Gavriil E, Pavlidis G, 'The fog of information: the EU AI Act and legal strategies against AI-fuelled disinformation' (2025) 2 EJPLT.
DOI: https://doi.org/10.57230/EJPLT252EGGP

EJPLT
ISSN 2704-8012

The fog of information: the EU AI Act and legal strategies against AI-fuelled disinformation

La nebbia dell'informazione: l'AI Act dell'UE e le strategie giuridiche contro la disinformazione alimentata dall'IA

ELENI GAVRIIL
Lecturer in International Economic Law and Human Rights
Neapolis University Pafos

GEORGIOS PAVLIDIS
Associate Professor of International and EU Law, UNESCO Chair
Neapolis University Pafos

Abstract

The rise of generative AI has complicated longstanding challenges in the fight against disinformation, enabling the creation and spread of synthetic content at unprecedented speed and scale. While disinformation is not new, AI's ability to impersonate credible sources, manipulate audiovisual material, and craft persuasive narratives across languages raises urgent legal and regulatory concerns. This article examines the provisions of the EU Artificial Intelligence Act (AI Act), which can provide innovative responses to AI-driven disinformation. The AI Act does not address disinformation in a direct manner, but it establishes mechanisms that are relevant to this emerging problem. The article situates the AI Act within the broade EU regulatory landscape, which also includes the Digital Services Act, the Code of Practice on Disinformation, and the EU Charter of Fundamental Rights. The focus is on the risk-based classification and transparency rules of the AI Act, which can enhance democratic resilience, if they are paired with the necessary safeguards.

This study was prepared in the context of the Jean Monnet Centre of Excellence AI-2-TRACE-CRIME (Project 101175740) and was funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

Abstract
L'ascesa dell'intelligenza artificiale generativa ha complicato sfide già note nella lotta contro la disinformazione, rendendo possibile la creazione e la diffusione di contenuti sintetici con una velocità e una portata senza precedenti. Sebbene la disinformazione non sia un fenomeno nuovo, la capacità dell'IA di imitare fonti credibili, manipolare contenuti audiovisivi e costruire narrazioni persuasive in più lingue pone interrogativi giuridici e regolatori urgenti. Questo articolo analizza il Regolamento europeo sull'intelligenza artificiale (AI Act), che rappresenta una risposta innovativa, sebbene indiretta, alla disinformazione generata dall'IA. Pur senza affrontare esplicitamente il fenomeno, l'AI Act introduce strumenti rilevanti per i nuovi disordini informativi. L'analisi colloca il Regolamento nel più ampio panorama normativo dell'UE, che comprende il Digital Services Act, il Codice di buone pratiche sulla disinformazione e la Carta dei diritti fondamentali dell'Unione europea. L'attenzione è posta sul sistema di classificazione basato sul rischio e sulle norme di trasparenza previste dall'AI Act, che possono rafforzare la resilienza democratica se accompagnate dalle necessarie garanzie.

Keywords: AI Act; disinformation; generative AI; synthetic content; EU law; risk-based regulation; transparency obligations; freedom of expression; digital governance; platform accountability

Summary:

1. Introduction.

Our time has been described as a post-truth era,[1] in which public opinion seems to be guided more by emotions and personal beliefs than by objective facts.[2] For the purposes of this article, the term 'disinformation' is defined as the intentional spread of false or misleading information, which has become a serious threat to modern democracies. The problem seems to have been magnified by the rise and viral nature of digital platforms, as well as the emergence of large language models (LLMs), deepfakes, and other generative AI technologies. What used to require manual effort can now be automated and scaled up. Disinformation can also be tailored to individuals and situations with great speed and reach. For example, in the run-up to Slovakia's 2024 parliamentary elections, AI-generated audio recordings were circulated to impersonate a candidate, containing fake statements.[3] In the United States, during the 2024 primary season, there have been deepfake videos using President Biden's voice, urging US citizens not to vote. This illustrates how

