

UDK: 004.8 : 004.056
004.8 : 342.7
004.8 : 342.727
004.738.5 : 34 (094.5)
Preliminary communication

Hedvig Szabó, PostDoc researcher

Deák Ferenc Faculty of Law and Political Sciences, Széchenyi István University, Hungary

E-mail address: szabo.hedvig@sze.hu

ORCID-ID: 0009-0000-8403-8943

LEGAL REGULATION OF ARTIFICIAL INTELLIGENCE- BASED DISINFORMATION AND DEEPPAKE TECHNOLOGIES IN THE EUROPEAN UNION PROTECTION OF DEMOCRATIC ELECTORAL PROCESSES AND THE BALANCE WITH FREEDOM OF SPEECH*

ABSTRACT

Artificial intelligence-generated disinformation, including deepfake technologies, poses unprecedented challenges to democratic processes and electoral integrity in the European Union. This study examines the evolving legal framework regulating AI-based disinformation detection and content moderation, with particular attention to the tension between protecting democratic functioning and fundamental rights, namely freedom of expression. The analysis situates these regulatory developments within the broader context of the European approach to digital transformation, taking into account emerging risks associated with artificial intelligence technologies. The research examines three main dimensions of the regulatory environment. First, it analyzes substantive legal instruments addressing AI-generated disinformation, including the Digital Services Act (DSA), the Artificial Intelligence Act (AI Act), and the European Democracy Action Plan, evaluating their coherence and effectiveness in responding to technologically sophisticated manipulation techniques. Second, it examines the institutional architecture of implementation, particularly the role of national authorities, platform responsibilities, and the emerging framework of co-regulation between public and private actors. Third, it assesses the fundamental rights implications of automated content moderation systems, focusing on the risks of over-removal, algorithmic bias, and chilling effects on legitimate political discourse.

Keywords: *disinformation; artificial intelligence; Digital Services Act; deepfake; freedom of speech*

1. INTRODUCTION

The digital age has fundamentally transformed the nature of political communication and the dynamics of shaping democratic public opinion. Unlike the traditional media environment, where gatekeepers and editorial standards provided a certain filtering mechanism, online platforms have enabled the immediate and global dissemination of information with virtually no oversight. This democratization has brought many positive changes while also creating new vulnerabilities that disinformation campaigns can exploit.¹

The rapid development of artificial intelligence technologies, particularly in the area of generative AI and deepfake capabilities, has exponentially increased the potential impact of disinformation. Deepfake technology creates synthetic media that convincingly depicts real individuals saying or doing things they never actually said or did. University College London has identified deepfake technology as one of the most significant societal threats of our time, blurring the line between reality and fiction, raising serious concerns regarding privacy, ethical implications, and security threats.²

The European Union has developed the world's first comprehensive regulatory framework to address these challenges, which includes the Digital Services Act (DSA), the Artificial Intelligence Act (AI Act), the Regulation on the Transparency of Political Advertising, and the European Democracy Action Plan. Together, these regulatory instruments apply a risk-based approach that attempts to balance the protection of democratic institutions with respect for freedom of expression and the right to access information.³

This study goes beyond the existing literature by examining the EU's regulatory instruments on disinformation and artificial intelligence not merely in isolation, but within an explicitly electoral and fundamental rights-focused interpretive framework. While most research on the Digital Services Act (DSA) and the Artificial Intelligence Act (AI Act) primarily concentrates on technology-neutral risk management logic, market compliance issues, or general fundamental rights impacts, this study examines the specific challenges that AI-based disinformation

* This paper was supported by the STARTING_25 funding scheme of the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund.

¹ Vosoughi, S., Roy, D., Aral, S., *The spread of true and false news online*, Science, Vol. 359, No. 6380, 2018, p. 1146.

² Caldwell, M., Andrews, J. T. A., Tanay, T., Griffin, L. D., *AI-enabled future crime*, Crime Science, Vol. 9, No. 1, 2020, Article 14.

³ Senftleben, M., *Guardians of the UGC galaxy*, JIPITEC, Vol. 14, No. 3, 2023, p. 270.

- particularly deepfake technologies - pose to the integrity of democratic electoral processes, and within what constitutional boundaries the handling of these challenges can remain legitimate in the EU fundamental rights system.

The novelty of the research lies in its integrated connection of the imperative to protect electoral integrity, the structural risks arising from the operation of automated content moderation systems, and the EU model of co-regulation as the dominant paradigm of legislative intervention. The study argues that strengthening democratic resilience is not solely a matter of suppressing disinformation content, but is equally dependent on institutional and technological decisions that determine the proportionality, transparency, and remedial guarantees of fundamental rights restrictions. Through this approach, the study contributes to the discourse on interpreting the DSA and AI Act by highlighting that the handling of AI-based disinformation in the electoral context becomes one of the EU legal system's most sensitive fundamental rights test areas in the digital age.

