

Nicoleta Annemarie MUNTEANU, PhD 
Lucian Blaga University of Sibiu, Romania
nicoleta.munteanu@ulbsibiu.ro

TACKLING AI-GENERATED DISINFORMATION AND REGULATORY RESPONSES: EU–USA COMPARATIVE INSIGHTS

ABSTRACT: *The rapid advancement of generative artificial intelligence (AI) technologies has intensified concerns regarding the proliferation of disinformation, with significant implications for democratic processes, public trust, and international security. This article examines regulatory approaches to AI-generated disinformation in the European Union (EU) and the United States of America (USA), highlighting both convergences and divergences in policy frameworks, institutional responses, and normative orientations. Our research is contingent upon temporal proximity, with the data employed being dated November 2025. While the EU has pursued a proactive, rules-based strategy, most notably through the Digital Services Act and the AI Act, the United States has relied on a more decentralized, sectoral, and market-driven model, emphasizing innovation and self-regulation. By conducting a comparative analysis, the article identifies key strengths and limitations of each approach, particularly in balancing freedom of expression with the need to mitigate harmful content. Furthermore, it explores the extent to which transatlantic cooperation can foster more coherent responses to the global challenges posed by AI-driven disinformation. The findings suggest that although contextual differences shape regulatory choices, mutual learning and coordinated governance mechanisms are essential to effectively address this evolving phenomenon. Ultimately, the study underscores the importance of harmonizing ethical, legal, and technological strategies to safeguard democratic resilience in the digital age.*

KEYWORDS: European Union; AI-generated disinformation; regulatory frameworks; the United States of America; First Amendment

1. Introduction

The accelerated development of generative artificial intelligence (AI) technologies has expanded the frontiers of information production, while simultaneously intensifying concerns about the creation and dissemination of disinformation. The capacity of AI systems to generate highly realistic text, images, audio, and video presents both opportunities and threats for democratic societies. On the one hand, AI can facilitate innovation and enhance digital communication; on the other, it can amplify the spread of false or misleading content at unprecedented speed and scale. Against this backdrop, regulatory responses are emerging as a critical field of inquiry, particularly within the European Union (EU) and the United States of America (USA), two of the most influential global actors in shaping digital governance. The central research question of this study is how do the European Union and the United States regulate the risks associated with AI-generated disinformation, and how do their respective regulatory approaches differ with regard to regulatory architecture, allocation of responsibility, transparency, fundamental-rights safeguards, and enforcement? The study further

addresses three subsidiary research questions - to what extent do the EU and US frameworks address AI-generated disinformation directly, and to what extent do they regulate the broader digital-platform and AI ecosystem through measures that indirectly affect disinformation? How do differences in constitutional and institutional structures, particularly the EU's fundamental-rights and risk-based regulatory framework and the US constitutional protection of speech, shape the design and permissible scope of regulatory intervention? What mechanisms of EU–US cooperation exists for addressing AI-enabled foreign information manipulation and interference, and to what extent do these mechanisms provide a basis for regulatory interoperability?

Rather than testing causal hypotheses, the study develops three analytical propositions that guide the comparative inquiry: the EU and US regulatory approaches differ primarily in the distribution of regulatory authority and responsibility, with the EU relying more extensively on horizontal regulatory instruments and formal platform obligations, while the US framework is characterized by a combination of federal measures, agency action, judicial constraints, voluntary standards, and state-level legislation. The principal divergence between the two jurisdictions is not whether AI-generated disinformation constitutes a regulatory concern, but how that concern is reconciled with fundamental-rights protections, particularly freedom of expression in the United States and the broader fundamental-rights framework applicable within the European Union. Despite differences in regulatory architecture, the EU and US frameworks contain areas of functional convergence, particularly concerning transparency, platform accountability, AI safety, electoral integrity, and foreign information manipulation, creating potential grounds for transatlantic regulatory cooperation. These propositions are treated as analytical expectations rather than hypotheses subject to statistical testing. The purpose of the comparative analysis is to determine whether the documentary and legal evidence supports, qualifies, or contradicts these propositions.

While both actors recognize the risks associated with AI-enabled manipulation of information, their responses differ considerably. The EU has advanced a comprehensive and proactive model of regulation, exemplified by the Artificial Intelligence Act, which aims to classify AI systems according to risk levels and impose obligations on developers and deployers. Complementary initiatives, such as the Digital Services Act, reinforce accountability by requiring platforms to mitigate systemic risks, including the spread of disinformation. Conversely, the USA has pursued a more fragmented approach, prioritizing innovation and freedom of expression. The recent Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Federal Register, 2023) could have represented a step toward a more coordinated national strategy; it was issued on October 2023. According to our research topic, this document stated that irresponsible AI could exacerbate societal harms, including disinformation, but mostly focused on cybersecurity issues, which are undoubtable related to the informational environment. In this context, The Executive Order was considered to be a Congress's inaction consequence, due to the after 2020 election polarization and

also 2022 byelections (Pelc, P. 2024). This Executive Order was rescinded on January 2025 (White House, 2025). Even so, from our perspective, it would have remained less prescriptive than the EU framework. The methodology employed in this article is a comparative legal and policy analysis, combining document review of official regulations, executive measures, and judicial cases with scholarly commentary on AI governance.

Special attention is paid to the intersection between AI regulation and constitutional protections, particularly in the USA, where the First Amendment imposes strong safeguards for freedom of speech. Several court cases have begun to address issues at the nexus of AI technologies and free expression, illustrating both the opportunities and constraints for regulatory intervention in the American context. Arguments in favour of regulation emphasize the need to protect democratic institutions, safeguard electoral processes, and prevent societal harm caused by synthetic media. Critics, however, caution that overly restrictive regulation could stifle innovation, raise enforcement challenges, and risk infringing upon fundamental rights. This comparative inquiry also considers the prospects for transatlantic cooperation in addressing AI-generated disinformation.