[1] S Lewandowsky, U KH Ecker and J Cook, 'Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era' (2017) 6 Journal of Applied Research in Memory and Cognition 353.
[2] C Galletta Horner and others, 'Emotions: The Unexplored Fuel of Fake News on Social Media', *Fake News on the Internet* (Routledge 2023) 1039-1066.
[3] M Meaker, 'Slovakia's Election Deepfakes Show AI Is a Danger to Democracy' (*Wired*, 10 March 2023) <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/> accessed 11 July 2025.

generative AI can be misused to undermine public trust and electoral processes in modern democracies.[4] A similar pattern has been observed in Singapore, where the announcement of the 2025 general elections was followed by a spike in AI-generated political content online, as part of attempts to affect voters' ability to make informed decisions.[5]

A complex regulatory challenge emerges, as AI is integrated into the information ecosystem. Could legal frameworks keep pace with the tactics of those who exploit AI to distort public discourse, without compromising freedom of expression and opinion? Prompt-driven propaganda, lifelike synthetic avatars, and deceptively real fake videos are part of today's reality, they are no longer speculative. Not surprisingly, there are widespread concerns and calls for regulatory action. Against this backdrop, the article examines how the European Union is responding to the threat of AI-powered disinformation, with an emphasis on the Artificial Intelligence Act (AI Act). It assesses the role of the AI Act within the EU's wider digital governance agenda and considers how it might strengthen the regulatory response to algorithmically amplified manipulation. At the same time, the analysis highlights the Act's potential and its shortcomings in addressing this fast-moving threat landscape.

2. Disinformation meets AI: Definitions and Regulatory Tensions.

Disinformation involves the intentional creation and the spread of content that is demonstrably false or misleading. Disinformation is designed to deceive the public and potentially inflict harm.[6] However, its objective is not always to persuade but often to disrupt, to erode trust and social consensus, and to weaken democratic institutions. One example emerged during the COVID-19 pandemic, when coordinated efforts to falsely link the spread of the virus to 5G technology gained traction online.[7] These claims had no scientific basis, but they circulated widely on social media. They prompted vandalism of telecom infrastructure in several countries, and they contributed to public confusion at a time that was critical. Among the elements that set AI-driven disinformation apart from earlier forms, we can mention its scale, speed, and precision. Today's disinformation is powered by generative AI and deep learning models; it can be produced rapidly, while it can achieve high levels of realism and emotional appeal. The tools of 'deepfakes', i.e. hyper-realistic audio or video forgeries, have already been used to impersonate public figures, but also to interfere with elections, as well as to fabricate evidence of atrocities. These tools can mimic voices, facial expressions, and writing styles, leading the public to question the

[4] S Bond, 'How AI Deepfakes Polluted Elections in 2024' (NPR, 21 December 2024) <https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections> accessed 11 July 2025.

[5] T Davina, 'GE2025: Surge in AI-Generated Videos Related to Elections after Polls Called' (CNA, 23 April 2025) <https://www.channelnewsasia.com/singapore/ge2025-deepfakes-ai-generated-tiktok-politicians-disinformation-5078011> accessed 11 July 2025.

[6] D Fallis, 'What Is Disinformation?' (2015) 63 Library Trends 401; E Kapantai and others, 'A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework' (2021) 23 New Media & Society 1301.

[7] A Bruns, S Harrington and E Hurcombe, '"Corona? 5G? Or Both?": The Dynamics of COVID-19/5G Conspiracy Theories on Facebook' (2020) 177 Media International Australia 12.

authenticity of images, videos, and even first-hand accounts. In one instance, a 2022 video appeared online showing Ukrainian President Volodymyr Zelensky announcing his surrender to Russian forces. The video was unconvincing and easy to debunk, but it was shared across numerous social platforms and their users. This highlights how synthetic content, even at lower quality, can be weaponised to confuse the public and distort the understanding of political and military events.[8]