2. THEORETICAL AND CONCEPTUAL FRAMEWORKS

2.1. Conceptual Definition of Disinformation

Precise definition of the concept of disinformation is fundamental to the legitimacy of regulatory interventions. The literature and EU documents generally distinguish disinformation (intentionally disseminated false or misleading information), misinformation (unintentional dissemination of false information), and malinformation (real information disseminated with harmful intent). The EU approach primarily focuses on disinformation, which it defines as verifiably false or misleading information created, disseminated, and published for economic gain or to deliberately deceive the public, and which is likely to cause public harm.⁴

It is important to emphasize that disinformation itself does not necessarily constitute illegal content. Recital 84 of the DSA preamble specifically points out that very large online platforms must also pay attention to information that is not illegal but contributes to systemic risks as defined in this regulation, including misleading or deceptive content, including disinformation. This distinction is of central importance to understanding the constitutional limits of regulation, as

⁴ European Commission, *Tackling online disinformation: a European approach*, COM(2018) 236 final, 2018.

restricting freedom of expression requires greater justification in the case of non-illegal content.⁵

2.2. Deepfake Technology and Its Legal Challenges

According to Article 3(60) of the AI Act, deepfake means AI-generated or manipulated image, audio, or video content that depicts existing persons, objects, places, entities, or events and appears deceptively authentic or real to a person. This definition goes beyond the depiction of persons and includes objects, organizations, and events, recognizing that deepfake technology can be used not only for deceptive depictions of individuals but also of institutions.⁶

The availability and sophistication of deepfake technology has increased dramatically in recent years. In 2023, a 3000% increase in deepfake fraud attempts was recorded, while during the 2024 election year, the use of deepfakes for political purposes was documented in numerous countries. In Slovakia, AI-generated audio recordings spread just days before the 2023 elections, falsely depicting a liberal candidate discussing election fraud. In India's 2024 general elections, deepfakes depicting celebrities criticizing the prime minister went viral on WhatsApp and YouTube.⁷⁸

2.3. Theoretical Foundations of Fundamental Rights Balance

The tension between combating disinformation and protecting freedom of expression is one of the most complex issues in fundamental rights doctrine. According to the consistent practice of the European Court of Human Rights (ECtHR), freedom of expression is one of the essential foundations of a democratic society, which applies to all information and ideas, including those that offend, shock, or disturb. However, this broad protection is not absolute: Article 10(2) of the European Convention on Human Rights allows restrictions provided

⁵ Husovec, M., *The Digital Services Act's red line: What the Commission can and cannot do about disinformation*, Journal of Media Law, Vol. 16, No. 1, 2024, p. 48.

⁶ Pawelec, M., *Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions*, Digital Society, Vol. 1, No. 2, 2022, Article 19.

⁷ Roy, E., Laskar, K. A., Reyaz, M., *Retroactive gatekeepers: Fact-checkers, electoral disinformation, and audience engagement in India*, Emerging Media, Vol. 3, No. 4, 2025, p. 585.

⁸ Macaulay, T., *Deepfake fraud attempts are up 3000% in 2023 - here's why*, The Next Web, 15 November 2023.

they are prescribed by law, serve legitimate aims, and are necessary in a democratic society.⁹

Article 11 of the EU Charter of Fundamental Rights provides similar protection, but in the EU context, the practice of the Court of Justice of the European Union (CJEU) is also of decisive importance. In Case C-401/19 Poland v. Parliament and Council (2022), the CJEU specifically addressed the relationship between online content moderation and freedom of expression, finding that preventive filtering obligations restrict an important means of disseminating online content and thus constitute a restriction of Article 11 of the Charter. At the same time, the Court upheld the justifiability of such restrictions with appropriate safeguards.

3. EU REGULATORY FRAMEWORK

3.1. Digital Services Act (DSA)

Regulation (EU) 2022/2065, the Digital Services Act (DSA), became fully applicable in the EU in February 2024 and is the world's first comprehensive online safety legislation applying a systematic approach to a wide range of online services and risks. The regulation aims to create a safe, predictable, and reliable online environment that promotes innovation while protecting fundamental rights enshrined in the EU Charter of Fundamental Rights, such as freedom of expression and information or non-discrimination.¹⁰

The DSA applies a multi-layered regulatory approach: it sets basic obligations for all services, enhanced obligations for online platforms, and the strictest requirements for very large online platforms (VLOPs) and very large online search engines (VLOSEs). VLOPs are those platforms whose average monthly active users in the Union reach or exceed 45 million. In April 2023, the European Commission designated the first 17 VLOPs, including Facebook, Instagram, LinkedIn, Snapchat, TikTok, Twitter (now X), and YouTube platforms, as well as 2 VLOSEs.¹¹

⁹ Gerards, J., *How to improve the necessity test of the European Court of Human Rights*, International Journal of Constitutional Law, Vol. 11, No. 2, 2013, p. 467.

¹⁰ Ó Fathaigh, R., Buijs, D., van Hoboken, J., *The regulation of disinformation under the Digital Services Act*, Media and Communication, Vol. 13, 2025.

