

UDK: 004.738.5 : 007
004.738.5 : 316.77
34 : 004.8 (4-67 EU)(094.5)
Preliminary communication

Julia Małgorzata Kalus, PhD student

Department of Innovation and Digitalisation in Law, University of Vienna, Austria

E-mail: julia.malgorzata.kalus@univie.ac.at

ORCID: 0009-0008-9326-3999

DEEPFAKE - A PROBLEM ITSELF OR A SYMPTOM OF A DEEPER PROBLEM?

ABSTRACT

The paper tackles one of the most pressing issues regarding social media governance, namely, deep fakes and the spread of disinformation on social media platforms. First, the paper presents deepfakes in a broader perspective of synthetic content and aims to answer the question of whether deepfakes themselves are the problem or rather a symptom of a much more complex issue that lies at the root of the business model of social media platforms and the way they are governed. Additionally, the paper focuses on the threats that deepfakes pose to democracy. Secondly, the paper aims to outline the current regulatory framework dealing with deepfakes and disinformation, and it also touches upon the recently published Code of Practice on marking and labelling of AI-generated content. It critically examines the Code and its possible limitations. Lastly, the paper considers how the approach of public interest technology could be used as a possible theoretical framework to reframe the narrative on deepfake regulation.

Keywords: deepfakes, platform governance, social media platforms, EU law, disinformation, Digital Services Act, Artificial Intelligence Act, transparency

1. INTRODUCTION

1.1. Problem statement

The topic of deepfakes has been present in public debate and academic discourse for a couple of years. Some academics started to underline positive aspects of this technology and even claimed that some of the concerns around deepfakes are exaggerated.¹ Nonetheless, some studies conducted in the past years suggest that

¹ Rauchfleisch, A.; Vogler, D.; de Seta, G., *Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions*, Social Media and Society, volume 11, 09/2025, p. 3.

deepfakes indeed can contribute to the amplification of disinformation.² This is connected not only to their potential to deceive but also to the fact that they negatively influence public trust in media and institutions.³ At the same time, European Union (EU) legislation addressing deepfakes on online platforms has introduced mechanisms such as transparency obligations and risk assessments, which either give platforms a wide discretion in defining their own compliance or which are not addressed to platforms. This paper argues that this approach to deepfake regulation might be insufficient, considering the threats posed by them.

1.2. Research question and contribution

This paper aims to present the concept of deepfake disinformation in a broader context, not focusing on this phenomenon as a content problem that can be addressed through labelling obligation, but rather as an infrastructure problem, rooted in platforms' design. The central research question of this paper is to what extent the current EU legal framework addresses the topic of deepfakes through regulating content. In doing so, the paper aims to contribute to the current legal and policy scholarship on deepfakes and social media platforms by adopting the angle that goes beyond treating deepfakes as just a form of disinformation, and towards the structural role of platform design and recommender systems.

1.3. Outline of the paper

The structure of the paper is the following. The first section of the paper includes research questions, an outline, and the methodology used to develop the paper. In the second section, the paper defines deepfakes and presents them in a broader context of synthetic media. This section provides a definition of a deepfake and provides illustrative examples. Additionally, the section discusses threats that deepfakes pose to democracy, as well as presents deepfakes as a form of disinformation. The third section of the paper focuses on legal analysis. The section starts with a brief introduction of the legal frameworks and then discusses and critically examines each of them. Firstly, the Artificial Intelligence Act (AIA)⁴

² Weikmann, T.; Egelhofer, J.; Lecheler, S, *Beyond Credibility: The Effects of Different Forms of Visual Disinformation*, Journalism & Mass Communication Quarterly, volume 102, 12/2025, p. 1035.

³ Pawelec, M., *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*, Digital Society, volume 19, 1/2022, p. 18.

⁴ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down

is analysed together with the first draft of the Code of Practice on marking and labelling of AI-generated content (Code of practice on marking and labelling).⁵ Secondly, the Digital Services Act (DSA),⁶ together with the 2022 Strengthened Code of Practice on Disinformation, which has transitioned into the Code of Conduct on Disinformation (Code of Conduct)⁷ is addressed. Each of the legal instruments is assessed against the problems identified in the second section and presents limitations of an approach that targets content without addressing the underlying problems in the infrastructure. Lastly, the fourth section of the paper concludes the discussion and considers how the public interest technology theory could be applied to the existing legal framework to strengthen the focus on public interest in social media regulation.

1.4. Methodology

The paper adopts an interdisciplinary approach, combining legal primary sources with sources stemming from communication science, psychology, political science, and media studies. Nonetheless, the primary method used in the paper is a legal doctrinal analysis. Legal doctrinal analysis is used to examine the normative content and structure of the European Union secondary law, precisely the AIA and the DSA, as well as soft law instruments adopted as part of these legal frameworks, namely the recently published first draft of the Code of Practice on the marking and labelling of AI-generated content and the Code of Conduct on Disinformation. These instruments were selected as they represent the EU's primary regulatory tools for addressing AI-generated disinformation. The paper analysed these legal instruments in two ways: systematically, by situating them within the broader EU digital regulatory framework, and teleologically, by reading their provisions considering their stated objectives. The EU legislation was accessed through the official EUR-Lex database, and the soft law was accessed through the official website of the European Commission. Additionally, the paper

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) 2016/797and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L, 2024/1689 (hereinafter: AIA).

⁵ European Commission, First Draft Code of Practice on Transparency of AI-Generated Content, February 2026, <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content>, 1 March, 2026.

⁶ 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 277/1 (hereinafter: DSA).

⁷ European Commission, The Code of Conduct on Disinformation, February 2026, p. 1, 2025 <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>, 26 February, 2026.

is based on the latest state of the art in legal scholarship. Academic legal literature on the matter was identified based on well-established legal databases such as, JIPITEC, Verfassungsblog, Tech Policy Press, HeinOnline, Sage Journals, Cambridge Core, as well as by using Google Scholar.

Before the legal analysis, the paper builds on scholarship from communication science, political science, psychology, and media studies (section 2 of the paper) to establish the empirical and conceptual foundation of the research question. The author identified relevant academic literature in these fields by using databases such as Spring Nature, Sage Journals, SSRN, IEEE Xplore, Taylor Francis Online, and Google Scholar. The identified literature serves two purposes. First, to identify the threats deepfakes pose to democratic discourse and electoral integrity and their capacity to discourage political participation. Secondly, to argue that deepfakes can be understood as a form of disinformation, which explains the inclusion of the Code of Conduct on Disinformation among the instruments under analysis.

2. DEEPFAKES, DEMOCRACY AND DISINFORMATION

2.1. Deepfakes: definition and characteristics

According to the AIA, a deepfake is an "AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful".⁸ The adjective "deep" is used in the definition because this form of content is often created with the use of deep learning.⁹ Deepfakes are part of a broader category of synthetic media,¹⁰ which also encompasses media such as augmented reality, humanoid robots,¹¹ or virtual influencers.¹² While a comprehensive analysis of synthetic media and its nuances is beyond the scope of this paper,¹³ it may briefly be defined as content produced through the application of Artificial Intelligence (AI) technologies.¹⁴

⁸ Art. 3(60) AIA.

