off approach and the Parliament's earlier stance of establishing uniform rules for all generative AI systems.

Specifically, providers of "standard" GPAIs, when generating synthetic audio, image, video, or text content, must ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated (Article 50, paragraph 2). In addition providers shall (Article 53): draw up and keep up-to-date the technical documentation; comply with Union copyright law; make a sufficiently detailed summary of the content used for training publicly available.

Conversely, providers of open-source GPAIs must only comply with copyright rules and those regarding the synthesis of content used for training.

The obligations for providers of GPAI with systemic risk are more extensive than for "standard".

In addition to those already stipulated for "standard" GPAIs, providers shall perform model evaluation to identify and mitigate systemic risk; assess and mitigate possible systemic risks at the Union level; report without undue delay to the AI Office and, as appropriate, to national competent authorities, relevant information about severe incidents and possible corrective measures to address them; ensure an adequate level of cybersecurity protection.

2.2. Governance²

The governance of the artificial intelligence market is complex due to AI's diverse applications in both public and private sectors across national and EU levels. It is important to note that responsibility for actions like fundamental rights impact assessment and conformity evaluation falls on the entrepreneur's initiative, reflecting a confusing blend by the Commission of individual centrality and a trend toward system privatization.

The governance system could have taken three forms: complete decentralisation, relying on national oversight systems like telecommunications, complete centralised supervision disregarding national variations, or a mixed system with tasks entrusted to the Commission in a designated Directorate-General and to European agencies. The governance of the AI Act takes different forms depending on the level considered and even the type of intelligence, as shown by the case of GPAIs. Examples of entirely European Independent Authorities, independent from National Governments and the Commission, were established in 2010 with the European Banking Authority, the European Insurance and Occupational Pensions Authority, and the European Securities and Markets Authority.

While traditional AI systems have an intermediate governance form balancing decentralisation and centralisation, generative AI systems follow a fully centralised approach. This led to a complex governance structure: the original proposal involved three authorities (Commission, National Authorities, and AI Board), but the final act expanded to at least 5 (including the Advisory Forum and Scientific Panel).

The entire AI system obediently follows a government-centric approach. It is justified solely by the fact that AI will become the engine of public policies that must firmly remain in the hands of the current political majorities.

² This part is due to Professor Giovanna De Minico.

As for the Commission, it has the power to have the final and definitive say on corrective measures proposed by national supervisory authorities, confirming an approach centred on the community executive

As for the AI Board, it is ensured only functional independence; indeed, genetic independence from the Commission and Member States is not required. Evidence of this is the freedom of each State to send whomever they want and the presence of the Commission on the Board. Consequently, even functional independence is at risk, as already evidenced by the comparison between European Parliament amendments (Article 56) and the final text of Article 65. This rule states: "The Board shall be organised and operated to safeguard the objectivity and impartiality of its activities". The original formulation of Article 56 stated: "The 'European Artificial Intelligence Office' (...) shall be an independent body of the Union. It shall have legal personality" and the Office "act independently when carrying out its tasks or exercising its powers" (Article 56 quater). Currently, in terms of its structure, the Office is integrated within the administrative framework of the Directorate-General Communications Networks, Content and Technology (DG-CNECT) of the Commission. It does not have operational autonomy from DG-CNECT. Furthermore, national authorities, no dedicated infrastructures or technical, financial, or human resources are provided.

Regarding National Supervisory Authorities, it is permissible for the State to designate them within an affiliated entity such as the Government (as emphasised in the Privacy Commissioner's letter dated March 25, 2024).

In Italy, the *agencification* approach is confirmed by the legislative initiative on a delegated law regarding artificial intelligence approved by Council of Ministers on April 23, 2024. In the draft, Article 18 designates the two Authorities: the Agency for Digital Italy (AgID) and the National Cybersecurity Agency (ACN). In both cases, these government agencies achieve functional independence only with respect to the regulated entities but not the representative political body. The Government's choice seems clear: to maintain control over "intelligent policies" in its hands, rejecting the model of independent authorities.

In addition to functional independence, organisational independence was expressly requested in the Parliament's amendments (Article 59, par. 1, EP).