¹¹ European Commission, *Digital Services Act: Commission designates first set of Very Large Online Platforms and Search Engines*, IP/23/2413, 25 April 2023.

Article 34 of the DSA requires VLOPs to conduct annual risk assessments regarding systemic risks arising from the design, functioning, and use of their services. Under Article 35, platforms must take appropriate, proportionate, and effective risk mitigation measures. Disinformation is specifically included among the systemic risks that service providers must address, although the regulation emphasizes that risk mitigation measures do not necessarily mean content removal - demonetization, warning labels, or fact-checking references may also be applied.¹²

In February 2025, the European Commission and the European Board for Digital Services approved the integration of the Code of Practice on Disinformation into the DSA framework. From July 1, 2025, the Code will enjoy the status of a code of conduct under the DSA, meaning that the commitments contained therein become auditable and serve as a relevant reference point in assessing DSA compliance. This step makes the previously voluntary self-regulatory tool part of the legal framework, potentially with enforcement consequences for non-compliant platforms.¹³

3.2. Artificial Intelligence Act (AI Act)

Regulation (EU) 2024/1689, the AI Act, entered into force in August 2024 and is the world's first comprehensive AI regulation. The regulation applies a risk-based approach: it prohibits manipulative and exploitative AI systems, sets strict requirements for high-risk systems, and prescribes transparency obligations for certain AI applications, including systems generating deepfakes.¹⁴

Article 50 of the AI Act establishes specific transparency obligations for deepfakes. Providers of AI systems generating synthetic audio, image, video, or text content must ensure that the outputs are marked in machine-readable format and are recognizable as artificially generated or manipulated. Additionally, users employing deepfakes are obliged to publicly disclose that the content is artificially generated or manipulated. This transparency obligation does not apply to use authorized for the detection, prevention, or investigation of criminal offenses.¹⁵

¹² Husovec, M., *The Digital Services Act's red line*, *op. cit.*, note 6, p. 49.

¹³ Mündges, S., Park, K., *But did they really? Platforms' compliance with the Code of Practice on Disinformation in review*, 2024.

¹⁴ Laux, J., Wachter, S., Mittelstadt, B., *Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk*, *Regulation & Governance*, Vol. 18, No. 1, 2024, p. 5.

¹⁵ Łabuz, M., *op. cit.*, note 3.

It is important to note that the AI Act does not completely prohibit deepfakes. Instead, it imposes strict transparency requirements on both providers and users. The regulation's full applicability is due from August 2, 2026, although certain provisions - such as obligations regarding prohibited AI practices and AI literacy - entered into force earlier. In December 2025, the Commission published the first draft of a code of practice on marking and labeling AI-generated content, providing practical guidance on watermarking, metadata, and other technical measures.¹⁶

3.3. Regulation on Transparency of Political Advertising

Regulation (EU) 2024/900 on the transparency and targeting of political advertising (TTPA) entered into force on April 9, 2024, with most provisions applicable from October 10, 2025. The regulation aims to contribute to the proper functioning of the internal market for political advertising and to open, fair political debate, addressing concerns about information manipulation and foreign interference in elections.¹⁷

The regulation requires that political advertisements be clearly marked as political advertisements and provide information about who financed them, which election or referendum they relate to, and whether targeting or ad delivery techniques were used. Targeted political advertising in an online environment is only permitted under strict conditions: data must be collected from the data subject and may only be used if the data subject has given explicit and separate consent for their use for political advertising purposes. Special categories of data - such as data relating to racial or ethnic origin or political opinions - cannot be used for profiling.¹⁸

To prevent foreign interference, it is prohibited to provide political advertising services to sponsors from third countries during the last three months before elections or referendums. In the event of violations, sanctions established by Member States may reach up to 6% of the service provider's annual worldwide turnover. The regulation also establishes a public European repository of online

¹⁶ Meding, L., Burchardi, A., Veale, M., *What constitutes a deep fake? The blurry line between legitimate processing and manipulation under the EU AI Act*, Proceedings of the ACM Symposium on Computer Science and Law, ACM, 2025.

¹⁷ Nikiforov, L., *Transparency in targeting of political advertising: Challenges remain*, SSRN Working Paper, 1 November 2024.

¹⁸ Stephenson, R., Rinceanu, J., Bovermann, M. A., *Regulating political advertising in the EU: Transparency without accountability*, Euclid, 2025(1), p. 91.

political advertising, where all online political advertisements will be publicly accessible.