⁹ Tolosana, R. *et al.*, *Deepfakes and beyond: A Survey of face manipulation and fake detection*, Information Fusion, volume 64, 12/2020, p. 131.

¹⁰ Van der Sloot, B.; Wagenveld, Y., *Deepfakes: regulatory challenges for the synthetic society*, Computer Law and Security Review, volume 46, 9/2022, p. 1-2.

¹¹ Van der Sloot, B., *Regulating the Synthetic Society*, Hart Publishing, Oxford 2024, p. 14.

¹² Kalpokas, I., *Problematising reality: the promises and perils of synthetic media*, SN Soc Sci, volume 1, 11/2020, p. 1-2.

¹³ See: Millièrè, R., *Deep Learning and Synthetic Media*, Synthese, volume 200, 5/2022.

¹⁴ Kalpokas, I., *op. cit.* note 12, p. 1-2.

Deepfakes are no longer only a domain of technically skilled actors, as the tools necessary for their creation have become progressively more accessible and user-friendly.¹⁵ One of the examples of such tools is the DeepFaceLab, which can be used for face swaps and is available as an open-source software for Windows and macOS.¹⁶ The ease and speed connected to the creation and dissemination of deepfakes are what make them a dangerous tool. This is because tools that do not require programming expertise have made production accessible to virtually anyone, and a deepfake can reach millions of online users through algorithmic amplification before any correction is issued.¹⁷ Deepfakes are considered to be very persuasive mainly because of the use of AI in their production, and in particular, a technique called a generative adversarial network.¹⁸

In the next section, the focus is on the negative impact of deepfakes on democracy, and attention is given to political disinformation. However, it must be acknowledged that 90% of deepfakes present online are a form of violence against women.¹⁹ This form of non-consensual pornographic content also has an indirect impact on democracy as it discourages women from participation in public life.²⁰ One of the recent examples of such a malicious use of deepfake technology is the recent scandal in which German television presenter Collien Fernandes became a victim of a deepfake campaign orchestrated by her husband.²¹ Nonetheless, this paper will not focus specifically on this form of deepfakes since a separate legal framework is considered for this form of deepfakes, as recently, deepfake pornography has gotten attention from the European Parliament (EP). The EP announced that it will call for stronger enforcement of EU law to combat the use

¹⁵ Labuz, M., *A Teleological Interpretation of the Definition of Deep Fakes in the EU Artificial Intelligence Act—A Purpose-Based Approach to Potential Problems With the Word "Existing"*, Policy and Internet, volume 17, 1/2025 p. 2.

¹⁶ DeepFaceLab, How DeepFaceLab Works, <https://deepfacelab.org/#choose>, 31 March 2026.

¹⁷ Momeni, M., *Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation*, Journal of Creative Communications, volume 20, 10/2024, p. 47.

¹⁸ Nieweglowska, M.; Stellato, C.; Sloman, S. A., *Deepfakes: Vehicles for Radicalization, Not Persuasion*, Current Directions in Psychological Science, volume 32, 3/2023, p. 237.

¹⁹ Kira, B., Deepfakes, The Weaponisation of AI Against Women and Possible Solutions, Verfassungsblog, June 2024, <https://verfassungsblog.de/deepfakes-ncid-ai-regulation/>, 12 February, 2026.

²⁰ Shukla, V., Deepfakes and Elections: The Risk to Women's Political Participation, TechPolicy Press, February 2024, <https://www.techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation/>, 20 February, 2026.

²¹ Parker, J.; Völk, K., How deepfake porn scandal surrounding TV star rocked Germany, BBC News, 29 March, 2026, <https://www.bbc.com/news/articles/c9v1m4j47e2o>, 31 March, 2026.

of AI tools to create illegal sexual content.²² Therefore, this malicious use of deepfakes falls outside the scope of this paper. Nonetheless, the author wanted to acknowledge this malicious form of deepfakes.

2.2. Threats posed by deepfakes to democracy

Trust is an essential element required for the effective functioning of democratic processes.²³ If a shared perception of reality is lacking, it is extremely difficult to tackle national and global problems, such as the financial crisis or the rule of law crisis.²⁴ Deepfakes lead to citizens questioning whether what they see is "true", which consequently contributes to a decay of trust in shared facts and truths.²⁵ Therefore, they significantly contribute to the rise of a so-called post-truth society, where it is impossible to distinguish between facts and opinions.²⁶ Some might argue that the spread of half-truths in the digital sphere is not a novelty,²⁷ but the high quality of the content generated by deepfake technology makes it an unprecedentedly dangerous tool.²⁸ It must be pointed out that deepfakes contribute to citizens losing trust not only in democratic processes, such as elections, but also in each other.²⁹ This phenomenon can be described as a 'societal trust decay'.³⁰ As a result of a loss of trust in each other, individuals might feel discouraged from participating in democratic processes.³¹

Deepfakes can change attitudes towards politicians and evoke strong emotions.³² One of the examples of a deepfake that was met with a strong reaction from the

²² European Parliament, Tackling AI deepfakes and sexual exploitation on social media (EP) Briefing, European Parliament, 15 January, 2026, <https://www.europarl.europa.eu/news/en/agenda/plenary-news/2026-01-19/8/tackling-ai-deepfakes-and-sexual-exploitation-on-social-media>, 15 February, 2026.

²³ Chesney, R.; Danielle, C., *Deepfakes: A Looming Crisis for National Security, Democracy and Privacy?* Lawfare, February 2018, <https://www.lawfaremedia.org/article/deepfakes-looming-crisis-national-security-democracy-and-privacy>, accessed 26 February, 2026.

²⁴ *Ibid.*

²⁵ Pawelec, M., *op. cit.* note 3, p. 4.

²⁶ Chesney, R.; Citron, D., *Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics*, Foreign Affairs, volume 98, 2/2019, p. 155.

²⁷ Barber, A., *Freedom of expression meets deepfakes*, Synthese, volume 202, 7/2023, p. 14.

²⁸ *Ibid.*

²⁹ Pawelec, M., *op. cit.* note 3, p. 12.

³⁰ *Ibid.*, p. 19.

³¹ *Ibid.*, p. 18.

³² Dobber, Tom *et al.*, *Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?*, The International Journal of Press/Politics, volume 26, 1/2021, p. 70.

public is a video in which Ukrainian President Volodymyr Zelensky was encouraging Ukrainians to surrender.³³ This is an example of a malicious deepfake manipulation, as it could alter public perception of Ukraine's commitment to defeating Russia's aggression and consequently negatively affect European security. In the current Iranian war, deepfakes are being used as pro-war propaganda. Social media researchers identified a wave of deepfakes on online platforms, in particular on TikTok, portraying women in military uniforms saying "Habibi, come to Iran".³⁴ These deepfakes were spread on social media platforms through fake accounts. Even though deepfakes in question were a clear example of disinformation, as women are not allowed to serve in the Iranian army, deepfakes got a lot of reactions and positive comments, such as "yes".³⁵ This example of a deepfakes campaign shows how deepfakes can reach users and gather their reaction through the use of fake social media accounts led by fake influencers³⁶, and it shows how social media infrastructure can be used to spread disinformation.