Article 70 (AI Act adopted) only recognises functional independence, which seems like a

superficial rule because it is unclear how to ensure functional independence without first guaranteeing organisational independence, known as genetic independence. Additionally, Article 70 does not specify which entities should respect this independence. Previously, Article 59, par. 4, stated that "members of each national supervisory authority, (...), shall neither seek nor take instructions from anybody and shall refrain from any action incompatible with their duties". Removing this part of the rule suggests that independence is only aimed at those being regulated, not political representatives. This suggests that the AI Act has accepted a partial risk of capture because it only focuses on regulating solid entities.

The National Authority possesses regulatory powers, allowing it to mandate actions such as suspensions, corrective measures, or removing AI systems from the market. Could personal ablation be constructed as a complex administrative action with unequal powers? To this question, we respond affirmatively because the National Authority has the authority to propose the action, with the Commission having the final say. Therefore, two authorities - the National and the Community - intervene in the same decision but at different times and with different contributions: one proposes, and the other finalises the procedure. This procedural collaboration achieves coordination between authorities operating simultaneously within the National Union network.

In a summary overview, the governance of the AI Act reserves central political authority for the Commission, whose *verbum* is communicated downstream to the National Authorities. Then, it returns to the Commission itself, with an AI Board intervening occasionally to address gaps in the discourse.

In contrast to the system described above, the generative system governance is centralised between the AI Office and the Commission.

The Commission shall have exclusive powers to supervise general-purpose AI models and request measures and shall entrust the implementation of these tasks to the AI Office, a European Agency of the Commission.

The AI Office plays a central role in developing a Code of Practice and monitoring its application.

The Scientific Panel serves as a qualified advisory body to the AI Office.

Unlike traditional AI models, governance for generative models is fully Eurocentric to the extent that the Commission fulfils its administrative role and possesses law enforcement powers. The centralisation of governance is at its maximum.

While centralisation under the Commission is justified by the sector's sensitivity and the need for a unified implementation approach, it raises concerns about deviating from the independent European authorities established in 2010.

3. Boundaries

3.1. Disadvantages of the AI Act

The qualification of so-called 'systemic risk' will initially depend on the capability, either based on a quantitative threshold of the cumulative amount of compute used for its training measured in floating point operations (FLOPs set as 10^25) or based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel. It is presumed that a model trained with large amounts of data and advanced complexity has foreseeable adverse effects on public health, safety, public security, fundamental rights, or the society as a whole that can be propagated at scale across the value chain.

Upon closer examination, this definition of systemic risk consider the combination of the probability of a harmful event occurring and the severity of that harm, as well as the values and assets of constitutional relevance. This appears consistent with the framework of the AI Act, which prohibits all AI systems whose use is deemed unacceptable because it contradicts Union values. However, there needs to be coherence between the means, represented by numerical indicators such as FLOPs, and the end, which is the protection of common constitutional values. These parameters «describe the foundational model but not its impact on society, safety or fundamental rights» (Helberger et al.).

Another criticism concerns the quantitative and qualitative reduction of provider obligations, even for GPAIs with systemic risk. It is surprising to note that for such models, the activity of demonstrating compliance before placing the AI system into the market or service is not accompanied by guarantees of a prior conformity assessment for the provider and a fundamental rights impact assessment for the deployer. Compliance can be demonstrated by relying on codes of practice within the meaning of Article 56 until a harmonised standard is published.

The AI Office, in collaboration with the Board, encourages the drafting of the Code. They aim to ensure that the Codes of Practice comprehensively address the obligations in Articles 53 and 55. However, all providers of general-purpose AI models

and relevant national competent authorities will intervene. Civil society organisations, industry, academia, and other relevant stakeholders, such as downstream providers and independent experts, may also support the process (Article 56).

Co-regulation in the technology sector has shown various reasons for fallibility, with the primary concern being the risk of capture by regulated solid entities

Lastly, there is an emphasis on the responsible behaviour of providers but a need for proper division of responsibility with 'downstream' users. It should be considered that the distribution of responsibilities along the value chain should involve multiple parties, each with different responsibilities, particularly considering the user's role depending on whether they use the output for professional purposes (Hacker, Engel, Mauer).

3.2. Addressing Challenges Within and Beyond the AI Act

Beyond the regulatory aspects addressed in the AI Act, the proliferation of this new type of AI also presents interpreters with the issue of the rapid pace of technological transformations. From a constitutional law perspective, it is essential to clarify the categories involved to assess whether the discipline outlined in the article respects fundamental liberties.

Consider that the prerogative of the most widespread GPAIs – such as chat GPT – is communication.