The spread of AI-generated content can be accelerated by algorithmic curation, which influences what users encounter online. The problem is that social media platforms depend on recommendation algorithms that prioritise engagement. This means that they often elevate content that provokes strong emotional reactions (outrage, fear, or humour), without regard for truthfulness. Therefore, emotionally charged disinformation is more likely to be surfaced and shared, while accurate but less sensational information may be overlooked by the social media platforms.[9] This creates an environment where falsehoods are likely to outperform factual content, not because they are more convincing, but because their provocative nature is favoured by platform algorithms.[10] Over time, such patterns can distort collective understanding, promote polarized approaches, and diminish trust in authoritative sources. A striking example occurred in the aftermath of the 2023 earthquake in Turkey and Syria. Social media feeds were filled with AI-generated images depicting destroyed buildings and suffering civilians.[11] Many of these were circulated by accounts aiming to attract attention, clicks, or even fraudulent donations. Despite their lack of authenticity, the emotional impact of these visuals led to widespread engagement, which often outpaced the reach of verified updates from emergency services. This underscores how generative AI, when paired with algorithmic amplification, can be exploited during crises to mislead the public.

At the regulatory level, legal systems are confronted with the difficult task of finding the right balance. Disinformation poses a threat to democratic dialogue, social cohesion, and public safety. It can fuel hatred, provoke violence, as well as manipulate public opinion through coordinated campaigns.[12] Yet, efforts to regulate disinformation carry risks too, such as the risk of overreach and the risk of suppressing lawful expression. This is especially problematic in areas where concepts such as "falsehood," "harm," or "public interest" are subjective and politically charged.[13] The legal challenge in this context lies in distinguishing between deliberate manipulation and honest mistake, and in designing responses that address the underlying intent of

[8] J Wakefield, 'Deepfake Presidents Used in Russia-Ukraine War' (*BBC*, 18 March 2022) <https://www.bbc.com/news/technology-60780142> accessed 11 July 2025.
[9] J Serrano-Puche, 'Digital Disinformation and Emotions: Exploring the Social Risks of Affective Polarization' (2021) 31 International Review of Sociology 231.
[10] N Bontridder, Y Poullet, 'The Role of Artificial Intelligence in Disinformation' (2021) 3 Data & Policy e32.
[11] S Khatsenkova, 'What's Real and What's Fake: Debunking Misleading Content about the Turkey Earthquake' (Euronews, 2 July 2023) <https://www.euronews.com/2023/02/07/whats-real-and-whats-fake-debunking-misleading-content-about-the-turkey-earthquake> accessed 10 July 2025.
[12] S McKay, C Tenove, 'Disinformation as a Threat to Deliberative Democracy' (2021) 74 Political Research Quarterly 703.
[13] L Gielow Jacobs, 'Freedom of Speech and Regulation of Fake News' (2022) 70 The American Journal of Comparative Law i278.

disinformation, rather than punishing those who merely share inaccurate or controversial views. A credible regulatory approach must focus on systemic harms, such as the algorithmic amplification of coordinated inauthentic behaviour or the orchestration of state-sponsored influence campaigns. At the same time, it must also preserve core rights, including freedom of expression, media pluralism, and political dissent.[14] France provides an interesting case study. In 2018, it enacted legislation aimed at curbing "manipulation of information" in the run-up to elections.[15] The provisions of that law empower courts to block or remove content deemed false and deliberately spread at scale within three months of a national vote. The measure was introduced to safeguard electoral integrity, but it has raised serious concerns and criticism from journalists, academics, and civil society groups, who argue that its vague terms risks enabling excessive state control over political speech.[16] The French example highlights the ongoing tension between countering manipulation and preserving the openness and contestation that are very important for democratic life.[17]

3. Regulatory Tools in the EU AI Act for Combating AI-Driven Disinformation.

The EU AI Act,[18] formally adopted in 2024, is the EU's flagship regulatory response to the opportunities and risks posed by AI systems.[19] The AI Act introduces a risk-based approach to AI governance and it classifies systems into four tiers: unacceptable risk (prohibited), high-risk, limited risk, and minimal risk.[20] The AI Act applies both to providers (those who develop AI systems) and deployers (those who use them), including public and private entities. Moreover, the scope of the AI Act is horizontal, cutting across sectors, and includes general-purpose AI models (GPAI), which is a class particularly relevant for disinformation concerns. Although the Act does not explicitly target disinformation, its architecture includes tools that can mitigate disinformation's supply chains and propagation mechanisms, particularly when powered by GPAI. Some sets of provisions are of particular relevance. First, there are provisions that address GPAI, including foundation models and