3.4. European Democracy Action Plan

The European Commission presented the European Democracy Action Plan (EDAP) in December 2020, which announced further steps, including legislative measures, as a non-legislative initiative. The Action Plan has three main pillars: protecting the integrity of elections and promoting democratic participation, strengthening media freedom and media pluralism, and combating disinformation, foreign interference, and information influence operations.¹⁹

Within the framework of the Action Plan, the EU has intensified its response to disinformation threats both domestically and internationally. This includes establishing an operational mechanism through the European Cooperation Network on Elections, strengthening the Code of Practice on Disinformation, and supporting civil society and media freedom. The plan also sets out measures to protect journalists from abusive lawsuits (SLAPP) and to strengthen transparency in media ownership.²⁰

4. INSTITUTIONAL ARCHITECTURE OF IMPLEMENTATION

4.1. Multi-level Implementation Mechanisms

The DSA creates a multi-level enforcement system in which national-level Digital Services Coordinators (DSCs) and the European Commission have shared competences. Member States were obliged by the February 2024 deadline to designate an independent regulatory authority as a DSC, which is responsible for enforcing the rules on intermediary services, hosting services, and non-VLOP platforms in their territory. The designated DSCs have different backgrounds - some are new bodies, others are existing communication, media regulatory, consumer protection, or competition authorities.²¹

¹⁹ European Commission, *European Democracy Action Plan*, COM(2020) 790 final, 2020.

²⁰ Haapala, I., Roch, C., *EU democratic resilience in times of information warfare: A study of the European Democracy Action Plan*, European Foreign Affairs Review, Vol. 30, No. 1, 2025, p. 85.

²¹ Rangoni, B., Thatcher, M., *National de-delegation in multi-level settings: Independent regulatory agencies in Europe*, Governance, Vol. 36, No. 1, 2023, p. 85.

Regarding the supervision of VLOPs and VLOSEs, the European Commission acts as the exclusive enforcement authority. In 2023, the Commission established the European Centre for Algorithmic Transparency (ECAT), which supports enforcement efforts. The Commission has initiated several proceedings against major platforms, including against Meta, TikTok, and X (formerly Twitter), due to concerns that they are not complying with the DSA's requirements on electoral integrity. In March 2024, the Commission issued guidelines for VLOPs and VLOSEs on recommended measures to mitigate systemic risks affecting electoral processes.²²²³²⁴

4.2. The Co-regulation Model

The EU's approach to addressing disinformation is based on a co-regulation model that combines mandatory legal requirements and industry self-regulatory initiatives. Article 45 of the DSA allows for the development of codes of conduct serving specific public interest objectives. The Code of Practice on Disinformation, strengthened in 2022, was specifically designed to be recognized as a code of conduct under the DSA, which was realized in 2025.²⁵

Signatories to the Code - including online platforms, search engines, the advertising industry, fact-checking organizations, and civil society organizations - commit to implementing various measures, such as diverting advertising revenue from disinformation content, ensuring transparency in political advertising, providing users with tools and features to recognize disinformation, and ensuring researchers' access to data. This model provides flexibility for platforms in choosing how to comply while establishing common standards and control mechanisms. However, due to the Code's voluntary nature, platforms do not face meaningful sanctions for non-compliance with commitments, and enforcement is often fragmented and performative.²⁶

²² Sundorph, E., *European Centre for Algorithmic Transparency (ECAT): Issue 3*, Publications Office of the European Union, 2024.

²³ Miller, G., *EU tests DSA with investigative proceedings against Musk's X*, TechPolicy Press, 19 December 2023.

²⁴ Broughton Micova, S., Schnurr, D., *Systemic risk in digital services: Benchmarks for evaluating the management of risks to electoral processes*, CERRE, 2024.

²⁵ Oleart, A., Bouza García, L., *From self to co-regulation in the EU's approach to disinformation*, International Review of Public Policy, Vol. 7, No. 2, 2025, p. 240.

²⁶ Shattock, E., *Self-regulation 2.0? A critical reflection of the European fight against disinformation*, Harvard Kennedy School Misinformation Review, 2021.

5. FUNDAMENTAL RIGHTS IMPLICATIONS AND RISKS

5.1. Risks of Algorithmic Content Moderation

A natural consequence of the systemic risk assessment and mitigation required by the DSA is that platforms increasingly rely on algorithmic systems for large-scale content moderation. However, this development carries significant fundamental rights risks. According to a 2024 study by Vanderbilt University's Future of Free Speech research institute, platforms may over-moderate content to avoid DSA fines, which can lead to unjustified removal of lawful content.²⁷

One of the most serious problems with algorithmic content moderation is inconsistency and arbitrariness. Research has shown that different machine learning models can produce different classification results for the same content, leading to a phenomenon called predictive multiplicity. This means that the restriction of freedom of expression - which theoretically requires justification - may depend on the technical decision of model selection rather than objective legal criteria. The opacity of such arbitrary decisions poses a threat to fundamental freedoms, as restrictions must be justifiable.²⁸

The reinforcement of bias is another concern. Research has documented that algorithmic content moderation disproportionately affects certain social groups. Messages containing minority identity markers (e.g., Arab, Black, LGBTQ+), explicit profanity, or specific racial and ethnic dialects (e.g., African American English) are more likely to be classified as toxic, regardless of their communication context. The EU Fundamental Rights Agency expressed concern in 2023 that large amounts of misogyny and racism bypass content moderation systems, while algorithms are prone to errors and may ultimately popularize online hatred over time.²⁹

²⁷ Alkiviadou, N., *Platform liability, hate speech and the fundamental right to free speech*, Information & Communications Technology Law, Vol. 34, No. 2, 2024, p. 3.