2. Research Overview

The phenomenon of AI-generated disinformation has rapidly become a focal point of scholarly inquiry and policy debates. Researchers from multiple disciplines, such as law, political science, communication studies, and computer science have sought to conceptualize, assess, and mitigate the risks posed by generative AI technologies. A growing body of literature identifies AI-enabled disinformation as distinct from earlier forms of misinformation due to its scale, speed, and sophistication, with particular emphasis on the capacity of large language models and generative adversarial networks (GANs) to produce convincing synthetic media (Romanishyn et al., 2025). Scholars in media studies (Wardle, C., & Derakhshan, H., 2017), have highlighted the societal harms of disinformation ecosystems, while legal scholars have focused on the regulatory dilemmas that arise when balancing technological governance with fundamental rights (Balkin, 2018). In contemporary research and applications, artificial intelligence (AI) utilizes electromagnetic principles to process textual data with exceptional precision, often producing outputs that are virtually indistinguishable from those generated by humans. These systems, referred to as large language models (LLMs), have garnered significant attention due to their transformative potential in natural language processing and computational linguistics. The capabilities of LLMs not only demonstrate remarkable technical advancement but also raise important considerations regarding their impact on communication, information dissemination, and human-machine interaction (FLoridi, L 2023).

Practitioners, particularly within governmental and transnational institutions, have adopted complementary approaches. The European Union has taken a normative, rules-based stance, building upon its existing

regulatory architecture for digital platforms. The EU Code of Practice on Disinformation, followed by the Digital Services Act and the proposed Artificial Intelligence Act, reflects a layered strategy combining self-regulation, co-regulation, and legally binding obligations. We will include these aspects later in our study. Researchers analysing these measures (Helberger, N., Pierson, J., & Poell, T. 2018) emphasize their ambition to operationalize accountability and transparency standards across the digital ecosystem. In contrast, the United States has historically relied on a market-driven and innovation-oriented framework. Academic debates centre on the constraints imposed by the First Amendment, which complicates governmental regulation of disinformation, including when generated by AI. Practitioners in both jurisdictions have underscored the necessity of transatlantic cooperation. Policy research underscores that AI-generated disinformation circulates within a transnational information environment, where fragmented or uncoordinated regulatory responses risk undermining overall effectiveness in countering such threats. Scholars emphasize the need for coordinated governance frameworks to address the scale and cross-border nature of AI-driven disinformation campaigns (Gorwa, R., 2019; European Commission, 2023). Apart from the disrupting and divisive impact of disinformation campaigns in the political debate, influence operations impinge on the autonomy of citizens themselves” (Ignatidou, S., 2019).

Recent empirical studies offer promising evidence that targeted interventions can meaningfully curb the circulation of misinformation and disinformation. Pennycook, G., et al. (2021), using large-scale experimental designs, showed that brief accuracy-focused prompts can substantially enhance the reliability of information individuals choose to disseminate online. Their findings suggest that prompting users to reflect on the veracity of a headline prior to sharing significantly reduces the spread of false or misleading content. This indicates that misinformation often proliferates not through deliberate intent, but rather through momentary inattention to accuracy within the rapid, habitual dynamics of digital communication (Pennycook, C et.al., 2021).

Together, these academic and practical perspectives reveal both convergences, in recognizing the systemic risks of AI-generated disinformation, and divergence, rooted in legal traditions, institutional logics, and political cultures. From AI chatbots, to deepfake media, as elements parts of the AI-driven disinformation techniques, lead to the risk that they may substantially reinforce misleading narratives (Romanishyn, A. et. al., 2025). Our study reveals a reality that was identified at the beginning of 1996, by John Perry Barlow in the already well-known Declaration of the Independence of Cyberspace, noticing that the human will build a new society shaped by human thought within the digital world: ”we will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before” (Barlow, J. P., 1996).

3. Problem Statement

The rise of generative artificial intelligence has amplified long-standing challenges associated with misinformation and disinformation by enabling the creation of realistic fake news, persuasive synthetic texts, and hyper-realistic deepfakes. Unlike earlier waves of disinformation, AI-generated content can be produced at scale, personalized to target specific audiences, and disseminated rapidly across digital platforms, thereby threatening the integrity of democratic processes, public trust, and social cohesion.

Beyond its informational impact, AI-driven disinformation intersects with cybersecurity risks (Ummer, K., A., 2025), as malicious actors increasingly deploy generative models to craft phishing emails, social engineering campaigns, and malware that evade traditional detection systems. These threats highlight the dual-use nature of AI, where tools developed for innovation can be weaponized to undermine digital security infrastructures. Economically, AI-generated disinformation exacerbates concerns regarding job displacement and broader market disruptions, as automation and content manipulation technologies introduce new vulnerabilities for industries reliant on trust, authenticity, and reputation. Simultaneously, bias and discrimination embedded within AI systems risk perpetuating and amplifying existing social inequalities, particularly when deployed in politically sensitive or adversarial contexts. Furthermore, privacy violations present a critical challenge. The generation and dissemination of synthetic data raise profound questions regarding ownership, consent, and compliance with legal frameworks such as the EU's General Data Protection Regulation (GDPR). These multifaceted risks demand robust regulatory responses.

Thus, divergences between the EU's rules-based governance and the USA's market-driven, First Amendment constrained approach complicate the development of coherent global strategies. Addressing AI-generated disinformation therefore requires not only national regulation, but also coordinated transatlantic solutions. Joint EU–USA efforts, including dialogues under the Trade and Technology Council (TTC) - we will refer to this later in our study - signal an emerging platform for harmonizing approaches. Coordinated transatlantic governance can mitigate cross-border disinformation threats while setting global standards. EU and US cooperation in tackling disinformation should rest on a shared normative foundation, informed by mutually principles and a collective commitment to common democratic values (Ignatidou, S., 2019).

4. Methodology

This study employs a comparative qualitative research design, combining comparative legal analysis, qualitative document analysis, and structured policy analysis. The objective is to identify and explain similarities and differences between the European Union and the United States in their regulatory responses to AI-

generated disinformation, with particular attention to institutional architecture, regulatory instruments, platform accountability, transparency requirements, fundamental rights, and mechanisms of transatlantic cooperation. The study is temporally delimited to developments available up to 30 November 2025. This temporal boundary is important because the regulatory environment surrounding artificial intelligence is undergoing rapid change. Accordingly, the analysis distinguishes between instruments that had been adopted, instruments that had entered into application, and policy initiatives or legislative proposals that remained under development as of the research cut-off date. The empirical material consists primarily of primary legal and policy documents. For the European Union, the corpus includes the Digital Services Act (DSA), Regulation (EU) 2024/1689 establishing the Artificial Intelligence Act (AI Act), the EU Code of Practice on Disinformation and its subsequent integration into the DSA framework, European Commission communications and policy documents, relevant enforcement decisions and proceedings, and documents relating to EU–US cooperation. For the United States, the corpus includes federal legislation and executive measures, relevant agency frameworks and guidance, Supreme Court and lower-court decisions concerning digital speech and platform regulation, selected state legislation addressing AI-generated or manipulated political content, and documents concerning transatlantic cooperation.