[14] F Nuñez, 'Disinformation Legislation and Freedom of Expression' (2020) 10 UC Irvine Law Review 783.
[15] I Couzigou, 'The French Legislation Against Digital Information Manipulation in Electoral Campaigns: A Scope Limited by Freedom of Expression' (2021) 20 Election Law Journal: Rules, Politics, and Policy 98.
[16] A Harcourt, 'Legal and Regulatory Responses to Misinformation and Populism' in T Howard and S Waisbord (eds), *The Routledge Companion to Media Disinformation and Populism* (Routledge 2021); MR Fiorentino, 'France Passes Controversial "fake News" Law' (*Euronews*, 22 November 2018) <https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law> accessed 11 July 2025.
[17] C Marsden, T Meyer and I Brown, 'Platform Values and Democratic Elections: How Can the Law Regulate Digital Disinformation?' (2020) 36 Computer Law & Security Review 105373.
[18] 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) (2024/1689).
[19] G Pavlidis, 'Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI' (2024) 16 Law, Innovation and Technology 293.
[20] M Ebers, 'Truly Risk-Based Regulation of Artificial Intelligence How to Implement the EU's AI Act' [2024] European Journal of Risk Regulation 1.

generative AI systems such as large language models.[21] Second, there are prohibitions for specific AI practices deemed incompatible with Union values, including manipulative techniques that exploit vulnerabilities.[22] Finally, there are post-market monitoring and enforcement mechanisms, which aim to ensure continued compliance and adapt to emerging risks.[23]

The AI Act introduces obligations for providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union. These obligations are imposed to providers irrespective of whether they are established or located within the Union or in a third country.[24] The term 'provider', as defined in Article 2(3) of the Act, refers to a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model. The Act also applies to deployers of AI systems that have their place of establishment or are located within the Union, as well as to providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the Union. The term 'deployer', as defined in Article 2(4) of the Act, means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity. Moreover, the Act covers importers and distributors of AI systems, product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark. Authorised representatives of providers, which are not established in the Union also covered.

Key requirements under the AI Act include disclosure that the content has been generated by AI, technical documentation on the training data, architecture, and performance, as well as the implementation of safeguards to prevent the generation of illegal content.[25] The Act creates distinct obligations for providers and deployers; nevertheless, the line between them can blur in the context of disinformation. For example, a malicious actor using a public API of an LLM to mass-generate election propaganda could be technically classified as a deployer. However, such person would exert de facto control over the outputs. Currently, most obligations fall on providers, while deployers remain lightly regulated, unless their use of the AI system can be qualified as high-risk. This seems to create an accountability gap, particularly for malicious or negligent uses of GPAI in campaigns of disinformation.

For its part, Article 50 (2) of the AI Act requires providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, to ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or

[21] OJ Gstrein, A Zwitter and N Haleem, 'General-Purpose AI Regulation and the European Union AI Act' (2024) 13 Internet Policy Review 1.
[22] RJ Neuwirth, 'Prohibited Artificial Intelligence Practices in the Proposed EU Artificial Intelligence Act (AIA)' (2023) 48 Computer Law & Security Review 105798.
[23] J Mokander and others, 'Conformity Assessments and Post-Market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation' (2022) 32 Minds and Machines 241.
[24] M Czerniawski, 'Towards the Effective Extraterritorial Enforcement of the AI Act' in Jaap-Henk Hoepman and others (eds), *Privacy Symposium* 2024 (Springer Nature Switzerland 2025).
[25] N Díaz-Rodríguez and others, 'Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation' (2023) 99 Information Fusion 101896.

manipulated. This is important given the proliferation of AI systems capable of producing vast quantities of synthetic content poses a challenge to the integrity of the information environment. As these systems advance, one can expect that the output would become increasingly indistinguishable from human-created content. This would exacerbate the risks of misinformation, large-scale manipulation, fraud, impersonation, and consumer deception. For this reason, it is important to require providers of AI systems, particularly those capable of generating or manipulating content, to incorporate technical solutions that allow the machine-readable identification of outputs that have been generated or modified by AI. These solutions may include watermarks, metadata tagging, cryptographic provenance methods, logging, or other appropriate techniques. [26] To be effective, such tools must be reliable, interoperable, and technically robust, while also adaptable to different content types, the costs of implementation and the generally acknowledged state of the art, as required by Article 50(2) of the AI Act. Importantly, this requirement must not be intended to extend to assistive AI tools that perform standard editing functions without substantially altering the semantic content or structure of the input.