²⁸ Gomez, J. F., Machado, C. V., Paes, L. M., Calmon, F. P., *Algorithmic arbitrariness in content moderation*, Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, 2024, p. 2235.

²⁹ EU Fundamental Rights Agency (FRA), *Online content moderation: Current challenges in detecting hate speech*, <https://fra.europa.eu/en/publication/2023/online-content-moderation>, 29 November 2023.

5.2. Over-removal and Chilling Effect

The phenomenon of over-removal, when lawful content is mistakenly removed, is particularly concerning in the context of political speech. The ECtHR has consistently emphasized that political expression enjoys enhanced protection and should be evaluated from the perspective of its contribution to democratic debate. However, automated filters are often unable to understand context and nuance, which can lead to mistaken removal of legitimate political criticism.³⁰

The concept of chilling effect has long been recognized in freedom of expression jurisprudence. This phenomenon occurs when individuals introduce voluntary restrictions on their expressions out of fear of illegal proceedings or sanctions, even if their expression would actually be protected. In the context of online content moderation, the chilling effect occurs when users avoid discussing certain topics or expressing opinions because they fear their content will be removed or their accounts suspended. Multi-stage complaint and dispute resolution processes can frustrate users and lead them to abandon well-founded claims for restoration of their uploads.³¹

5.3. Relevant Case Law of the CJEU and ECtHR

The CJEU's judgment in Case C-18/18 *Glawischnig-Piesczek v. Facebook Ireland* (2019) is one of the landmark decisions on the relationship between content moderation and freedom of expression. The Court ruled that when a court orders the removal of an unlawful post, it may also order the removal of identical content and content that is equivalent in content. This judgment raised concerns about the use of automated filters, which can lead to context-independent removals.³²

In Case C-401/19 *Poland v. Parliament and Council* (2022), the CJEU's Grand Chamber examined the constitutionality of Article 17 of the Copyright Directive from the perspective of freedom of expression. The Grand Chamber found that preventive filtering obligations restrict an important means of disseminating online content, thereby constituting a restriction of Article 11 of the Charter.

³⁰ Gorwa, R., Binns, R., Katzenbach, C., *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*, Big Data & Society, Vol. 7, No. 1, 2020, p. 5.

³¹ Are, C., *'Dysfunctional' appeals and failures of algorithmic justice in Instagram and TikTok content moderation*, Information, Communication & Society, Vol. 28, No. 11, 2025, p. 2000.

³² Keller, D., *Facebook filters, fundamental rights, and the CJEU's Glawischnig-Piesczek ruling*, GRUR International, Vol. 69, No. 6, 2020, p. 617.

However, the Court upheld the justifiability of the restriction, emphasizing that Member States must provide appropriate safeguards for the protection of freedom of expression, including ensuring that platforms only block manifestly unlawful content.³³

The ECtHR's judgment in *Sanchez v. France* (2023) raised new questions regarding social media content moderation. The Grand Chamber found that a politician's criminal liability for hate speech comments written by others on their Facebook page did not violate Article 10 of the Convention because the politician did not exercise appropriate vigilance regarding the comments. This judgment raised concerns that it holds public figures responsible for others' content, which may encourage them to preventively restrict commenting options.³⁴

6. SPECIFICITIES OF THE ELECTORAL CONTEXT

6.1. The Imperative of Protecting Electoral Integrity

The protection of electoral integrity is a legitimate public interest objective that can justify certain restrictions on freedom of expression. The United States Supreme Court in *Burson v. Freeman* (1992) recognized that protecting the integrity of elections is a compelling state interest that can justify restrictions - such as prohibiting campaign activities near polling places. However, in other cases, such as *McIntyre v. Ohio Elections Commission* (1995), the Court found that certain electoral rules go too far and violate the First Amendment.

In the EU context, the DSA's electoral guidelines issued in March 2024 provide recommendations for VLOPs and VLOSEs to mitigate systemic risks affecting electoral processes before, during, and after elections. While not legally binding, the guidelines clarify the DSA text and provide practical, current examples of how services can meet their risk mitigation obligations under the DSA. They pay particular attention to measures on online political advertising and generative AI.³⁵

6.2. The Impact of Deepfakes on Elections: Empirical Experience

³³ Senftleben, M., *Guardians of the UGC galaxy*, JIPITEC, Vol. 14, No. 3, 2023, p. 270.

³⁴ Husovec, M., Grote, T., et al., *Grand confusion after Sanchez v. France*, Common Market Law Review, 2024.

³⁵ Broughton Micova, S., Schnurr, D., *op. cit.*, note 25.