2.3. Deepfakes as a disinformation tool

The previous sub-section briefly presented threats that deepfakes pose to democratic societies. Even though it has not yet been measured whether a particular deepfake had an impact on the change of a person's political views, and it is doubtful whether such a measurement would even be possible, it has been established that deepfakes contribute to the amplification of disinformation. This section will consider whether deepfakes fulfil the definition of disinformation.

One of the first definitions of disinformation in European Law appeared in the 2018 version of the Code of Practice on Disinformation.³⁷ Based on definitions of disinformation in the EU law, including 2018 Code of Practice on Disinformation,

³³ Twomey, J. J.; Linehan, C.; Murphy, G., Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine, *The Conversation*, October 2023, 2023, <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393>, 31 March, 2026.

³⁴ Marco Silva, Flood of AI soldiers push pro-Iran messages during war, *BBC News*, 12 March 2026, <https://www.bbc.com/news/videos/cy9gw900dp9o>, 1 April, 2026.

³⁵ Berger, E., 'They feel true': political deepfakes are growing in influence – even if people know they aren't real, *The Guardian*, 28 March, 2026, <https://www.theguardian.com/technology/2026/mar/28/military-deepfakes-ai-propaganda-money>, 1 April 2026.

³⁶ *Ibid.*

³⁷ Bleyer-Simon, K.; Reviglio, U., Defining Disinformation across EU and VLOP Policies, *European Digital Media Observatory*, October 2024, p.11, <https://edmo.eu/wp-content/uploads/2024/10/EDMO-Report-%E2%80%93-Defining-Disinformation-across-EU-and-VLOP-Policies.pdf>, 1 April, 2026.

it is possible to distinguish key elements that need to be fulfilled for a content to be considered disinformation: (i) content must be verifiably false or misleading, (ii) it must have the potential to cause harm to society, (iii) it is spread intentionally (iv) possibly for economic or political gain.³⁸ When it comes to deepfakes, the first element is fulfilled as at the core of the deepfake definition lies the fact that they falsely resemble reality, and are therefore misleading. Considering the discussion in the previous section, in which potential risks of deepfakes to democracy and internal security were presented, the second element is also fulfilled. The last element relating to political and economic gain seems to also be fulfilled, as in most cases, political deepfakes are spread to serve a specific political agenda.³⁹

Some scholars argue that deepfakes are not a new threat and that they rather belong to the broader disinformation landscape.⁴⁰ At the same time, it is debatable whether deepfakes alone could actually contribute to the spread of disinformation. Some studies suggest that, to be successful, deepfake campaigns would need to be precisely targeted,⁴¹ which strongly links their success in disinformation to social media infrastructure,⁴² and fake social media accounts, bots, and recommender systems.⁴³

Additionally, studies conducted in the past years confirm that algorithms on social media platforms contribute to the spread of emotionally charged content.⁴⁴ One study, which focused on platforms such as TikTok, Instagram and Facebook, showed that the negative and emotionally charged content around port security in Split was prioritised by algorithms on these platforms.⁴⁵ Additionally, the study

³⁸ *Ibid.*, p. 3.

³⁹ Szabo, L.V.; Bader, S., *Deepfake content in political communication*, Convergence: The International Journal of Research into New Media Technologies, volume 0, 2026, p. 10.

⁴⁰ Łabuz, M.; Nehring, C., *On the way to deep fake democracy? Deep fakes in election campaigns in 2023*, year 2024, 23:454–473 European Political Science, volume 23, 12/2024, p. 465.

⁴¹ Fathaigh, R. et.al., *Microtargeted propaganda by foreign actors: An interdisciplinary exploration*, Maastricht Journal of European and Comparative Law, volume 28, 11/ 2021, Vol. 28(6) 2021, p. 866.

⁴² Birrer, A.; Just, N., *What We Know and Don't Know about Deepfakes: An Investigation into the State of the Research and Regulatory*, New Media and Society, volume 27, 12/2025, p. 6828.

⁴³ Von Sikorski, Ch.; Hameleers, M., *Disinformation in the Age of Artificial Intelligence (AI): Implications for Journalism and Mass Communication*, Journalism & Mass Communication Quarterly, volume 102, 11/2025, p. 944.

⁴⁴ See: Gombar, M.; M. Boban, *Research on the Impact of Algorithmic Echo Chambers on Perceptions and Attitudes of Social Network Users in a Digital Society*, MIPRO 48th ICT and Electronics Convention, Opatija, Croatia, June 2025, <https://ieeexplore.ieee.org/document/11131918>, 1 April, 2026.

⁴⁵ *Ibid.*, p. 1028-1029.

showed that algorithms are not a neutral technology, but have an active role in spreading disinformation.⁴⁶ The results of the study are one of the testimonies that the discussion on deepfake disinformation should not be focused only on addressing the content, but rather the overall infrastructure for its spread.

3. REGULATORY FRAMEWORK

In the last section, the topic of deepfakes has been framed as a symptom of a disease rather than an illness. It was argued that even though deepfakes are a sophisticated method of disinformation, and their effects can be very harmful for public trust in democratic societies, the problem seems to lie in the channels that focus on distributing this form of content.

This section will deal with the analysis of the EU Law that addresses deepfakes and content on social media platforms. First, the AIA will be analysed, and specifically, Article 50(4) of the AIA will be addressed, as well as the Code of Practice on marking and labelling of AI-generated content. The next sub-section will focus on the DSA, which is a crucial puzzle in the regulatory landscape regarding deepfakes, as they are mainly spread on major technological platforms.⁴⁷

3.1. The Artificial Intelligence Act

The AIA includes the first explicit definition of a deepfake in European Law.⁴⁸ Additionally, Article 50 of the AIA imposes a transparency obligation on deployers of AI systems capable of producing deepfakes. It is worth noting that the AIA is a market-focused legislation, meaning that it primarily regulates the process of developing and bringing AI systems to the market,⁴⁹ and was designed as a product safety regulation.⁵⁰ Therefore, even though ensuring a high level of protection of democracy is mentioned in the scope of this legislation,⁵¹ it is not its main objective. Additionally, the AIA does not focus on disinformation explicitly in its provisions. At the same time, the EU legislator has acknowledged the potentially

⁴⁶ *Ibid.*, p. 1032.

⁴⁷ Soundarya, B.; Gururaj, H. L., *Deepfake detection: critical review of state-of-the-art approaches and future perspectives*, Discover Applied Sciences, volume 8, 1/2026, p. 20.

⁴⁸ See: Art. 3(60) AIA.

⁴⁹ Almada, M.; Petit, N., *The EU AI Act: Between the Rock of Product Safety and the Hard Place of*

Fundamental Rights, Common Market Law Review, volume 62, 2/2025, p. 104.

⁵⁰ Art. 1(1) AIA.