In addition to the obligations imposed on AI providers, those who deploy AI systems must also respect the principle of transparency when they use AI tools to generate or alter visual or audio content that closely mimics real people or events in a way that could mislead audiences into believing that the content is genuine. [27] Article 50(4) of the AI Act requires deployers to clearly and prominently label such content to indicate its AI origin. The obligation, however, is not absolute. It must be balanced against the need to protect freedom of expression and artistic freedom, particularly in cases involving satire, fiction, or creative uses that are clearly not intended to deceive. [28] In those cases, the transparency requirement may be fulfilled through disclosure methods that are appropriate to the use and do not interfere with the audience's experience or the expressive purpose of the work. A comparable duty of transparency may also apply to AI-generated or manipulated text, especially when the content is presented as informing the public on matters of general interest. That said, this obligation may be waived where the material has undergone human editorial review and there is a clearly identifiable person or organisation that assumes responsibility for its publication. [29]

While these transparency obligations are binding, their effective implementation may be promoted by Union-level codes of practice. Under Article 50(7) of the AI Act, the AI Office must facilitate the drawing up of such codes of practice, which will then be approved by the Commission as implementing acts. These codes of practice may include practical arrangements

[26] AI Act (n 18) recital 133.
[27] J Botha and H Pieterse, 'Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security' (presented at 15th International Conference on Cyber Warfare and Security, 2020) <https://www.researchgate.net/publication/341454354_Fake_News_and_Deepfakes_A_Dangerous_Threat_for_21st_Century_Information_Security> accessed 11 July 2025.
[28] R Cejudo, 'Ethical Problems of the Use of Deepfakes in the Arts and Culture' in F Lara and J Deckers (eds), *Ethics of Artificial Intelligence* (Springer Nature Switzerland 2023) <https://doi.org/10.1007/978-3-031-48135-2_7> accessed 11 July 2025.
[29] G Hu, 'Challenges for Enforcing Editorial Policies on AI-Generated Papers' (2024) 31 Accountability in Research 978.

for content detection, as well as mechanisms for flagging or verifying the authenticity of information. They may also include collaboration mechanisms among stakeholders across the content distribution chain. Such initiatives have the potential to increase the capacity of users and intermediaries to identify AI-generated content and assess its trustworthiness.[30]

These transparency obligations gain particular significance in view of the Digital Services Act (DSA)[31] and the systemic risks that the DSA identifies, particularly those that threaten democratic institutions, electoral integrity, and the quality of civic discourse. Mechanisms for labelling AI-generated content form part of the broader regulatory strategy of the EU that requires very large online platforms and search engines to identify and mitigate risks, including those posed by synthetic media. It is important to note that the transparency duties set out in the AI Act operate independently, without prejudice to the content moderation rules established by the DSA. The obligation to disclose that content is artificially generated is separate from the legal assessment of whether the content is unlawful. That assessment is made solely on the basis of the applicable standards governing illegality, not on the mere fact that the content in question originated from an AI system.

The AI Act also introduces explicit prohibitions on certain uses of AI. These prohibitions cover the deployment of subliminal techniques and manipulative systems that exploit the vulnerabilities of specific groups, as outlined in Article 5(1) of the AI Act. In principle, this prohibition could cover disinformation campaigns targeting minors or marginalised communities. However, the thresholds for what constitutes "manipulation" or "exploitation" under the Act are relatively high. This means that political deepfakes or emotionally charged narratives aimed at a broad audience could fall outside the scope of this prohibition. One might argue that such content could fall within Article 5(1) of the AI Act where it causes significant harm or is used in a targeted and systematic manner. Yet, there is a broader issue of legal classification. Most instances of AI-generated disinformation do not meet the criteria for "unacceptable risk" under the AI Act, while they do not clearly fall within the high-risk categories. As it stands, the AI Act seems to be more finely tuned to addressing technical harms, such as those arising from the use of AI in medical devices or biometric surveillance, than to the harms associated with the manipulation of information, which are more dependent on the context.