Secondary sources, including peer-reviewed academic literature, institutional reports, legal commentary, and policy studies, are used to contextualise and interpret the primary materials. Secondary literature is not treated as equivalent to primary law; rather, it is used to identify competing interpretations, conceptual debates, and broader implications of the regulatory approaches examined.

The comparative analysis proceeds through a common analytical framework applied to both jurisdictions. Five dimensions are used as comparative criteria. The first one regulatory architecture presents the degree to which AI-generated disinformation is addressed through comprehensive horizontal regulation, sector-specific legislation, executive measures, voluntary frameworks, or combinations of these instruments; the second one - regulatory target and allocation of responsibility analysis whether obligations are primarily imposed on AI developers, deployers, online platforms, advertisers, political actors, intermediaries, or public authorities; the third one - transparency, traceability and content provenance provide the extent to which the regulatory framework requires disclosure, labelling, watermarking, provenance information, risk reporting, or access to relevant data; the fourth one - fundamental rights and safeguards explains the manner in which the framework addresses freedom of expression, privacy, equality, media freedom, electoral integrity, and procedural safeguards; the fifth one - enforcement and institutional capacity presents the authorities responsible for implementation and enforcement, available sanctions or remedies, supervisory mechanisms, and the degree of coordination among relevant institutions. A sixth dimension, transatlantic cooperation, is examined separately in order to assess the extent to which the two systems provide mechanisms for policy coordination, common

standards, information exchange, and joint responses to foreign information manipulation and interference.

The qualitative analysis follows a structured comparison rather than a simple chronological description. For each jurisdiction, the relevant legal and policy instruments are first examined individually and subsequently compared according to the common criteria. This enables the study to identify both institutional differences and functional similarities that may not be apparent from a description of individual regulatory instruments. The analysis does not attempt to measure the causal effectiveness of the EU or US regulatory systems in reducing the volume of disinformation. Such an assessment would require longitudinal behavioural and platform-level datasets that are beyond the scope of the present study. Instead, the article evaluates the regulatory design, scope, implementation mechanisms, rights safeguards, and institutional arrangements established to address AI-generated disinformation. Consequently, terms such as "effectiveness" are used in a limited regulatory sense, referring to the capacity of a framework to establish identifiable obligations, oversight mechanisms, transparency requirements, and mitigation mechanisms, rather than to a statistically demonstrated reduction in disinformation. This research design allows the study to answer the research questions through systematic comparison while avoiding an unsupported claim of quantitative measurement.

5. Preexisting Regulatory Approaches

This section of our research aims to present, in parallel, the recent instruments established at both the European Union and United States levels, designed to address the phenomena of misinformation, disinformation, and malinformation (MDM) (Masrek, M. N., et al., 2024).

a. European Union

In the case of the European Union, this effort began in 2015 with the project materialized within the East StratCom Task Force, established to respond to disinformation campaigns originating from the Russian Federation (European Council, 2015). In the spring of 2017, with the objective of curbing the spread of fake news, a conceptual framework was developed through the definition of the three terms, followed by a public consultation in the autumn of the same year (European Commission, 2018).

The chronology of events leading to the formulation of an EU strategy has been documented by author in previous research (Munteanu, N., A., 2019), updated according to the evolution of the European Union's strategy (Munteanu, N., A., 2022), with the aim of incorporating the most recent developments. The significant regulatory milestones generated between 2015 and present includes: 2016 "Joint Framework on Countering Hybrid Threats. A European Union Response (European Commission, 2016); 2017 "Joint Communication on

Resilience, Deterrence and Defence: Building strong cyber security in Europe (European Commission, 2017); 2018 “Joint Communication to the European Parliament, the European Council and the Council. Increasing resilience and bolstering capabilities to address hybrid threats (European Commission, 2018); 2018 “Code of Practice against Disinformation (European Commission, 2018); 2018 ”European External Action Service”; 2018 “Action Plan against Disinformation (European Commission, 2018); 2019 - Rapid Alert System (European Council 2020); 2019 “EU Reports on Progress in Fighting Disinformation Ahead of European Council” (European Commission, 2019); 2020 ”Digital Service Act” (European Commission, 2020); 2022 ”Code of Practice on Disinformation” (European Commission, 2022); 2024 ”Artificial Intelligence Act” (European Commission, 2024).

Following three years of intensive legislative negotiation, the Council of the European Union formally approved the European Union Artificial Intelligence (AI) Act on 21 May 2024, entered into force on August 2024. We mention that the rules will become applicable until August 2026. It was subsequently published in the EU Official Journal on 12 July 2024, marking it as the first comprehensive AI regulatory framework to complete the full EU legislative procedure. Originating from a proposal by the European Commission, the AI Act seeks to simultaneously promote the development and adoption of AI technologies while safeguarding fundamental rights and ensuring a human-centric orientation across the EU. The final legislation comprises approximately 50,000 words, organized into 180 recitals, 113 Articles, and 13 annexes, establishing a holistic, risk-based regulatory system that applies to all stakeholders within the AI ecosystem, including developers, exporters, and deployers (Lathan&Watkins, 2024). Given its expansive scope and regulatory ambition, the AI Act, much like the General Data Protection Regulation (GDPR) before it, is anticipated to exert influence well beyond EU borders and to serve as a benchmark for AI governance in the years ahead.

The year of 2025 is characterized by the gradual application of AI Act, mainly concerning the transparency matters and standardization of the detecting artificial content mechanisms. In the spring of 2025, the European Commission proposed an AI Continent Action Plan (European Commission, 2025), based on five strategic actions, to achieve competitiveness and productivity, and to assure the security, sovereignty and democracy. From our point of view this could be considered as one of the most important steps, including in terms of a possible comparison with the USA strategy.