We need to raise the following questions: Does generative AI produce ideas? Is it a new form of media? Or a digital private communication?

Article 15 of the Constitution protects the limitation of the freedom of communication by the guarantees of legal reservation and jurisdiction, but with a significant difference compared to Article 21. The limits expressed (good conduct) and unexpressed (protection of personality rights such as reputation and privacy) would not apply to communications as they do to Article 21. Furthermore, protecting freedom and correspondence from undue interference would extend to the recipient and the sender (in our case, OpenAI, Google, and others).

The constitutional coverage of Article 21 would imply extending the guarantees of the press medium to chat GPT as well: the prohibition of censorship, the possibility of adopting inhibitory acts with the guarantees of legal reservation and jurisdiction, and finally, the possibility of limiting its contents.

Finally, similarly to the issue of advertising information, it could be argued that informational content has an economic purpose. Therefore, it would be more appropriate to adopt the limits of Article 41. Thus, it cannot be carried out in contrast to social utility or in a way that harms health, the environment, security, freedom, or human dignity. Upon closer examination, the protection here is also twofold: towards the end-user citizen and other commercial operators.

Providing a definitive answer to the question of constitutional coverage requires further examination. It is essential to consider the communicative context and the communicating subject: We are within the protective sphere of Article 15 when the intention is to maintain the secrecy of the content of virtual correspondence, the recipients are specific and immutable, and the means are suitable for achieving secrecy. It is precisely this last requirement that leans towards the category of Article 21.

When considering the fine line with Article 41, we must examine the purpose of freedom of expression. If it serves an economic aim, such as profiling, then the broader protection of Article 21 may not apply. (Ruffolo).

4. Next steps: Implementing the Regulatory Framework in the Digital Ecosystem³

Upon closer examination, we found that adopting the AI Act only addresses some of the issues raised by generative AI systems.

It could be a reason for adding the general discipline of the AI Act with some particular disciplines, such as Digital Services Act – Regulation 2022/2065 – and Digital Market Act – Regulation 2022/1925, which follow a community goal: to create a single digital ecosystem.

This cumulation operation cannot happen automatically, but it should be conducted with some questions showing us how to integrate the disciplines.

1) Is there a difference in the passive legitimates of the two disciplines? By passive legitimates, we mean the recipients of the rules of the Digital Service Act and the AI Act. The former are the platforms that are identified in the discipline because of a quantity: "several average monthly active recipients of the service in the Union equal to or higher than 45 million" (Article 33, DSA).

The criterion for identifying GPAIs at systemic risk relates to computational capacity: "when the cumulative amount of computation used for its training measured in floating point operations is greater than 10(^25)" (Article 51 AI Act). Unlike for the DSA in this case, the number of clients is disregarded because the taxable entity is identified because of the capacity to input data and occupy the data world.

The second difference is an objective one that concerns the service. Platforms render an intermediary service: they connect those who generate information with those who receive information, and this encounter between supply and demand happens on the platform. So, the platform does not put its hand on the information; it hosts it, rationalizes it, organizes it, and categorizes it, but someone else is generating the idea. Generative AI, on the other hand, does precisely what the platform does not do; that is, it occupies that space vacated by the platform because the AI is not a host; it is the author of an idea of its own can argue about whether Chat-GPT creates the idea out of thin air, or whether the idea is generated by fishing around the network, how it articulates it, and so on but it is still something that involves creative energy. Even though Chat-GPT, unlike the human mind, does not create from anything but from a background, it still realizes a vital, active, innovative contribution that is not there in the platform.

So I have come to say that there is a subjective and objective difference, and because of this difference, to be strict, the discipline of DSA cannot be applied.

2) If we combine all the regulations, they would still fall under the guidelines of the AI Act, which has deliberately exempted GPAI. GPAI was only included in the final negotiations as a last resort, with less strict rules compared to other forms of intelligence. Essentially, GPAI has been shielded from excessive regulation. Those advocating for additional rules are going against the original purpose. Since EU laws are meant to be interpreted beyond just the words, this combination of regulations would contradict the *intentio legis* of the AI Act, burdening GPAI with regulations it was meant to avoid. While we might regret not imposing more regulations, it is worth acknowledging that GPAI has been protected as intended.