The AI Act introduces post-market monitoring obligations for AI systems, including mechanisms for incident reporting and the continuous updating of risk assessments. In theory, these mechanisms could help identify cases of disinformation. National supervisory authorities and the newly established European AI Office are entrusted with the tasks of oversight and coordination. However, the Act does not establish a clear mechanism for redress in cases where AI-generated disinformation harms reputations or disrupts electoral processes. Without specific provisions targeting disinformation, enforcement

[30] RM Ballardini, K He and T Roos, 'Chapter 8: AI-Generated Content: Authorship and Inventorship in the Age of Artificial Intelligence' in T Pihlajarinne, J Vesala and O Honkkila (eds), *Online Distribution of Content in the EU* (Edward Elgar Publishing 2019).

[31] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277, 27.10.2022, 1–102.

is likely to remain reactive, relying on broader regulatory cooperation.

4. Legal Strategies Against AI-Fuelled Disinformation.

As previously noted, the DSA offers a complementary layer to the AI Act by addressing a different but still crucial dimension of the digital landscape: content dissemination. Whereas the AI Act regulates the design, development, and deployment of AI systems, particularly those classified as high-risk, the DSA focuses on how content is distributed and moderated across digital platforms. Under the DSA, Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs)—those with more than 45 million users in the EU—are subject to specific due diligence obligations.[32] These include mandatory risk assessments, algorithmic transparency, data access for vetted researchers, and independent audits, especially in relation to systemic risks. These risks include disinformation, algorithmic manipulation, and threats to democratic discourse. In the context of the EU regulatory approach, the DSA addresses the amplification and circulation of content in online spaces, while the AI Act governs how such content might be synthetically generated. When applied together, the two legislative instruments seek to address both the creation and the rapid spread of AI-driven disinformation. Nonetheless, there are still important regulatory gaps. Coordination between enforcement authorities (the European Commission, national Digital Services Coordinators,[33] and the newly created European AI Office) is not yet developed. Moreover, there are still questions around overlapping mandates, mechanisms for joint enforcement in cross-border cases, and coherent oversight of hybrid systems, which can merge generative AI technologies with large-scale infrastructures of content dissemination.

Some real-world scenarios can illustrate the intersection of the AI Act and the DSA. For example, a generative AI chatbot integrated into a VLOP could begin generating persuasive yet false claims about an emerging public health crisis. Since the platform's algorithms would amplify and disseminate this content to millions of users, the impact would become immediate and far-reaching. In such a case, the AI Act would apply to the AI provider, imposing technical safeguards such as output labelling, system traceability, and risk management procedures. At the same time, the DSA would require the platform to assess and mitigate the systemic risks posed by the spread of such disinformation, through transparency obligations, risk assessments, and possibly by modifying the content's algorithmic reach. In a second scenario, one could consider a political campaign using a generative AI tool to produce deepfake videos aimed at discrediting an opposing candidate. These videos could then spread through micro-targeted ads on a major social media

[32] B Genç-Gelgeç, 'Regulating Digital Platforms: Will the DSA Correct Its Predecessor's Deficiencies?' (2022) 18 Croatian Yearbook of European Law and Policy 25–60.

[33] P Van Cleynenbreugel and P Mattioli, 'Digital Services Coordinators and Other Competent Authorities in the Digital Services Act: Streamlined Enforcement Coordination Lost?' (European Law Blog, 30 November 2023) <https://www.europeanlawblog.eu/pub/digital-services-coordinators-and-other-competent-authorities-in-the-digital-services-act-streamlined-enforcement-coordination-lost/release/1> accessed 11 July 2025.

platform. In this case, the AI Act would impose obligations on the developer of the generative AI system, such as maintaining documentation, ensuring content authenticity, and preventing misuse. At the same time, the DSA would require the hosting platform to act swiftly, by removing or demoting harmful content and adjusting recommender systems, particularly in the lead-up to elections. These two examples highlight the need for coherent enforcement and coordination between the two regimes, particularly in scenarios where content creation and content dissemination are intertwined.