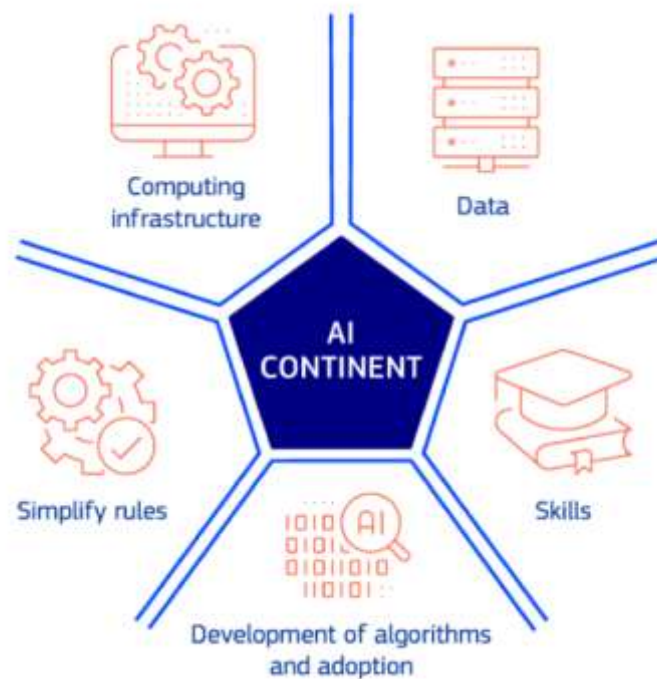


Fig. 1 - AI Continent Action Plan Strategic Areas (European Commission, 2025)

From a competition perspective, but also within the cooperation approach, the fact that the European Union provided a budget of 10 billion euro from 2021 to 2027 to train and finetune AI models, taking shape in at least 13 operational AI factories by 2026 represent not only a target, but mostly a necessity. Part of the computing infrastructure, European Union extend its ambitions to five AI gigafactories (European Commission, 2025) four times powerful than AI factories (20 billion euro allotted), in the context to encouraging the investments to triple the European Union's data centre till 2030-2032. Related to our study it is important to notice also the role of the algorithms and its connexions with the informational field, from an informatic perspective. The European Commission released on July 2025 the "General Purpose AI Code of Practice", entered into application on August 2025 (European Commission, 2025), representing a mechanism overseen by 14 independent experts, designed to support industry in meeting the requirements of the AI Act. The rules will start being enforced by the Commission's AI Office one year after adoption, in 2026 for new models and two years after adoption, 2027 for models already on the market. The goal is to ensure that all general-purpose AI models available in the EU, especially the most advanced ones, meet strong safety and transparency standards. This month, November 2025, the European Commission started the project, under the AI Act, aimed to create a code of practice on marking and labelling AI-generated content, based on a public consultation started in September 2025, for a month. This initiative's purposes are to facilitate the manage the AI-generated outputs,

including synthetic, video, visual media, content, in formats that can be readily interpreted by machines, enhancing the detection and traceability. The Code will also provide guidance to entities deploying deepfakes or other AI-generated materials, to ensure transparent disclosure of AI involvement, especially in communications related to the matters of public significance. The requirements are scheduled to take effect in August 2026 and complete the already existing regulatory frameworks (European Commission, 2025). The European Union has established itself as a frontrunner in advancing the implementation of watermarking techniques for generative AI models (Souveraion, T., 2025).

As part of the research methodology we have proposed, we considered it appropriate to present, by way of illustration Cases that involved social media platforms and the implications of digital speech. Within the European Union jurisprudence there are cases and regulatory enforcement actions concerning social-media platforms that implicate questions of digital speech, particularly where content moderation practices rely on algorithmic or automated systems. In the 2024 case of European Commission against Meta as provider for Facebook and Instagram, frameworks (European Commission, 2025), the platform is under the suspicion of breaching the Digital Services Act. In this context we mention that since 2023 Facebook and Instagram are part of the VLOPs (Very Large Online Platforms), under the rule of complying with the obligations set out in the DSA (European Commission, 2023). The main elements of the ongoing proceedings is focused on disinformation and deceptive advertisements, visibility of political content, the mechanism to flag illegal content.

Another case significant for our research is *Meta Platforms Ireland v Commission/TikTok Technology v Commission* (Court of Justice of the European Union, 2025). Under the Digital Services Act (DSA), the European Commission is assigned specific supervisory responsibilities over certain service providers that qualify as very large online platforms or very large online search engines, a status triggered once their user base in the European Union surpasses a defined quantitative threshold. To finance the execution of these regulatory functions, the Commission imposes an annual supervisory levy on each designated provider. The amount payable is calculated in proportion to the platforms or service's average number of monthly users, ensuring that the fee reflects the scale of the service subject to oversight. The general Court annuls the decisions of the Commission related to the supervisory fee applicable on Instagram, TikTok and Facebook. In 2025 a Berlin Court decided in favor of two German NGOs, in a lawsuit against X. The social media was accused that violates the EU's Digital Service Act, concrete the aspect that compels platform to provide, without undue delay, public data to the researches (Desmarais, A., 2025). The DSA obliges platforms designated as "very large" to grant qualified non-profit entities, organisations, and research associations access to their internal data when such access is necessary to conduct studies aimed at detecting, identifying, and analysing systemic risks arising on social media.

The case is also important from a pecuniary point of view, because although X offers only a minimal level of free API (Application Programming Interface) access, anyone requiring substantial data volumes must purchase a paid tier, its “Pro access”, which can reach as high as \$5,000 (€4,794) per month. According to those two NGOs representative, this pricing makes meaningful data access effectively unattainable. It is particularly noteworthy, within the framework of this case, that The European Commission’s preliminary inquiry into X reached a similar conclusion. In its 2024 press statement, the Commission observed that the platform seemingly discourages researchers from pursuing their investigative work or effectively compels them to incur excessively high costs in order to do so (Pinge, A., 2024).

In this regard, it is deemed appropriate to mention few of the European Union actions: in 2024 The European Commission has rendered TikTok’s undertakings to permanently discontinue the TikTok Lite Rewards programme within the European Union legally binding. These commitments, submitted by TikTok in response to the concerns identified by the Commission are intended to ensure the platform’s compliance with the obligations set forth under the DSA. This decision represents the first instance in which the European Commission has concluded proceedings under the article 71 of the DSA, doing so 105 days subsequent to their initiation. It also constitutes the first occasion on which the Commission has accepted commitments from a designated online platform in the context of formal proceedings initiated under the DSA (European Commission, 2024). In 2024 the European Commission has informed X that its activity is against the DSA, in terms of advertising transparency, data access for researchers, dark patterns. The purpose was to ensure that all the platforms, including X comply with EU legislation (Vestager, M., 2024).

b. The United States of America

The USA implemented several legislative measures to combat disinformation and misinformation, such as National Artificial Intelligence Initiative Act of 2020 (U.S. Congress, 2020); Educating against Misinformation and Disinformation Act 2021-2022 (U.S. Congress, 2022); Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (White House, October 2023); Legislation against Deepfakes (Federal Trade Commission, February 2024); State Department’s Framework to Counter Foreign State Information Manipulation, (U.S. Department of State, January 2024). Since the beginning of 2025 the situation related to AI is in a continuous moving (White House, 2025).