⁵¹ *Ibid.*

harmful effects of synthetic content on the information landscape and public knowledge.⁵² Further, Recital 134 explicitly addresses deepfakes and underlines the importance of transparency and labelling of deepfake content.⁵³

3.1.1. Actors in the AIA

To understand the transparency obligations in the AIA that apply to AI-generated content, including deepfakes, it is necessary to explain the main actors in this legislation. The AIA applies to providers and deployers of AI systems. Both are any natural and legal persons, or other bodies such as public authorities.⁵⁴ A provider of an AI system is an actor that develops an AI system or a general-purpose AI model and places it on the market,⁵⁵ whereas a deployer uses it under its authority, unless it is in the course of a personal activity.⁵⁶

In the context of deepfakes, providers of AI systems would usually be companies that develop AI systems that can be used to produce deepfakes, such as the previously mentioned DeepFaceLab. Deployers of AI systems are actors that use the AI system, in this case, natural persons creating deepfakes. What raises enforcement concerns is the general exemption to the scope of the AIA in Article 2(10), which states that the regulation does not apply to persons who use AI systems in a purely private capacity.⁵⁷ It is claimed that since deepfakes are mostly created with the intent to be published publicly, this exception should not be applicable.⁵⁸ Therefore, a person creating a deepfake would be considered a deployer, and the provisions of the AIA should apply to a deepfake creator.⁵⁹ This interpretation is reasonable and would be preferred in order to ensure compliance with transparency obligations. Nonetheless, it is difficult to assess what exactly qualifies as a "personal activity" and it could potentially lead to overreliance on this exemption,⁶⁰ and consequently limit adherence to transparency obligations.

⁵² Recital 133 AIA.

⁵³ Periello, L.E., *Blurred realities: Legal strategies for the deepfake era*, Maastricht Journal of European and Comparative Law, volume 0, 3/2026. p. 15.

⁵⁴ Arts. 3(3),3(4) AIA

⁵⁵ Art. 3(3) AIA.

⁵⁶ Article 3(4) AIA.

⁵⁷ Periello, L.E., *op. cit.* note 54, p. 25.

⁵⁸ *Ibid.*

⁵⁹ *Ibid.*

⁶⁰ *Ibid.*

3.1.2. Do deepfakes qualify as high-risk?

The AIA adopts a risk-based approach and qualifies AI systems into categories based on the level of risk they pose to fundamental rights and safety of individuals.⁶¹ An AI system can be classified as a minimal risk,⁶² limited risk,⁶³ high-risk,⁶⁴ or unacceptable risk (prohibited).⁶⁵ Depending on the risk level, the AIA assigns different obligations to providers and deployers.

When it comes to high-risk AI systems, Article 6(2) of the AIA states that AI systems can be qualified as high-risk if they are included in Annex III. In the context of political disinformation, it can be observed that point 8 in Annex III seems to be the most relevant, as it refers to "administration of justice and democratic processes".⁶⁶ However, it applies explicitly to AI systems that have the potential to influence the outcome of elections or a referendum.⁶⁷ Therefore, it could be understood that this provision relates to AI systems that are directly used in voting systems, such as systems used for vote counting or systems used in the assistance with the voting procedure.⁶⁸ Additionally, recital 62, read in conjunction with Article 6 AIA and point 8(b) of Annex III AIA, specifies that it relates to the right to vote as per Article 39 of the EU Charter of Fundamental Rights. Therefore, it can be stated that this provision does not apply to AI systems used for the purpose of political disinformation.⁶⁹ As analysed in section two of this paper, there are diverging outcomes of studies on whether deepfakes are a successful disinformation tool. Consequently, considering recital 62, it can be stated that AI systems used to generate political disinformation do not qualify as high-risk. However, if they are used to directly infringe on the person's right to vote, obligations for high-risk AI systems would apply. Therefore, even though the AIA includes provisions of high-risk AI systems, it is not clear whether they would apply to the AI systems used to spread disinformation or content that qualifies as disinformation.

⁶¹ Gasser, U.; Palfrey, J., *Advanced introduction to Law and Digital Technologies*, Elgar Advanced Introductions, 2025, p. 58.

⁶² Ebers, M., *Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act*, European Journal of Risk Regulation, volume 16, 9/2025, p. 687.

⁶³ Recital 132 AIA.

⁶⁴ Art.6 AIA.

⁶⁵ Art. 5 AIA.

⁶⁶ Krack, N.; Dutkiewicz, L.; De Meyere, J., *Generative Artificial Intelligence and Disinformation*, in Ginsborg, L.; Gori, P. (eds), *Disinformation: A Multi-Disciplinary Analysis*, Springer Nature, Switzerland 2026, p. 275

⁶⁷ Annex III Point 8 AIA

⁶⁸ Krack, N., Dutkiewicz, L. De Meyere, J., *op. cit.* note 65, p. 276.

⁶⁹ *Ibid.*

Consequently, AI systems used to generate deepfakes and spread deepfake disinformation would most likely not be classified as high-risk obligations, and the provisions for high-risk AI systems would likely not apply to them. The scope of this provision must be further specified by the case law of the Court of Justice of the European Union.

3.1.3. Transparency obligations for providers and deployers of AI systems generating deepfakes

Another important provision in the AI regarding deepfakes is Article 50, which establishes transparency obligations for certain AI systems. Article 50 imposes obligations on both providers and deployers of AI systems that interact with natural persons or generate synthetic content. Most relevantly for this paper, Article 50(4) requires "deployers of an AI system that generates or manipulates image, audio, or video content constituting a deep fake"⁷⁰ to disclose that the content has been artificially manipulated. This provision does not apply in cases in which authorization has been given by a person presented in images, videos, or audio content, or if it was *inter alia* produced to exercise the freedom of expression or serve as a parody.⁷¹ Even though the inclusion of this provision is a noble development and a step towards a more transparent online environment, the provision remains problematic. Many open-ended questions arise when reading the text of the provision, and there are several critical points that are connected to the discussion in the previous section of deepfakes being a symptom rather than a disease.

First, assessing what qualifies as a legitimate speech and what deepfakes can be regarded as a freedom of expression is problematic. The AIA does not state clearly where the line lies between, for instance, artistic expression and a deepfake that must be labelled.⁷² Lack of clarity on what constitutes a parody or artistic expression could lead to harmful abuses of this exemption.⁷³

Secondly, there is a general concern regarding the enforcement of transparency obligations. The territorial scope of the AIA, included in Article 2(1)(a)(b)(c), covers not only actors based in the EU, but also actors who deploy or provide an AI system whose output reaches the EU. Consequently, deepfakes reaching EU

⁷⁰ Art. 50(4) AIA.

⁷¹ *Ibid.*

⁷² Periello, L.E., *op. cit.* note 54, p. 25.

⁷³ *Ibid.*

citizens must be subjected to transparency obligations.⁷⁴ It means that in theory, the scope of the AIA is extraterritorial. However, tracking individuals who produce and share deepfakes and enforcing the transparency obligation across international borders, especially given the decentralized channels, risks being a Sisyphean task.⁷⁵ In order to provide for effective enforcement, measures such as cross-border mechanisms or international agreements should be in place.⁷⁶ More detailed discussion on the extraterritorial enforcement of the transparency obligations applicable to deepfakes constitutes another research field, which falls outside the scope of this paper.