During the 2024 super election year, when more than half of the world's population participated in elections, numerous cases documented the use of deepfakes for electoral purposes. The most widely cited case occurred before the January 2024 New Hampshire primary, when AI-generated robocalls - imitating President Joe Biden's voice - called on voters not to vote. The Federal Communications Commission subsequently declared robocalls containing AI-generated voices illegal.³⁶

However, it is important to nuance the assessment of the deepfake threat. The Alan Turing Institute's study found that AI-based disinformation did not actually have a significant impact on 2024 election results in the United Kingdom and EU. While there were viral deepfakes, users amplifying the content generally came from a small number of people identifying with ideological narratives embedded in that content. In other words, the most controversial voices on platforms generally do not influence undecided voters.³⁷

The Brennan Center for Justice analysis pointed out that the long-term impact of deepfakes goes beyond immediate disinformation. The liar's dividend phenomenon means that the existence of deepfakes allows anyone - politicians, corporations, or other influential figures - to avoid accountability by questioning the authenticity of credible evidence. Over time, this uncertainty weakens democratic institutions, promotes alienation, and makes societies more vulnerable to manipulation.³⁸

7. CHALLENGES AND RECOMMENDATIONS

7.1. Challenges of Regulatory Coherence

Despite the EU's comprehensive regulatory architecture, significant challenges remain in ensuring coherent and effective implementation. Addressing potential overlaps and contradictions between the AI Act and DSA requires ongoing attention. For example, both regulations prescribe transparency obligations for AI-

³⁶ Federal Communications Commission, *Declaratory Ruling on AI-generated robocalls*, FCC 24-17, 8 February 2024.

³⁷ The Alan Turing Institute, *Turing event at the Houses of Parliament explores impact of AI disinformation on elections*, <https://www.turing.ac.uk/news/>, 30 January 2025.

³⁸ Schiff, K. J., Schiff, D. S., Bueno, N. S., *The liar's dividend: Can politicians claim misinformation to evade accountability?*, American Political Science Review, Vol. 118, No. 2, 2024, p. 650.

generated content but apply different terminology and approaches. Member States have different regulatory frameworks and implementation capacities, which can lead to uneven application.

The AI Act does not specify which regulatory authority should act as a supervisory body, so Member States must establish their own national sanction systems, which can lead to different enforcement practices. Some Member States - such as Spain - have created dedicated regulatory bodies to supervise AI developments, while others - such as France and Germany - do not yet have specific legislation or bodies for regulating artificial intelligence. This fragmentation may weaken the effectiveness of the EU-level response.³⁹

7.2. Technological Challenges

Despite the rapid development of deepfake detection technologies, significant technical limitations remain. Generative AI models are becoming increasingly sophisticated, while detection systems always struggle with a step behind. Current detection systems are particularly weak in identifying voice-based deepfakes, which are increasingly popular in electoral disinformation campaigns, as evidenced by the Slovak and United States cases.⁴⁰

Implementation of the watermarking and metadata requirements prescribed by the AI Act also poses challenges. Malicious actors may intentionally remove metadata and detection markings. Additionally, enforcement on global platforms - where content originates from and is directed to different jurisdictions - raises practical difficulties. Organizations cannot rely solely on upstream compliance; they need independent verification capabilities to identify synthetic content regardless of source or labeling status.⁴¹

7.3. Recommendations for Legislators and Platforms

Based on the research, the following recommendations can be formulated for effective and fundamental rights-compliant disinformation regulation. First,

³⁹ Future of Life Institute, *Overview of all AI Act national implementation plans*, <https://artificialintelligenceact.eu/national-implementation-plans>, 19 May 2025.

⁴⁰ Sharma, K., *AI voice cloning and deepfakes are here, and they are dangerous*, Scientific American, 15 January 2026.

⁴¹ Feltes, N., *Article 50 AI Act: Do the transparency provisions improve upon the Commission's draft?*, JIPITEC, Vol. 16, 2025, p. 230.

strengthening transparency obligations for algorithmic content moderation systems is necessary. Platforms should publicly make available descriptions of the main parameters of algorithmic decision-making, error rates, and breakdowns by affected user groups. This would enable independent researchers and regulators to identify patterns of bias and over-removal.

Second, strengthening human oversight requirements is essential, particularly regarding political content and electoral periods. The DSA requires ensuring human oversight in content moderation decisions, but further attention is needed to specify and monitor compliance with these requirements. In the case of political speech - where understanding context is critical - purely automated decision-making can be particularly problematic.

Third, ensuring effective remedial mechanisms for cases of erroneous removal is indispensable. Article 20 of the DSA requires platforms to establish internal complaint handling mechanisms, and Article 21 allows for out-of-court dispute resolution. However, evaluating the effectiveness of these mechanisms and strengthening them as needed is an ongoing task. Users must be provided with clear, prompt, and cost-free procedures to request restoration of their content.