Several U.S. states have enacted legislation addressing deepfakes and artificial intelligence. For example, since 2019 Texas and California (Brennan Centre for Justice, 2023) passed laws prohibiting malicious deepfakes in elections, while Virginia, in the same year, criminalized deepfake pornography. More recent bills, such as New York’s digital impersonation laws, expand protections against AI-enabled identity misuse.

Wisconsin requires disclaimers on AI-generated political content. Arizona has introduced measures that allow civil actions against the creators of AI-generated disinformation targeting political candidates or officials (Voting Rights Lab). Florida imposes criminal penalties for failing to include disclaimers on such content (Brennan Centre for Justice, 2024).

i. AI and the First Amendment

The First Amendment of the United States Constitution safeguards essential civil liberties, encompassing freedom of speech, freedom of the press, the right to assemble, the right to petition the government for redress, and the protection of religious practice. These constitutional protections are particularly significant in contemporary debates surrounding the dissemination of disinformation and misinformation, as they frame the legal and ethical boundaries for regulating speech while maintaining democratic principles. Two key ideas help us understand when the First Amendment permits legal regulation of the people and organizations that use Big Data, algorithms, and artificial intelligence. The first is the concept of information fiduciaries. The second is the concept of algorithmic nuisance (Balkin, J. M., 2018). As AI becomes more integrated into society, its relationship with the First Amendment raises numerous legal and ethical questions:

Freedom of speech AI as a content generator. A central question in the study of AI-generated content is whether such speech is afforded protection under the First Amendment. This category encompasses a broad range of expressions created by AI, including scholarly articles, social media posts, and other digital content (Balkin, J. M., 2020; Klonick, K., 2022). The rapid advancement of generative AI models, such as large language models and multimodal systems, raises novel legal and ethical challenges, as courts have not yet definitively established whether AI-generated content enjoys the same constitutional safeguards as human speech (Citron, D. K., & Norton, H. L., 2021). Scholars argue that the distinction between human and machine-generated expression may have profound implications for content regulation, freedom of expression, and accountability in digital spaces (Gasser, U., & Almeida, V. A. F., 2017). Until judicial precedent emerges, policymakers and legal practitioners must navigate the uncertainties surrounding AI-generated content, balancing innovation with constitutional protections.

In terms of *regulation and censorship*, the platforms employing AI for content moderation face the challenge of balancing the removal of harmful or illegal content with the protection of users' freedom of speech. While AI systems can efficiently detect and eliminate content such as misinformation and hate speech, they are not infallible. These systems may misinterpret context, leading to over-censorship or the suppression of legitimate expression (Hakami A., & Tazel, R., 2024). Additionally, biases inherent in AI algorithms can disproportionately affect marginalized communities, undermining democratic values (Balendra, S., 2025). Therefore, it is crucial for platforms to implement transparent, accountable, and equitable moderation practices that uphold both safety and free expression (Kamande, J., 2024).

Freedom of Press. AI in Journalism Artificial intelligence is increasingly integrated into journalism, facilitating tasks such as data analysis, content creation, and distribution. While AI tools can enhance journalistic efficiency and innovation, their use raises complex legal questions, particularly concerning First Amendment protections. Courts have established that computer code constitutes speech under the First Amendment, thereby extending protections to AI-generated content (Foundation for Individual Rights in Education, 2024). However, AI itself lacks constitutional rights; thus, the speech rights pertain to the human creators or users of the AI tools (Balkin, J. M., 2020). This distinction is crucial as it influences the application of speech regulations and the assignment of liability for AI-generated content (Kazaryan, A., 2025). As AI continues to shape journalism, legal frameworks must adapt to balance innovation with the preservation of free speech principles.

Accountability. The integration of artificial intelligence into journalism raises significant concerns regarding accountability for errors and biases in AI-generated content. AI systems, lacking moral agency, can produce content that is inaccurate or biased, yet attributing responsibility is complex. Studies indicate that AI-generated news often exhibits gender and racial biases, with some models demonstrating higher levels of discrimination against women and Black individuals (Fang, X., et al., 2024). The absence of clear accountability frameworks exacerbates these issues, underscoring the need for transparent and responsible AI usage in journalism (Morosoli, S., et al., 2025).

Bias and Discrimination. Algorithmic bias in artificial intelligence systems poses significant challenges to the First Amendment, particularly concerning free speech and equal protection. AI models, trained on vast datasets, can inadvertently perpetuate racial, gender, and ideological biases, leading to discriminatory outcomes in areas like hiring, law enforcement, and media representation. While the First Amendment protects individuals' rights to free speech, it does not extend these protections to AI systems themselves (Salib, P., 2024). This distinction is crucial, as it allows for regulation of AI outputs to prevent harm without infringing on constitutional rights. However, the challenge lies in balancing regulation with the preservation of free expression, ensuring that efforts to mitigate bias do not inadvertently suppress legitimate speech (Kazaryan, A., 2025). As AI continues to influence public discourse, ongoing legal and ethical considerations are essential to uphold both non-discrimination and free-speech principles.

Private Companies. Social media platforms and other private technology companies operate under different legal constraints than government actors, as the First Amendment restricts only governmental suppression of speech (Cornell Law, 2023). These entities possess broad authority to develop and enforce content moderation policies, including the management of disinformation and misinformation, often leveraging artificial intelligence tools to automate detection and removal processes (Gorwa R. et al., 2020). AI-based moderation must navigate the tension between minimizing harm and preserving free expression, highlighting the ethical and technical challenges of algorithmic governance (Napoli, 2019). The cooperation between U.S.

authorities and major tech platforms such as Google, Microsoft, Twitter, Facebook, Tik Tok, Instagram, related to AI exists in several areas including ethical standards, research, security and regulation: Ethical AI Guidelines, under the umbrella of U.S: National Institute of Standards and Technology (NIST).