To conclude, Article 50(4) addresses the symptom, being, for the purpose of this paper, the existence of an AI-generated content that could harm democratic processes, without engaging with the structural problems and issues related to the infrastructure. It is because the regulators decided to include deepfakes as a limited risk category and, therefore, subjected them to mostly transparency obligations.⁷⁷ It must be noted that according to Article 50(6), AIA does not exclude that transparency obligations go hand in hand with high-risk obligations. Therefore, it is not unlikely that in the future some deepfake applications might be considered high-risk as well.⁷⁸

3.1.4. Code of Practice on marking and labelling of AI-generated content

To support the implementation of the obligations in Article 50(4) of the AIA, the AI Office, a body designated by the European Commission that gathers experts on AI,⁷⁹ initiated the process of creating the Code of Practice on marking and labelling of AI-generated content. The aim is to allow providers and deployers to prepare for compliance before the rules take effect in August 2026.⁸⁰ The Code of Practice on marking and labelling finds its legal basis in Article 95 of the AIA. Adherence to the Code of Practice is voluntary, and it is addressed to its signatories.⁸¹ At the time of writing this paper, only the first draft of the Code of Practice published in February 2026 is available. According to the European

⁷⁴ *Ibid.*

⁷⁵ *Ibid.*, p. 27.

⁷⁶ *Ibid.*

⁷⁷ *Ibid.*, p. 23.

⁷⁸ *Ibid.*, p. 22.

⁷⁹ Art. 64(1) AIA.

⁸⁰ European Commission, *op. cit.* note 5.

⁸¹ European Commission, First Draft Code of Practice on Transparency of AI-Generated Content, February 2026, p. 2, <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content> 1 March, 2026.

Commission, the second draft of the Code of Practice is foreseen to be published in March 2026.⁸² Therefore, the content of the code discussed in this paper might not be up to date once it is published. Nonetheless, the Code of Practice is still worth examining. The limitations of the code, such as its voluntary nature, show the overall approach of the Commission to the regulation of AI-generated content, which is focused largely on transparency obligations and which lacks clear guidance on enforcement.

The Code on Practice proposes an EU-wide icon that would signal that the content is a deepfake.⁸³ This initiative is a positive development. Nonetheless, it stems from the idea that there is a willingness from deployers of the AI systems for the purpose of deepfake creation to comply with legal obligations. However, the aim of deepfake creators is often to destabilise and spread disinformation.⁸⁴ Therefore, specifications in relation to labelling the content, yet again, do not solve the root of the problem. Additionally, a lot of emphasis in the Code of Practice is placed on AI literacy.⁸⁵ While it is a measure that must go hand in hand with other legal obligations, the burden cannot be solely placed on the individual seeing a deepfake, considering the power imbalance between users of platforms and platform administrators. The DSA aims to mitigate these imbalances, which is why both legislations are jointly analysed in this paper.

3.2. The Digital Services Act (DSA)

The DSA became an enforceable law in February 2024.⁸⁶ The topic of deepfakes is not directly addressed in the legislation, and in contrast to the AIA, the DSA does not include an explicit definition of a deepfake. It is worth mentioning that the DSA does not focus on AI but rather on digital intermediaries.⁸⁷ The legislation distinguishes between Very Large Online Platforms (VLOPS) and Very Large Search Engines (VLSE). Platforms having more than 45 million users per month qualify as VLOPS.⁸⁸ The word 'disinformation' appears in the DSA thirteen times,⁸⁹ but only in the recitals of the regulation, and it does not include an explicit definition of disinformation in its provisions.

⁸² *Ibid.*

⁸³ *Ibid.*, p. 24 (Measure 1.2.2).

⁸⁴ Fathaigh, R. *et.al.*, *op. cit.* note 40, p. 857.

⁸⁵ European Commission, *op. cit.* note 79., p. 21(recital e).

⁸⁶ Art. 93 DSA.

⁸⁷ See: Art. 2(1) DSA.

⁸⁸ Art. 33(1) DSA.

⁸⁹ Recitals 2, 9, 69, 83, 84, 88, 95, 104, 106, 108 DSA.

3.2.1. Risk assessment and risk mitigation

One of the most important obligations of VLOPS is risk assessment, introduced in Article 34 of the DSA. The DSA distinguishes four categories of risks,⁹⁰ such as risks to democratic processes or risks posed by illegal content.⁹¹ The concept of risks is also a crucial element of the AIA.⁹² However, the DSA does not adopt a risk-based approach, but rather introduces a concept of systemic risk.⁹³ The understanding of the concept of systemic risk in the DSA is broad, and it does not include an exhaustive list of risks.⁹⁴ It is the platforms themselves that must identify and analyse them.⁹⁵

Next to the risk assessment, VLOPs are also required to introduce mitigation measures.⁹⁶ The DSA includes various examples of mitigation measures, such as awareness-raising,⁹⁷ or adapting the design of the interfaces.⁹⁸ Notably, the last mitigation measure listed in Article 35 refers to "generated or manipulated image, audio or video that appreciably resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful".⁹⁹ Although deepfakes are not explicitly mentioned in the provision, they clearly fall within its scope, as it indirectly sets out a definition of a deepfake, but without mentioning AI.¹⁰⁰ This proposed mitigation measure requires intermediaries to inform users that the content was manipulated by "prominent markings".¹⁰¹ Therefore, it is similar to Article 50(4) of the AIA. The transparency requirements, such as marking and labelling, might not be the most effective, as according to a recent psychological study, individuals tend to rely on the content of deepfakes even if they are made aware that it is not true or correct.¹⁰²

⁹⁰ Recital 79 DSA.

⁹¹ Art. 34 DSA.

⁹² Krack, N.; Dutkiewicz, L.; De Meyere, J., *op. cit.* note 65, p. 279.

⁹³ Recital 76 DSA.

⁹⁴ *Ibid.*

⁹⁵ Art. 34 (1) DSA

⁹⁶ Art. 35 DSA

⁹⁷ Art. 35(i) DSA.

⁹⁸ Art. 35(k) DSA.

⁹⁹ *Ibid.*

¹⁰⁰ Krack, N.; Dutkiewicz, L.; De Meyere, J., *op. cit.* note 65, p. 278.

¹⁰¹ Art. 35 (k) DSA.

¹⁰² Clark, S.; Lewandowsky, S., *The Continued Influence of AI-Generated Deepfake Videos despite Transparency Warnings*, Communications Psychology, volume 4, 1/2026, p. 10.