Fourth, developing media literacy and critical thinking throughout society is of paramount importance. Technical and legal solutions alone are not sufficient; citizens need to be able to recognize disinformation and manipulated content. The European Democracy Action Plan rightly emphasizes supporting media literacy programs, and intensifying these efforts is warranted.

8. CONCLUSION

This study has provided a comprehensive analysis of the legal regulation of AI-based disinformation and deepfake technologies in the European Union, with particular attention to the protection of democratic electoral processes and respect for freedom of expression. The research findings indicate that the EU has developed a comprehensive regulatory architecture - through the DSA, AI Act, Regulation on the Transparency of Political Advertising, and the European Democracy Action Plan - representing the world's most advanced framework for addressing these challenges.

However, significant challenges remain in ensuring proportionate and fundamental rights-respecting implementation. Reliance on AI-based detection

systems introduces new risks of error and opacity that may undermine the very democratic values that regulation seeks to protect. The inconsistency and bias of algorithmic content moderation, the phenomenon of over-removal, and the chilling effect on political speech are all problems requiring ongoing attention and fine-tuning.

The study argues that effective disinformation governance requires not only robust legal frameworks but also transparency obligations for detection algorithms, meaningful human oversight, and effective remedial mechanisms for erroneous content removals. Finding the balance between protecting democratic institutions and respecting freedom of expression remains an ongoing challenge requiring constant dialogue and cooperation among regulators, platforms, civil society, and researchers.

The pace of technological development exceeds the adaptive capacity of regulation, necessitating a flexible, risk-based approach. The EU's co-regulation model - which combines mandatory legal requirements and industry self-regulatory initiatives - offers a promising approach, but its success depends on the quality of implementation and continuous review.

Ultimately, democratic resilience depends not only on legal and technical tools but also on society's overall media literacy, critical thinking, and commitment to democratic values. Legal regulation is an important tool, but it is not sufficient alone to protect the democratic public sphere in the digital age.

REFERENCES

BOOKS AND ARTICLES

1. Alkiviadou, N., *Platform liability, hate speech and the fundamental right to free speech*, Information & Communications Technology Law, Vol. 34, No. 2, 2024, pp. 1-11.
2. Are, C., *'Dysfunctional' appeals and failures of algorithmic justice in Instagram and TikTok content moderation*, Information, Communication & Society, Vol. 28, No. 11, 2025, pp. 1997-2014.
3. Broughton Micova, S., Schnurr, D., *Systemic risk in digital services: Benchmarks for evaluating the management of risks to electoral processes*, Centre for Regulation in Europe (CERRE), 2024.

4. Caldwell, M., Andrews, J. T. A., Tanay, T., Griffin, L. D., *AI-enabled future crime*, Crime Science, Vol. 9, No. 1, 2020, Article 14.
5. Feltes, N., *Article 50 AI Act: Do the transparency provisions improve upon the Commission's draft?*, JIPITEC - Journal of Intellectual Property, Information Technology and E-Commerce Law, Vol. 16, 2025, pp. 222-240.
6. Gerards, J., *How to improve the necessity test of the European Court of Human Rights*, International Journal of Constitutional Law, Vol. 11, No. 2, 2013, pp. 466-490.
7. Gomez, J. F., Machado, C. V., Paes, L. M., Calmon, F. P., *Algorithmic arbitrariness in content moderation*, Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), ACM, 2024, pp. 2234-2253.
8. Gorwa, R., Binns, R., Katzenbach, C., *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*, Big Data & Society, Vol. 7, No. 1, 2020, pp. 1-15.
9. Haapala, I., Roch, C., *EU democratic resilience in times of information warfare: A study of the European Democracy Action Plan*, European Foreign Affairs Review, Vol. 30, No. 1, 2025, pp. 83-106.
10. Husovec, M., *The Digital Services Act's red line: What the Commission can and cannot do about disinformation*, Journal of Media Law, Vol. 16, No. 1, 2024, pp. 47-56.
11. Husovec, M., Grote, T., Mazhar, Y., Mikhaeil, C., Miñarro Escalona, H., Sinha Kumar, P., Sreenath, S., *Grand confusion after Sanchez v. France: Seven reasons for concern about Strasbourg jurisprudence on intermediaries*, Common Market Law Review, 2024.
12. Keller, D., *Facebook filters, fundamental rights, and the CJEU's Glawischnig-Piesczek ruling*, GRUR International, Vol. 69, No. 6, 2020, pp. 616-623.
13. Łabuz, M., *A teleological interpretation of the definition of deepfakes in the EU Artificial Intelligence Act*, Policy & Internet, 2025.
14. Laux, J., Wachter, S., Mittelstadt, B., *Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk*, Regulation & Governance, Vol. 18, No. 1, 2024, pp. 3-32.
15. Macaulay, T., *Deepfake fraud attempts are up 3000% in 2023 - here's why*, The Next Web, 15 November 2023.
16. Meding, L., Burchardi, A., Veale, M., *What constitutes a deep fake? The blurry line between legitimate processing and manipulation under the EU AI Act*, Proceedings of the ACM Symposium on Computer Science and Law, ACM, 2025.