From the perspective of the legal landscape, the intersection of the First Amendment and AI is an evolving area of law, including court cases and legislation and regulation. Related to court cases, there have been court cases directly addressing AI and the First Amendment, but future litigation is likely as AI continues to advance. *NetChoice v. Paxton* (2024) - this case involves the challenge against Texas' HB 20, a law that restricts social media platforms from moderating content based on viewpoints. AI plays a crucial role in content moderation, and the outcome of this case could impact how companies use AI to manage speech on their platforms (Cornell's Legal Information Institute, 2024). *Gonzalez v. Google LLC* (2023) - this Supreme Court case examines Section 230 of the Communications Decency Act, which provides immunity to online platforms for user-generated content. The role of AI in recommending and moderating content will be a critical aspect of this case (Supreme Court of the United States, 2023). *Knight First Amendment Institute v. Trump* (2019) - while not directly about AI, this case involved the use of social media algorithms. The court ruled that President Trump's blocking of users on Twitter violated the First Amendment. The implications of this ruling extend to how AI and algorithms used by social media platforms can impact free-speech rights (Supreme Court of the United States, 2019). *White Coat Waste Project v. Greater Richmond Transit Company* (2020). This case involved the use of AI to display advertisements on public buses. The Greater Richmond Transit Company (GRTC) rejected an advertisement from the White Coat Waste Project, leading to a lawsuit alleging First Amendment violations. The court ruled that the GRTC's advertising policy was too vague and constituted viewpoint discrimination, thus violating the First Amendment. This case underscores how AI can be used in public communications and the potential for First Amendment challenges (United States Court of Appeals for the Fourth Circuit, 2022). *Fields v. Twitter* (2018) - this case involved claims that Twitter's algorithms (a form of AI) aided terrorists by allowing their messages to be disseminated, raising questions about the liability of AI systems in facilitating harmful speech (United States Court of Appeals for the Ninth Circuit, 2018). *Hunt v. Board of Regents of the University of New Mexico* (2018) - Paul Hunt, a medical student, was suspended for a Facebook post criticizing the use of taxpayer funds for abortions. Hunt sued, claiming the university violated his First Amendment rights (United States District Court for the District of New Mexico, 2018).

Although AI was not the central issue, the case involved social media platforms and the implications of digital speech, which are increasingly moderated by AI algorithms. *Carpenter v. United States* (2018). This case addressed the use of cell-site location information (CSLI) by law enforcement without a warrant. While not directly about AI, it raised significant questions about privacy and the use of technology in surveillance. The Supreme Court ruled that accessing historical CSLI constitutes a search under the Fourth Amendment,

thus requiring a warrant. The decision has implications for AI technologies used in surveillance and data collection. These cases illustrate the complex and evolving legal landscape at the intersection of AI and First Amendment rights. As AI technology continues to develop, we can expect more legal challenges and decisions in this area.

Recent legal developments have begun to address the intersection of artificial intelligence (AI) and the First Amendment, particularly concerning the protection of AI-generated speech. In *Garcia v. Character Technologies*, a federal court declined to dismiss a lawsuit involving an AI chatbot, expressing scepticism about whether the chatbot's outputs qualify as "speech" under the First Amendment (United State District Court for the Middle District of Florida, 2025). Conversely, in *Walters v. OpenAI*, a Georgia state court suggested that AI-generated content might be entitled to First Amendment protections (United States Court of Appeals for the Eleventh Circuit, 2024; National Constitution Center, 2025). These conflicting rulings highlight the evolving nature of jurisprudence in this area. Legal scholars debate whether AI outputs should be considered speech, with some arguing that extending First Amendment rights to AI could hinder effective regulation (Atkinson et al. 2025). As AI continues to influence public discourse, courts will likely play a pivotal role in defining the scope of constitutional protections for AI-generated content. Related to legislation and regulation, Law-makers and regulators are beginning to address AI-related issues, but there is no comprehensive legal framework yet. This includes efforts to address algorithmic transparency and accountability.

6. Arguments in Favor of and Against Regulation

Considering the importance of objectivity in the research approach, we have analysed both the advantages and disadvantages associated with regulating AI-generated disinformation. Our analysis indicates that proponents of implementing a regulatory framework primarily emphasize four key elements: combating disinformation, safeguarding democracy, strengthening national security, and protecting public health from an informational perspective.

Regulating AI-generated misinformation can enhance societal resilience by mitigating the rapid spread of false content and protecting democratic processes (Helberger, N., Pierson, J., & Poell, T., 2018). It supports national security by reducing vulnerabilities to manipulation campaigns and safeguards public health through the containment of harmful disinformation (Floridi, L., & Chiriatti, M., 2020). Moreover, a clear regulatory framework fosters accountability among technology providers and promotes transparency in algorithmic content generation (European Commission, 2021).

Conversely, arguments against regulation centre on concerns related to freedom of expression, enforcement challenges, the risk of stifling innovation, and global regulatory discrepancies. We argue that these two

perspectives, advanced by both scholars and practitioners, should be addressed through four management criteria: transparency requirements, collaboration with technology companies, educational campaigns, and accountability measures. Regulating AI-generated misinformation may risk limiting freedom of expression, particularly in jurisdictions with strong speech protections like the United States (Balkin, J. M., 2020). Enforcement challenges arise due to the global, rapidly evolving nature of AI technologies, potentially leading to inconsistent application (Gasser, U., & Almeida, V. A. F., 2017). Overly restrictive regulations may also stifle innovation and hinder technological development, while international discrepancies complicate cross-border governance (Wagner, B., 2021).

7. The Eu and the USA Cooperation to Combat AI-Generated Disinformation

The EU–US Trade and Technology Council (TTC) has become a central platform for transatlantic cooperation in addressing AI-generated disinformation. Established in 2021, the TTC aims to align digital policies between the EU and the US, focusing on shared democratic values and human rights (European Commission, 2021). A key initiative under the TTC is the Joint Roadmap for Trustworthy AI, which emphasizes the development and deployment of AI technologies that are transparent, accountable, and respect fundamental rights (Centre for AI and Digital Policy, 2022). This roadmap serves as a foundational framework for both regions to harmonize their approaches to AI governance, ensuring that AI systems do not perpetuate or amplify disinformation.

Furthermore, the TTC has facilitated collaboration on combating foreign information manipulation and interference (FIMI), recognizing the threats posed by AI-driven disinformation campaigns to democratic processes (Office of the United States Trade Representative, 2023). By coordinating efforts, the EU and the US aim to enhance resilience against such threats, leveraging shared resources and expertise. Through these concerted efforts, the EU and the US demonstrate a commitment to not only advancing AI technologies but also ensuring they are developed and utilized in ways that uphold the integrity of democratic institutions and the dissemination of truthful information.