3.2.2. Recommender system transparency

So far, this section has focused on the obligations of VLOPs regarding risk assessment and risk mitigation. The DSA also includes an important provision dealing with obligations for recommender systems,¹⁰³ Article 27, which is arguably the most directly relevant provision for this paper's central argument. This Article requires intermediaries to present in their terms and conditions at least one option for a recommender system that does not include profiling, and to make their recommended systems transparent.¹⁰⁴ The DSA acknowledges that recommender systems play an important role in platforms' business models and it defines them as a tool that algorithmically suggests, ranks, and prioritises information and is capable of amplifying certain messages.¹⁰⁵

The option to allow individuals to decide not to have a recommender system based on profiling can be seen as a step towards empowering individuals, a process that is considered to be an important part of fair governance.¹⁰⁶ However, the provision does not prohibit systems that use engagement-maximising recommendations, which is the main mechanism that allows for emotionally engaging content, such as deepfakes, to be spread on social media platforms.¹⁰⁷ Therefore, the provision only focuses on transparency, rather than addressing the core of the problem. Secondly, there are some practical limitations to this provision and enforcement issues exemplified by researchers in non-governmental organisations, such as the EU Disinfo Lab.¹⁰⁸ One investigation conducted by the Dutch NGO, Bits of Freedom,¹⁰⁹ showed that even when users decide to disable profiling on Instagram, some of the features on the application, such as direct messages, become inaccessible.¹¹⁰ This case well exemplifies how big technological companies try to circumvent legal obligations imposed on them. Additionally, it is very unlikely

¹⁰³ Art. 27 DSA.

¹⁰⁴ Art. 27(3) DSA.

¹⁰⁵ Recital 70 DSA.

¹⁰⁶ Gasser U.; Palfrey, J., *Advanced Introduction to Law and Digital Technologies*, Elgar Advanced Introduction 2025, p. 120.

¹⁰⁷ Milli, S. *et al.*, *Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media*, *PNAS Nexus*, volume 4, 3/2025, p. 4-5.

¹⁰⁸ See: EU Disinfo Lab, About, <https://www.disinfo.eu/about-us/problem/>, 28 February, 2026.

¹⁰⁹ See: Zenger, R.; Austin, E.; Knaap, M., Judge in the Bits of Freedom vs. Meta lawsuit: Meta must respect users' choice, 2 Oktober 2025, <https://www.bitsoffreedom.nl/2025/10/02/judge-in-the-bits-of-freedom-vs-meta-lawsuit-meta-must->, 25 February, 2026.

¹¹⁰ See: Aşmane, K.L.; McNamee, J., Malicious semi-compliance as the platform approach to EU regulation, EU DisinfoLab, 25 November, 2025, <https://www.disinfo.eu/malicious-semi-compliance-as-platform-approach-to-eu-regulation/>, 28 February, 2026.

that an average user will read the provision and disable the profiling features.¹¹¹ The effectiveness of the provision is consequently not only based on the will of users to change the settings but also on how well the platforms inform users about this possibility.¹¹² This issue could be connected to the phenomenon of dark patterns. However, this discussion falls outside of the scope of this paper.

3.2.3. Code of Conduct on Disinformation: Can it be a hope for the future?

The 2022 Code of Practice on Disinformation is a crucial soft law when it comes to EU policy efforts in tackling disinformation. At the start of 2025, the Commission approved incorporating the 2022 Code of Practice, now reformulated as the Code of Conduct on Disinformation, into the DSA framework.¹¹³ The parties adhere to the Code of Conduct voluntarily.¹¹⁴ The transition into the Code of Conduct also strengthens the character of the document, making the commitments of the signatories auditable.¹¹⁵ However, adherence to the Code of Conduct remains voluntary.¹¹⁶ As of 13 February 2025, the Code has forty-two signatories, including major online platforms and search engines like Meta and Google, fact-checkers, and representatives of civil society organisations.¹¹⁷ Notably, online platform X withdrew its signature from the Code of Practice in 2023.¹¹⁸ The voluntary nature of the Code and lack of sanction for its implementation render it a rather insufficient mechanism. Additionally, the current trend of social media platforms, withdrawal from fact-checking practices, with Meta being a primary example,¹¹⁹ is a concerning development. The abolishment of Meta's fact-

¹¹¹ Hakkarainen, J.; Savolainen, L., *Individual choice, collective effects: recommender systems, law by design, and the DSA's double choice architecture*, Information, Communication and Society, volume 0, 12/2025, p. 6-7.

¹¹² *Ibid.*

¹¹³ European Commission, The Code of Conduct on Disinformation, 13 February, 2025, <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>, 20 February, 2026.

¹¹⁴ *Ibid.*, p. 1.

¹¹⁵ Ginsborg, L.; Gori, P. *Disinformation: a multi-disciplinary analysis*, Springer 2026, p. 557.

¹¹⁶ *Ibid.*

¹¹⁷ *Ibid.*

¹¹⁸ Polona, C., Fact-checking and content moderation, European Parliamentary Research Service, February 2025, p. 2, [https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769493/EPRS_ATA\(2025\)_12_February_2026](https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769493/EPRS_ATA(2025)_12_February_2026).

¹¹⁹ Körömi, C.; Haec, P; Cheslow, D, Zuck goes full Musk, dumps Facebook fact-checking program 7 January 2025, POLITICO, <https://www.politico.eu/article/mark-zuckerberg-full-elon-musk-dump-facebook-fact-checker/>, 20 February 2026.

checking mechanisms is momentarily limited only to the US.¹²⁰ Last year, Marc Zuckerberg openly called the EU's content moderation practices "censorship".¹²¹ It remains to be seen how Meta will continue complying with EU regulations. The voluntary nature of the Code of Conduct is even more problematic in the current political atmosphere, where companies fail to comply even with binding obligations. Lack of compliance with the DSA by X resulted in a 120 million euro fine, which was issued at the end of 2025.¹²² After the issuance of the fine, the CEO of X called for the abolishment of the EU and expressed his support for anti-democratic EU parties.¹²³ This case shows that platforms lack the willingness to comply with EU law. In these times, it is important that the EU addresses the core of the issue, being the platform's infrastructure.

3.3. Legal framework: Concluding Remarks

Article 50(4) of the AIA, including labelling and marking obligations, as well as the transformation of the 2022 Strengthened Code of Practice on Disinformation into the Code of Conduct in 2025, are needed improvements to the existing legal framework. However, currently, the AIA and the DSA focus mostly on transparency obligations. While DSA addresses the issue of recommender systems, the obligations in Art. 27 are not enforced, and platforms find ways to circumvent them. Additionally, a lot of discretion is left to platforms to identify risks and introduce mitigation measures. Transparency obligations in Article 50 raise challenges in relation to extraterritorial enforcement and application of freedom of expression exemption, as well as the personal use exemption. The official version of the Code of Practice on marking and labelling of AI-generated content, which is expected to be published in summer 2026, could offer more clarity on that. However, in the author's opinion, the code will likely uphold its transparency-focused nature and only address technical aspects of how the labelling should be designed.

¹²⁰ *Ibid.*

¹²¹ Powers, C., Zuckerberg calls on US to defend tech companies against European 'censorship', Euroactiv, 12 January, 2025, <https://www.euractiv.com/news/zuckerberg-calls-on-us-to-defend-tech-companies-against-european-censorship/>, 26 February, 2026.

¹²² European Commission, Commission fines X €120 million under the Digital Services Act, European Commission Press Corner, 5 December 2025, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2934, 1 April, 2026.

¹²³ European Digital Rights (EDRI), Press Release: EU stands up to Big Tech with €120 million fine to X, 8 December 2025, EDRI Resources, <https://edri.org/our-work/eu-stands-up-to-big-tech-with-e120-million-fine-to-x/>, 1 April 2026.