17. Miller, G., *EU tests DSA with investigative proceedings against Musk's X*, TechPolicy Press, 19 December 2023.
18. Mündges, S., Park, K., *But did they really? Platforms' compliance with the Code of Practice on Disinformation in review*, , 2024.
19. Nikiforov, L., *Transparency in targeting of political advertising: Challenges remain*, SSRN Working Paper, 1 November 2024.
20. Ó Fathaigh, R., Buijs, D., van Hoboken, J., *The regulation of disinformation under the Digital Services Act*, Media and Communication, Vol. 13, 2025.
21. Oleart, A., Bouza García, L., *From self to co-regulation in the EU's approach to disinformation: The framing power of Big Tech business lobbies in the lead to the Digital Services Act*, International Review of Public Policy, Vol. 7, No. 2, 2025, pp. 235-256.
22. Pawelec, M., *Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions*, Digital Society, Vol. 1, No. 2, 2022, Article 19.
23. Rangoni, B., Thatcher, M., *National de-delegation in multi-level settings: Independent regulatory agencies in Europe*, Governance, Vol. 36, No. 1, 2023, pp. 81-103.
24. Roy, E., Laskar, K. A., Reyaz, M., *Retroactive gatekeepers: Fact-checkers, electoral disinformation, and audience engagement in India*, Emerging Media, Vol. 3, No. 4, 2025, pp. 582-600.
25. Schiff, K. J., Schiff, D. S., Bueno, N. S., *The liar's dividend: Can politicians claim misinformation to evade accountability?*, American Political Science Review, Vol. 118, No. 2, 2024, pp. 648-666.
26. Senftleben, M., *Guardians of the UGC galaxy: Human rights obligations of online platforms, copyright holders, member states and the European Commission under the CDSM Directive and the Digital Services Act*, JIPITEC - Journal of Intellectual Property, Information Technology and E-Commerce Law, Vol. 14, No. 3, 2023, pp. 264-289.
27. Shattock, E., *Self-regulation 2.0? A critical reflection of the European fight against disinformation*, Harvard Kennedy School Misinformation Review, 2021.
28. Sharma, K., *AI voice cloning and deepfakes are here, and they are dangerous*, Scientific American, 15 January 2026.
29. Stephenson, R., Rinceanu, J., Bovermann, M. A., *Regulating political advertising in the EU: Transparency without accountability*, Eucrium - European Law Forum: Prevention • Investigation • Prosecution, 2025(1), pp. 90-97.
30. Sundorph, E., *European Centre for Algorithmic Transparency (ECAT): Issue 3*, Publications Office of the European Union (E. A. Pena Fernandez, Ed.; JRC139131), 2024.

31. van Drunen, M., Helberger, N., Ó Fathaigh, R., *The beginning of EU political advertising law: Unifying democratic visions through the internal market*, International Journal of Law and Information Technology, Vol. 30, No. 4, 2022, pp. 424-445.
32. Vosoughi, S., Roy, D., Aral, S., *The spread of true and false news online*, Science, Vol. 359, No. 6380, 2018, pp. 1146-1151.

LEGAL RULES

1. 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), Official Journal of the European Union, L 277, 1-102.
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Official Journal of the European Union, L 2024/1689.
3. Regulation (EU) 2024/900 of the European Parliament and of the Council of 13 March 2024 on the transparency and targeting of political advertising, Official Journal of the European Union, L 2024/900.
4. European Commission, Tackling online disinformation: a European approach (COM(2018) 236 final), 2018.
5. European Commission, European Democracy Action Plan (COM(2020) 790 final), 2020.
6. Federal Communications Commission, In the Matter of Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts (CG Docket No. 23-362, Declaratory Ruling, FCC 24-17), 8 February 2024.

COURT PRACTICE

1. Case C-18/18, Glawischnig-Piesczek v. Facebook Ireland, ECLI:EU:C:2019:821 (Court of Justice of the European Union, 3 October 2019).
2. Case C-401/19, Poland v. Parliament and Council, ECLI:EU:C:2022:297 (Court of Justice of the European Union, 26 April 2022).
3. Sanchez v. France, Application no. 45581/15 (European Court of Human Rights [Grand Chamber], 15 May 2023).

NETWORK RESOURCES

1. EU Fundamental Rights Agency (FRA), Online content moderation: Current challenges in detecting hate speech, <https://fra.europa.eu/en/publication/2023/online-content-moderation>, 29 November 2023.
2. Future of Life Institute, Overview of all AI Act national implementation plans, <https://artificialintelligenceact.eu/national-implementation-plans>, 19 May 2025.
3. The Alan Turing Institute, Turing event at the Houses of Parliament explores impact of AI disinformation on elections, <https://www.turing.ac.uk/news/turing-event-houses-parliament-explores-impact-ai-disinformation-elections>, 30 January 2025.