The European Union (EU) and the United States have increasingly recognized the need for transatlantic cooperation to combat AI-generated disinformation, particularly in the context of elections and democratic integrity. In 2023, both parties expressed deep concern over foreign information manipulation and interference (FIMI), emphasizing the importance of joint efforts to address disinformation challenges. This collaboration has been further institutionalized through the signing of the Framework Convention on Artificial Intelligence by the EU, US, and UK in September 2024 (Council of Europe Portal, 2024). This legally binding treaty aims

to align AI development with human rights, democracy, and the rule of law, setting forth principles for transparency and accountability in AI systems. By creating legally enforceable commitments, the treaty aims to guide AI development in accordance with human rights, democratic norms, and the rule of law, emphasizing principles of transparency, accountability, and ethical governance.

Moreover, the EU's Digital Services Act (DSA) and the US's AI initiatives have prompted discussions on creating a unified AI code of conduct. In June 2023, lawmakers from both regions agreed to draft such a code, focusing on mitigating AI-driven disinformation risks (Drake et al., 2023). These efforts are complemented by collaborative research projects under Horizon Europe, such as AI4Media, AI4Trust, TITAN, and vera.ai, which aim to develop tools and strategies to counter sophisticated AI-generated disinformation (EU Disinfo Lab, 2025; Munteanu N. A., 2022)

Despite these advancements, challenges persist. The closure of the US Global Engagement Centre in 2025, due to domestic concerns over censorship, has raised questions about the sustainability of joint disinformation initiatives (Congressional Research Service, 2024). Nevertheless, the EU and US continue to explore avenues for collaboration, recognizing that a coordinated approach is essential to effectively address the evolving threat of AI-generated disinformation.

8. Conclusions

The EU has introduced the Digital Services Act and the Artificial Intelligence Act, which aim to regulate digital platforms and AI technologies to prevent the spread of disinformation. The USA, through agencies like the Federal Trade Commission (FTC) and legislative initiatives, is working on policies to address the ethical use of AI and combat disinformation. EU and U.S. policymakers are aligned around two core themes for AI policy, enabling innovation and competition, ensuring trust and accountability.

During our research we noticed few *differences*; EU tends to focus on risk management and trust. U.S. tends to focus on the innovation and growth, greater research and development funding, and light-touch regulation. The Trade and Technology Council could play a role by exploring to what extent these approaches can be aligned behind a US-EU effort to enable safe and responsible AI innovation and adoption globally. Tackling AI-generated disinformation is a critical issue for both the European Union and the United States, as it poses significant risks to democratic processes, public health, and social cohesion. Here are some *strategies and initiatives* being considered or implemented by both regions to address this challenge: the potential threats posed by AI-generated content to the European Union (EU) and the United States (USA) can be understood from several perspectives; *misinformation and disinformation* related to the fact that AI-generated content can be used to create realistic fake news, deepfakes, and other forms of misinformation, which can influence public

opinion, interfere with elections, and create social unrest. This is a significant concern for both the EU and the USA. *Cybersecurity*: AI-generated phishing emails, malware, and other cyber threats can be more convincing and harder to detect, posing risks to individuals, businesses, and government institutions. *Economic Disruption*: Automation and AI-driven technologies can displace jobs, leading to economic and social challenges. While AI can create new opportunities, it also requires a shift in workforce skills and training. *Privacy*: AI systems that process and analyse large amounts of data can lead to privacy violations. This is a major concern, especially in the EU, which has strict data protection regulations under the General Data Protection Regulation (GDPR). *Bias and Discrimination*: AI systems can perpetuate and amplify existing biases if not properly managed. This can lead to unfair treatment and discrimination in areas such as hiring, lending, and law enforcement.

In response to these threats, both the EU and the USA are actively working on regulatory frameworks to ensure the safe and ethical use of AI. The EU has proposed the AI Act, which aims to regulate high-risk AI applications and promote trustworthy AI. The USA is also considering various legislative measures and has established initiatives to ensure AI development aligns with ethical guidelines and national security interests. Overall, while AI-generated content and applications present potential threats, they also offer significant opportunities for innovation and improvement in various sectors. The key lies in managing and mitigating the risks through robust regulation, ethical guidelines, and continuous oversight.

We consider in our study few *common challenges and solutions*, such as detection and verification, developing sophisticated AI and machine learning algorithms capable of identifying AI-generated disinformation, strengthening the role of independent fact-checkers and providing them with better tools and resources, platform accountability, by ensuring that social media platforms and tech companies take proactive measures to prevent the spread of disinformation; implementing transparent policies for content moderation and providing clear guidelines for users.

Our research had an important content in terms of the legal and ethical considerations. We consider that balancing the need to combat disinformation with the protection of free speech and privacy rights. Establishing clear and fair legal frameworks that address the complexities of AI-generated content. By addressing these issues through a combination of regulation, technology, public awareness, and international cooperation, the EU and USA can better tackle the growing threat of AI-generated disinformation.

Despite their regulatory differences, the EU and the USA share strategic interests in countering threats posed by malign actors and foreign interference campaigns. Joint initiatives, dialogues, and working groups suggest a growing recognition that only coordinated approaches can effectively mitigate the cross-border nature of AI-driven disinformation. By analysing the convergences and divergences between the EU and USA

approaches, this article seeks to contribute to a deeper understanding of how democratic societies can balance innovation, freedom of expression, and security in the age of generative AI.