4. CONCLUSION: PUBLIC INTEREST TECHNOLOGY AS A WAY FORWARD

The paper argued that deepfakes pose threats to democracy and can be a successful disinformation tool when paired with adequate dissemination tools. It was also analysed that algorithms of social media platforms tend to favour emotionally charged content and therefore amplify the spread of political deepfakes. Section 3 of this paper focused on the analysis of provisions of the AIA and the DSA, together with relevant soft laws, with an aim to address the research question of the paper and assess the extent to which these frameworks address deepfakes solely as a content issue. Based on the doctrinal analysis conducted, it can be stated that these two regulations are mainly focused on transparency obligations. Provisions that aim to address the core of the issue, such as Article 27 of the DSA, do not prohibit the use of algorithms that amplify radical content. In addition, as shown in the investigation of the NGO Bits of Freedom, platforms do not fully comply with this obligation. When it comes to the AIA, AI systems used to create deepfakes are only subject to transparency obligations that raise enforcement concerns.

To reframe the approach to regulating social media platforms and deepfakes, it is important to consider how to organise and better govern informational infrastructure. One of the promising theoretical frameworks that could be used in this context is the concept of public interest technology. Public Interest Technology is both a concept and a field of study, which is rooted in the idea that technology should be developed in the interest of the public, not to serve corporations or politics.¹²⁴ It is strongly connected to the public interest concept, which is one of the main concepts in political and legal theory and dates back to the works of Plato and Aristotle.¹²⁵ Within legal scholarship, public interest was described as a bridge between democratic values and regulation.¹²⁶

Relying on the concept of public interest technology in regulating social media platforms would put the welfare of the public at the centre, rather than the economic interests of platforms, which is currently the case. When it comes to the discussed legal frameworks, using public interest technology could mean

¹²⁴ Carayannis, E. G.; Erokhin, V., *Democracy, Environment, and Technology: The Cases of AI and Fusion Energy* Springer Nature, Switzerland 2025, p. 49.

¹²⁵ Züger, T.; Asghari, H., *AI for the public. How public interest theory shifts the discourse on AI*, AI & SOCIETY volume 38, 4/2023, p. 816.

¹²⁶ *Ibid.*

involving the public in the risk assessment process under Article 34 of the DSA¹²⁷ or prohibiting recommender systems that amplify the spread of emotionally driven content. Another possibility would be to include deepfakes in the category of high-risk AI systems and therefore impose more strict obligations on providers and deployers of AI systems that generate deepfakes. Alternatively, using the public interest technology approach could mean that the personal activity exception in Article 2(1) AIA would be read by courts in a restrictive manner, so it does not cover persons developing deepfakes. Even though limited transparency obligations are not an adequate measure, this would allow for some level of protection to be given to users of online platforms encountering deepfake content.

The development of decentralized platforms is also an idea that is often presented in the academic discourse.¹²⁸ Exploring how the decentralized social media platforms could fit into the European legal system and national laws of member states represents an idea for further research.

REFERENCES

BOOKS AND ARTICLES

1. Almada, M.; Petit, N., *The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights*, Common Market Law Review, volume 62, 2/2025, pp. 85-120.
2. Barber, A., *Freedom of expression meets deepfakes*, Synthese, volume 202, 7/2023, 1-17.
3. Birrer, A.; Just, N., *What We Know and Don't Know about Deepfakes: An Investigation into the State of the Research and Regulatory*, New Media and Society, volume 27, 12/2025, 6819-6838.
4. Carayannis, E. G.; Erokhin, V., *Democracy, Environment, and Technology: The Cases of AI and Fusion Energy* Springer Nature, Switzerland 2025.
5. Chesney, R.; Citron, D., *Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics*, Foreign Affairs, volume 98, 2/2019, 147-155.

¹²⁷ Correia de Carvalho, M., *It will be what we want it to be: Sociotechnical and Contested Systemic Risk at the Core of the EU's Regulation of Platforms' AI Systems*, Journal of Intellectual Property, Information Technology and Electronic Commerce Law, volume 16, 3/2025, p. 38.

¹²⁸ See: Bertelsmann Stiftung, *Decentralized Social Media Platforms as a Path to a More Resilient Information Ecosystem*, January 2026 <https://www.bertelsmann-stiftung.de>, 28 February 2026.

6. Clark, S.; Lewandowsky, S., *The Continued Influence of AI-Generated Deepfake Videos despite Transparency Warnings*, Communications Psychology, volume 4, 1/2026, 1-12.
7. Correia de Carvalho, M., *It will be what we want it to be: Sociotechnical and Contested Systemic Risk at the Core of the EU's Regulation of Platforms' AI Systems*, Journal of Intellectual Property, Information Technology and Electronic Commerce Law, volume 16, 3/2025, 35-72.
8. Dobber, Tom *et al.*, *Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes?*, The International Journal of Press/Politics, volume 26, 1/2021, 69-91.
9. Ebers, M., *Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act*, European Journal of Risk Regulation, volume 16, 9/2025, 684–703.
10. Fathaigh, R. *et.al.*, *Microtargeted propaganda by foreign actors: An interdisciplinary exploration*, Maastricht Journal of European and Comparative Law, volume 28, 11/ 2021, Vol. 28(6) 2021, 856- 877.
11. Gasser, U.; Palfrey, J., *Advanced introduction to Law and Digital Technologies*, Elgar Advanced Introductions, 2025.
12. Ginsborg, L.; Gori, P. *Disinformation: a multi-disciplinary analysis*, Springer 2026.
13. Hakkarainen, J.; Savolainen, L., *Individual choice, collective effects: recommender systems, law by design, and the DSA's double choice architecture*, Information, Communication and Society, volume 0, 12/2025, 1-18.
14. Kalpokas, I., *Problematising reality: the promises and perils of synthetic media*, SN Soc Sci, volume 1, 11/2020, 1-2.
15. Krack, N.; Dutkiewicz, L.; De Meyere, J., *Generative Artificial Intelligence and Disinformation*, in Ginsborg, L.; Gori, P. (eds), *Disinformation: A Multi-Disciplinary Analysis*, Springer Nature, Switzerland 2026.
16. Łabuz, M., *A Teleological Interpretation of the Definition of Deep Fakes in the EU Artificial Intelligence Act—A Purpose-Based Approach to Potential Problems With the Word "Existing"*, Policy and Internet, volume 17, 1/2025, 1-14.
17. Łabuz, M.; Nehring, C., *On the way to deep fake democracy? Deep fakes in election campaigns in 2023*, year 2024, 23:454–473 European Political Science, volume 23, 12/2024, 454–473
18. Milli, S. *et al.*, *Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media*, PNAS Nexus, volume 4, 3/2025, 1-10.
19. Millièrè, R., *Deep Learning and Synthetic Media*, Synthese, volume 200, 5/2022, 1-27.