In addition to the DSA, the Code of Practice on Disinformation constitutes an important instrument aimed at combating online disinformation. Established in 2018, the Code was significantly strengthened in 2022, in order to be recognised as a Code of Conduct under the DSA. In February 2025, the Commission and the European Board for Digital Services endorsed the integration of the Code into the framework of the DSA.[34] The Code brings together a wide range of signatories, including online platforms and industry associations, as well as advertisers, fact-checkers, civil society organisations, and researchers. The Code contains numerous commitments and measures, covering various areas, from the demonetisation of disinformation and the transparency of political advertising to the empowerment of users and cooperation with fact-checkers. The Code is not legally binding, but it has become an integral part of the EU's regulatory framework, particularly under the DSA. For VLOPs and VLOSEs, adherence to the Code is recognised as a mitigation measure that demonstrates compliance with systemic risk management obligations under the DSA. For its part, the European Commission monitors the implementation of the Code through the Transparency Centre[35]. The newly established European Board for Digital Services[36] is also expected to play a coordinating role. The AI Act does not formally incorporate the Code of Practice, buts its transparency, traceability, and risk mitigation obligations, have the potential to support the Code's commitments. For example, requirements to label AI-generated content can support the Code's objectives on the traceability and authenticity of information. This would create meaningful synergies between the EU's voluntary and binding frameworks.

Article 11 of the EU Charter of Fundamental Rights[37] guarantees the right to freedom of expression and information, which sets the outer limits of both the AI Act and the DSA. Enforcement actions under these two instruments, particularly in the context of disinformation, content moderation, or transparency, must respect this fundamental right.[38] Importantly, freedom of

[34] European Commission, 'Commission Endorses the Integration of the Voluntary Code of Practice on Disinformation into the Digital Services Act' (2025) Press Release <https://digital-strategy.ec.europa.eu/en/news/commission-endorses-integration-voluntary-code-practice-disinformation-digital-services-act> accessed 11 July 2025.

[35] For more information, see 'Transparency Center' <https://disinfocode.eu/> accessed 11 July 2025.

[36] 'The European Board for Digital Services Is an Independent Advisory Group That Has Been Established by the Digital Services Act, with Effect from 17 February 2024. It Is Composed of the Member States' Digital Services Coordinators and Chaired by the European Commission. For more information see <https://digital-strategy.ec.europa.eu/en/policies/dsa-board> accessed 11 July 2025.

[37] Charter of Fundamental Rights of the European Union OJ C 326, 26.10.2012, p. 391–407.

[38] A Kuczerawy, 'Fighting Online Disinformation: Did the EU Code of Practice Forget about Freedom of Expression?' in D Trottier and others (eds), *Disinformation and Digital Media as a Challenge for Democracy* (Intersentia 2020).

expression does not protect only the communication of factual information. It also covers satire, parody, artistic creation, political speech, and the airing of dissenting or minority views, even when such expression is controversial, provocative, or even offensive to some audiences. However, AI-generated content risks blurring the boundaries between expression and deception. In this context, regulatory responses to disinformation must be designed to avoid chilling legitimate speech, such as the use of AI tools to create parody videos, satirical memes, or artistic reinterpretations of public figures. If rules are applied too broadly or without flexibility, there is a risk of suppressing lawful content and content with cultural value.[39] For this reason, there is a need for a rights-sensitive approach to the implementation of the rules, while measures to combat disinformation should be precise and proportionate. The focus must remain on intentional deception and coordinated manipulation, not on the mere fact that the content is AI-generated.