Finally, let us ask what the purpose of DSA is. The DSA is for keeping the net clean of blatant malfeasance, misleading, hate speech, fake news, and

³ This part is due to Professor Giovanna De Minico.

so on. Provided that the DSA is not intended to derogate from the general principle that the platform is not the editor-in-chief of a newspaper, it is exempt from a prior and general control obligation, but it does put in place a punctual control obligation, thus not generalized, and not *ex-ante* but *ex-post*. This obligation of control signifies that the ultimate goal is to hold together a control that does not impose a generalized vigilance to which no platform will ever want to submit, surrendering fundamental freedoms, thus the right to say and have expressed those who then put news on the platforms.

Suppose we have a positive answer on positive DSA's purpose. In that case, it could be shifted to the Chat-GPT as well? Moreover, why not if this policy of cleaning up the network is so positive, even if some (De Minico d) see it as a form of censorship entrusted to the private entity? When the Commission designates platforms as "providers of large online platforms" (Articles 15 and 33 DSA), they become the subjects of a timely and subsequent obligation to control the information stored and transmitted by their platform to ensure a transparent and secure digital environment.

The fulfilment of this selective cleanup duty-which reconfirms the absence of a generalized duty of control by object and, over time, according to the philosophy of the e-Commerce Directive, now reproposed in Art. 8 DSA, should equalize the asymmetrical relationship between the platform owner and the author of the hosted content, just as it should put the author of the content and its recipient, the end user of the information flow, on the same level.

Without prejudice to my doubts about the suitability of this asymmetrical measure to equiordize misaligned social partners, I would instead call attention to a possible effect that could affect platforms. These, in order not to incur a liability judgment due to the maintenance of illicit content online, will be inclined to delete rather than preserve the ideas of others, based, moreover, on summary evaluations pronounced inaudita altera parte. Adding to this is a further consideration: the lack of an abstract and general definition of the concept of false leaves platforms free to confuse false with politically inappropriate news or news that does not conform to dominant thinking. No more complete is the prohibition of hate speech, which lacks a prior typification of hate speech, even regarding the necessary causal link between the saying and the activating effect toward the recipients of the prohibited conduct sought to be solicited. This normative gap in the DSA points back to tautological reasoning that hateful conduct is prohibited because it is prohibited. Rather, tautology can conceal dangerous liberticidal theses and easy slides toward only permissible speech: state speech. Thus, the control of permissibility conceals insidious forms of merit-based scrutiny of the manifestation of thought, opening the door to digital censorship on the Web.

This undergoes a radical change from a place unscathed by heteronomous interventions and free from information intermediaries to a space supervised by private individuals who have the keys to open and close the information agora in their hands. This risk cannot be avoided due the absence of the normative definition of falsehood, with the paradoxical consequence of a caesura between the offline environment, where the dissemination of false ideas does not constitute a crime unless it attacks other goods-interests other than the truth, and the virtual one, where instead the idea if false ceases to be the exercise of a right to become an illicit fact.

Therefore, I do believe that the DSA has aggravated the limit of lawfulness, making certain conduct that is lawful offline unlawful when it sees the playing field changed.

The most severe thing about this blank endorsement to platforms of the power to control the merit of others' ideas is the emergence of an unseen function assigned to private individuals, which in the real world does not even exist in terms of power referred to as a public authority. I prefer to trust in the beneficent virtues of the marketplace of ideas, which, in allowing the coexistence of the false with the true, lets citizens distinguish between the two entities because it believes in their ability to mature a responsible idea without being pre-addressed by those who claim to know for them what objective truth is.

This failure of the D.S.A. to predetermine the concept of forgery degrades abstractness into concreteness and generality into particularity. The law is resolved in the ordinal provision of the private strongman, while the equality of citizens before the law is attenuated in the concrete provision *ad certam personam*.

Adopting instead the point of view of those who consider it positive and desirable the extension of its contents to the A.I. Act, then this operation cannot be carried out by making an automatic addition of disciplines but can be applied in interpretation.

3) Who does the interpretation? The Commission has governance over the GPAIs and, therefore, could impose on them obligations arising from the digital service, not by automatic acquisition (Hacker, Engel, Mauer; Botero Arcila) but according to reasoning by

analogy *legis*, assuming that between the two cases, the regulated, the digital and the unregulated Chat-GPT, via is the identity of facts. At the end of this discussion, we have partially applied the rules of the DSA to chat, but only by interpretation and assuming that scrupulous reasoning is conducted.