20. Momeni, M., *Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation*, Journal of Creative Communications, volume 20, 10/2024, 41-56.
21. Nieweglowska, M.; Stellato, C.; Sloman, S. A., *Deepfakes: Vehicles for Radicalization, Not Persuasion*, Current Directions in Psychological Science, volume 32, 3/2023, 236-241
22. Pawelec, M., *Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions*, Digital Society, volume 19, 1/2022, 1-37.
23. Periello, L.E., *Blurred realities: Legal strategies for the deepfake era*, Maastricht Journal of European and Comparative Law, volume 0, 3/2026, 1-32.
24. Rauchfleisch, A.; Vogler, D.; de Seta, G., *Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions*, Social Media and Society, volume 11, 09/2025, 1-12.
25. Soundarya, B.; Gururaj, H. L., *Deepfake detection: critical review of state-of-the-art approaches and future perspectives*, Discover Applied Sciences, volume 8, 1/2026, 1-32.
26. Szabo, L.V.; Bader, S., *Deepfake content in political communication*, Convergence: The International Journal of Research into New Media Technologies, volume 0, 2026, 1-17.
27. Tolosana, R. et al., *Deepfakes and beyond: A Survey of face manipulation and fake detection*, Information Fusion, volume 64, 12/2020, 131-148.
28. Van der Sloot, B., *Regulating the Synthetic Society*, Hart Publishing, Oxford 2024
29. Van der Sloot, B.; Wagenveld, Y., *Deepfakes: regulatory challenges for the synthetic society*, Computer Law and Security Review, volume 46, 9/2022, 1-15.
30. Von Sikorski, Ch.; Hameleers, M., *Disinformation in the Age of Artificial Intelligence (AI): Implications for Journalism and Mass Communication*, Journalism & Mass Communication Quarterly, volume 102, 11/2025, 941–957,
31. Weikmann, T.; Egelhofer, J.; Lecheler, S., *Beyond Credibility: The Effects of Different Forms of Visual Disinformation*, Journalism & Mass Communication Quarterly, volume 102, 12/2025, 1020-1043.
32. Züger, T.; Asghari, H., *AI for the public. How public interest theory shifts the discourse on AI*, AI & SOCIETY volume 38, 4/2023, 815-828.

LEGAL RULES

1. Regulation (EU) 2024/1689 (Artificial Intelligence Act), OJ L 2024/1689.
2. Regulation (EU) 2022/2065 (Digital Services Act), OJ L 277/1.

NETWORK RESOURCES

1. Ašmane, K.L; McNamee, J., Malicious semi-compliance as the platform approach to EU regulation, EU DisinfoLab, 25 November, 2025, <https://www.disinfo.eu/malicious-semi-compliance-as-platform-approach-to-eu-regulation/>, 28 February, 2026
2. Berger, E., 'They feel true': political deepfakes are growing in influence – even if people know they aren't real, The Guardian, 28 March, 2026, <https://www.theguardian.com/technology/2026/mar/28/military-deepfakes-ai-propaganda-money>, 1 April 2026.
3. Bertelsmann Stiftung, Decentralized Social Media Platforms as a Path to a More Resilient Information Ecosystem, January 2026 <https://www.bertelsmann-stiftung.de>, 28 February 2026
4. Bleyer-Simon, K.; Reviglio, U., Defining Disinformation across EU and VLOP Policies, European Digital Media Observatory, October 2024, p.11, <https://edmo.eu/wp-content/uploads/2024/10/EDMO-Report-%E2%80%93-Defining-Disinformation-across-EU-and-VLOP-Policies.pdf>, 1 April, 2026.
5. Chesney, R.; Danielle, C., *Deepfakes: A Looming Crisis for National Security, Democracy and Privacy?* Lawfare, February 2018, <https://www.lawfaremedia.org/article/deepfakes-looming-crisis-national-security-democracy-and-privacy>, accessed 26 February, 2026.
6. DeepFaceLab, How DeepFaceLab Works, <https://deepfacelab.org/#choose>, 31 March 2026.
7. European Commission, Commission fines X €120 million under the Digital Services Act, European Commission Press Corner, 5 December 2025, https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2934, 1 April, 2026.
8. European Commission, The Code of Conduct on Disinformation, February 2026, p. 1, 2025 <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>, 26 February, 2026.
9. European Commission, First Draft Code of Practice on Transparency of AI-Generated Content, February 2026, p. 2, <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content> 1 March, 2026.

10. European Digital Rights (EDRi), Press Release: EU stands up to Big Tech with €120 million fine to X, 8 December 2025, EDRI Resources, <https://edri.org/our-work/eu-stands-up-to-big-tech-with-e120-million-fine-to-x/>, 1 April 2026.
11. EU Disinfo Lab, About, <https://www.disinfo.eu/about-us/problem/>, 28 February, 2026
12. European Parliament, Tackling AI deepfakes and sexual exploitation on social media (EP) Briefing, European Parliament, 15 January, 2026, <https://www.europarl.europa.eu/news/en/agenda/plenary-news/2026-01-19/8/tackling-ai-deepfakes-and-sexual-exploitation-on-social-media>, 15 February, 2026.
13. Gombar, M.; M. Boban, Research on the Impact of Algorithmic Echo Chambers on Perceptions and Attitudes of Social Network Users in a Digital Society, *MIPRO 48th ICT and Electronics Convention*, Opatija, Croatia, June 2025, <https://ieeexplore.ieee.org/document/11131918>, 1 April, 2026.
14. Kira, B., Deepfakes, The Weaponisation of AI Against Women and Possible Solutions, Verfassungsblog, June 2024, <https://verfassungsblog.de/deepfakes-ncid-ai-regulation/>, 12 February, 2026.
15. Körömi, C.; Haec, P; Cheslow, D, Zuck goes full Musk, dumps Facebook fact-checking program 7 January 2025, POLITICO, <https://www.politico.eu/article/mark-zuckerberg-full-elon-musk-dump-facebook-fact-checker/>, 20 February 2026.
16. Marco Silva, Flood of AI soldiers push pro-Iran messages during war, BBC News, 12 March 2026, <https://www.bbc.com/news/videos/cy9gw900dp9o>, 1 April, 2026.
17. Parker, J.; Völk, K., How deepfake porn scandal surrounding TV star rocked Germany, BBC News, 29 March, 2026, <https://www.bbc.com/news/articles/c9vlm4j47e2o>, 31 March, 2026.
18. Polona, C., Fact-checking and content moderation, European Parliamentary Research Service, February 2025, p. 2, [https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769493/EPRS_ATAG\(2025\)](https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/769493/EPRS_ATAG(2025)), 12 February, 2026.
19. Powers, C., Zuckerberg calls on US to defend tech companies against European 'censorship', Euroactiv, 12 January, 2025, <https://www.euractiv.com/news/zuckerberg-calls-on-us-to-defend-tech-companies-against-european-censorship/>, 26 February, 2026.
20. Shukla, V., Deepfakes and Elections: The Risk to Women's Political Participation, TechPolicy Press, February 2024, <https://www.techpolicy.press/deepfakes-and-elections-the-risk-to-womens-political-participation/>, 20 February, 2026.

21. Twomey, J. J.; Linehan, C.; Murphy, G., Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine, The Conversation, October 2023, 2023, <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393>, 31 March, 2026.