A real-world example, which has been widely discussed, concerns the viral image of Pope Francis wearing a white puffer jacket, generated by an AI tool and circulated in 2023.[40] The image was not created with the intent to deceive or cause harm, but its realism raised concerns about the importance of watermarking and provenance technologies. Some advocated for mandatory labelling of all AI-generated content, but critics cautioned that such blanket requirements could limit harmless or humorous uses of AI, particularly in the realm of popular culture. For example, an activist group could produce a satirical video with generative AI, exaggerating a political leader's statements during an election campaign. This might be labelled "manipulated media" by some, but it could also qualify as protected political expression, especially if its satirical nature is apparent to a reasonable viewer. In such cases, regulators and platforms must distinguish between manipulative disinformation and lawful forms of expression, [41] including exaggeration and parody. Moreover, procedural safeguards are essential for upholding freedom of expression. Human review of flagged content, accessible appeals processes, and transparency reporting can help guard against censorship by proxy, whether through government mandates or algorithmic content moderation.[42] The AI Act's obligations on transparency and record-keeping must also be interpreted in a manner consistent with fundamental rights. They should not place a disproportionate burden on artists, journalists, or civil society actors who use AI tools for legitimate purposes in a creative or critical context.

[39] Nuñez (n 14).
[40] D Tolentino, 'AI-Generated Images of Pope Francis in Puffer Jacket Fool the Internet' (NBC News, 27 March 2023) <https://www.nbcnews.com/tech/pope-francis-ai-generated-images-fool-internet-rcna76838> accessed 11 July 2025.
[41] I Khan, 'A/HRC/47/25: Disinformation and Freedom of Opinion and Expression - Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression' <https://www.ohchr.org/en/documents/thematic-reports/ahrc4725-disinformation-and-freedom-opinion-and-expression-report> accessed 11 July 2025.
[42] R Gorwa, R Binns and C Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945.

5. Critical Assessment and Future Directions.

The AI Act constitutes an important step forward in the EU's regulatory response to AI, but it is not a comprehensive solution to the problem of disinformation. Its provisions relevant to disinformation primarily address the supply side, i.e. the creation and dissemination of synthetic content. However, several key challenges remain. First, there is the problem of under-inclusiveness. There are many disinformation scenarios that fall outside the current classification framework of the AI Act. For instance, the malicious use of LLMs may not meet the thresholds for "high-risk" or "prohibited" systems under the AI Act, even when such use may cause real-world harm. Second, significant difficulties remain at the level of enforcement. Since disinformation is produced and distributed through opaque channels, it may be hard to trace specific outputs to a particular provider or deployer, particularly when open-source models or anonymised tools are involved.[43] Third, there is a problem of interpretative uncertainty. Key terms such as "manipulation," "transparency," and "exploitation" are open to legal and practical debate, and they are likely to remain contested. The precise meaning of these terms will need to be determined through regulatory guidance and the gradual development of case law.

Disinformation is not yet addressed through a unified legal framework at the EU level. The AI Act, the DSA, and the Charter of Fundamental Rights regulate different aspects of this fragmented landscape. There is a lack of dedicated regime targeting AI-driven misinformation risks, which risks opening the door to divergent national interpretations and enforcement across the EU Member States. Looking ahead, we argue that there is a need for a more integrated approach. This would involve aligning the enforcement of the AI Act with the systemic risk assessments under the DSA, as well as fostering technical interoperability through common standards for watermarking and metadata. This would also involve investing in complementary tools such as media literacy initiatives.[44] Such efforts should be supported by multi-stakeholder oversight structures that bring together regulators, civil society, and industry actors.

The AI Act is not crafted specifically to combat disinformation, but it introduces important tools, including transparency obligations, documentation requirements, and risk management processes. These tools can be used to address the supply side of synthetic and misleading content. To fight AI-driven disinformation, the EU must develop a rights-based, cross-cutting strategy that combines binding legal norms, soft governance frameworks, and technological safeguards. Moreover, effective implementation must be context-sensitive and transparent, seeking the delicate balance between safeguarding democratic integrity and upholding freedom of expression. As the fog of information continues to thicken, ensuring regulatory clarity is not merely a legal necessity: It is a democratic imperative.

---

[43] J Pielemeier, 'Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?' (2020) 2020 Utah Law Review 916.

[44] T Dame Adjin-Tettey, 'Combating Fake News, Disinformation, and Misinformation: Experimental Evidence for Media Literacy Education' (2022) 9 Cogent Arts & Humanities 2037229.