While maintaining a negative judgment on the content of the DSA, adopting instead the perspective of those who consider it positive and desirable to extend its contents to the AI Act, then such an operation cannot be carried out by automatically adding disciplines (Hacker, Engel, Mauer; Botero Arcila), but can be applied interpretively.

Who does the interpretation? The Commission, which has governance over GPAIs and could therefore impose obligations on them arising from the DSA, not through automatic acquisition but through analogical legal reasoning, admitting that between the two cases, the regulated digital pathway and the unregulated GPAI, there is a factual identity.

At the end of this discussion, we have partially applied the rules of the DSA to chat, but only interpretatively and provided that a careful reasoning is conducted.

Similarly, one should reason about the extension of the contents of the DMA.

In conclusion, Regulations in the digital environment provide solid foundations for addressing the risks posed by the rapid evolution of generative AI. Additionally, it is crucial for the European Commission to fully utilise its implementation power during the execution phase (Articles 290 and 291 TFEU), adapting the pace of technology to existing rules to protect fundamental rights.

Acknowledgements

The initiative falls within the scope of the FAIR Project FAIR (Future Artificial Intelligence Research), WP 3.8 – Resilient AI, Ethical - Legal and Societal Issues in Resilient AI Systems coordinated by Professor Giovanna De Minico.

References

- J. Bareis. "BigTech's Efforts to Derail the AI Act". VerfBlog, 2023/12/05.DOI: 10.59704/265f1afff8b3d2df
- [2] B. Botero Arcila. "Is it a Platform? Is it a Search Engine? It's Chat GPT! The European Liability Regime for Large Language Models". Journal of Free Speech Law, 3 (2023), pp. 455-488.

- Available at SSRN: https://ssrn.com/abstract=4539452
- [3] A. Cefaliello, M. Kullmann. "Offering False Security. How the Draft Artificial Intelligence Act Undermines Fundamental Workers Rights". European Labour Law Journal 4(2022), pp. 542-562.
- https://doi.org/10.1177/20319525221114474
 [4] G. De Minico *a*. "Too many rules or zero rules for the ChatGPT?" BioLaw Journal 2(2023), pp. 491-501. https://doi.org/10.15168/2284-4503-2723
- [5] G. De Minico *b.* "Una norma unica che tenga insieme ChatGpt e privacy". Il Sole 24 ore, June 14, 2023.
- [6] G. De Minico *c*. "The Challenge of the Virtual World for Independent Authorities". European Public Law 29:1(2023), pp. 1-26.
- [7] G. De Minico d. "Internet e le sue regole". Atti del Convegno Processi politici e nuove tecnologie tenutosi a Università degli Studi di Bari Aldo Moro nel 22 giugno 2023, edited by M. Calamo Specchia, Giappichelli, Torino, forthcoming.
- [8] G. De Minico e. "Nuova tecnica per nuove diseguaglianze. Case law: Disciplina Telecomunicazioni, Digital Services Act e Neurodiritti". Federalismi.it 6(2024).
- [9] L. Floridi, M. Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences". Minds & Machines 3 (2020), pp. 681–694. https://doi.org/10.1007/s11023-020-09548-1
- [10] P. Hacker. "AI Regulation in Europe: From the AI Act to Future Regulatory Challenges". Oxford Handbook of Algorithmic Governance and the Law edited by I. Ajunwa, J. Adams-Prassl, Oxford University Press, 2024.
- [11] P. Hacker, A. Engel, M. Mauer. "Regulating ChatGPT and other Large Generative AI Models". FAccT 7(2023) pp. 1112-1123 https://doi.org/10.48550/arXiv.2302.02337.
- [12] N. Helberger, N. Diakopoulos. "ChatGPT and the AI Act". Internet Policy Review. Journal On Internet Regulation 12:1 (2023), https://doi.org/10.14763/2023.1.1682.
- [13] C. Novelli, P. Hacker, J. Morley, J. Trondal, L. Floridi. "A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities," May 5, 2024, available at SSRN: https://ssrn.com/abstract=4817755 or http://dx.doi.org/10.2139/ssrn.4817755
- [14] U. Ruffolo. "Piattaforme, A.I. generativa e libertà di (formazione e) manifestazione del pensiero. Il caso ChatGPT". Giurisprudenza italiana (2024), pp. 472-480.