References

1. Atkinson, D.; Hwang, D. J.; Morrison, J., (2025). Intentionally Unintentional: GenAI Exceptionalism and the First Amendment. Cornell University. Computers and Society. <https://doi.org/10.48550/arXiv.2506.05211>.
2. Balendra, S. (2025). Meta's AI moderation and free speech: Ongoing challenges in the Global South. *Cambridge Forum on AI: Law and Governance*, No. 21, pp. 1–19 <https://doi:10.1017/cfl.2025.5>.
3. Balkin J. M. (2018). Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. Available at: https://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/51-3_Balkin.pdf.
4. Balkin, J. M. (2024). AI and the First Amendment. Interview for Yale Law School. Available at: <https://law.yale.edu/yls-today/news/ai-and-first-amendment-qa-jack-balkin>.
5. Balkin, J. M. (2020). *Free Speech in the Algorithmic Society*. *UC Davis Law Review*, No. 51, pp.1149–1210.
6. Barlow, J.P. (1996). A Declaration of the Independence of Cyberspace. Available at: <https://www.eff.org/cyberspace-independence>.
7. Barman, D. (2024). The dark side of language models: Generating multi-modal disinformation. *Journal of AI and Society*, 39 (4), pp. 567–580. Available at: <https://www.sciencedirect.com/science/article/pii/S2666827024000215>.
8. Bomboy, S. (2025). Lawsuit analyzes First Amendment protection for AI chatbots in civil case. Available at: <https://playwithchatgtp.com/lawsuit-analyzes-first-amendment-protection-for-ai-chatbots-in-civil-case-the-national-constitution-center/>.
9. Bontcheva, K. (2023). Generative AI and disinformation: Recent advances and challenges. *EDMO White Paper*. Available at: https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf.
10. Brennan Centre for Justice. (2023). Regulating AI Deepfakes and Synthetic Media in the Political Arena, Available at: <https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena>.
11. Brennan Centre for Justice. (2024). States Take the Lead in Regulating AI in Elections – Within Limits, Available at: <https://www.brennancenter.org/our-work/research-reports/states-take-lead-regulating-ai-elections-within-limits>.
12. Centre for AI and Digital Policy. (2022). EU-US Trade and Technology Council. TTC and Artificial Intelligence Policy. Available at: <https://www.caidp.org/resources/trade-and-technology-council/?utm>.
13. Citron, D. K., & Norton, H. L. (2021). Law and Digital Speech in the Age of AI. *Yale Law Journal Forum*, 130, pp. 1–28.
14. Congress.gov. (2020). H.R. 6216-National Artificial Intelligence Initiative Act of 2020 Available at: <https://www.congress.gov/bill/116th-congress/house-bill/6216>.
15. Congress.gov. (2022). Educating Against Misinformation and Disinformation Act. Available at: <https://www.congress.gov/bill/117th-congress/house-bill/6971?q=%7B%22search%22%3A%22Education+against+Misinformation+and+Disinformation+Act+2021-2022+%28H.R.+6971%2C+117th+Congress%29%22%7D&s=1&r=1>.
16. Congressional Research Service. (2024). Termination of the State Department's Global Engagement Center. Available at: <https://www.congress.gov/crs-product/IN12475?utm>.
17. Cornell's Legal Information Institute (2024), *NetChoice, LLC v. Paxton*. Available at: <https://www.oyez.org/cases/2023/22-555?utm>.

18. Council of Europe Portal, 2024. The Framework Convention on Artificial Intelligence. Available at: <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence?utm> .
19. Court of Justice of the European Union. (2025). Judgments of the General Court in Case T-55/24. Meta Platforms Ireland v Commission and T-58/24. TikTok Technology v Commission. Available at: <https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-09/cp250114en.pdf>.
20. Desmarais, A. (2025). A Berlin court ruled in favour of two German nonprofits, who sued X for breaching EU law by not giving them access to data to study how posts on the platform could influence the upcoming election. Available at: https://www.euronews.com/next/2025/02/07/we-must-defend-ourselves-german-non-profits-sue-x-for-access-to-public-data?utm_source=chatgpt.com.
21. Drake, M., Hansen, M., & Peets, L. (2023). EU and US Lawmakers Agree to Draft AI Code of Conduct. *Inside Privacy*. Available at: <https://www.insideprivacy.com/artificial-intelligence/eu-and-us-lawmakers-agree-to-draft-ai-code-of-conduct/?utm>.
22. European Commission (2018). The Action Plan against Disinformation. Available at: https://ec.europa.eu/commission/sites/beta-political/files/factsheet_disinfo_elex_140619_final.pdf.
23. European Commission, EU Code of Practice on Disinformation, Brussels, 2018.
24. European Commission, Joint Communication to the European Parliament and the Council, Joint Framework on Countering Hybrid Threats a European Union Response. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016JC0018&from=en>.
25. European Commission. (2017). Joint Communication to the European Parliament and the Council. Resilience, Deterrence and Defence: Building strong cyber security in Europe.
26. European Commission. (2018). Commission Recommendation of 12.02.2018 on enhancing the European nature and efficient conduct of the 2019 elections to the European Parliament, pp. 5-6. Available on: https://ec.europa.eu/commission/sites/betapolitical/files/recommendation-enhancingeuropean-nature-efficient-conduct-2019-elections_en.pdf .
27. European Commission. (2019). A Europe that protects: EU reports on progress in fight ahead of European Council, Brussels, 2019.
28. European Commission. (2020). Joint Communication to the European Parliament and the Council, Joint Framework on countering hybrid threats a European Union response. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016JC0018&from=en>.
29. European Commission. (2021). EU-US Trade and Technology Council (2021-2024). Available at: <https://digital-strategy.ec.europa.eu/en/node/12599/printable/pdf> .
30. European Commission. (2021). Proposal for a Regulation on Artificial Intelligence (AI Act).
31. European Commission. (2022). A strengthened EU Code of Practice on Disinformation. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/protecting-democracy/strengthened-eu-code-practice-disinformation_en.
32. European Commission. (2022). Digital Service Act. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en.
33. European Commission. (2023). DSA: Very large online platforms and search engines. Available at: <https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops>.
34. European Commission. (2023). *Tackling online disinformation: A European approach*. Brussels: European Commission. Available at: <https://digital-strategy.ec.europa.eu/en/policies/online-disinformation>.
- a. European Commission. (2024). TikTok commits to permanently withdraw TikTok Lite Rewards programme from the EU to comply with the Digital Services Act. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4161.
35. European Commission. (2024). AI Act enters into force. Available at: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en.
36. European Commission. (2024). Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act, Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373?utm.

37. European Commission. (2025). AI Continent Action Plan. Available at: <https://digital-strategy.ec.europa.eu/en/factpages/ai-continent-action-plan>.
38. European Commission. (2025). Commission launches work on a code of practice on marking and labelling AI-generated content, 05 November 2025, Available at: <https://digital-strategy.ec.europa.eu/en/news/commission-launches-work-code-practice-marking-and-labelling-ai-generated-content>.
39. European Commission. (2017). Joint Communication to the European Parliament and the Council. Resilience, Deterrence and Defence: Building Strong Cyber Security in Europe.
40. European Commission. (2018). Joint Communication to the European Parliament, the European Council and the Council, increasing resilience and bolstering capabilities to address hybrid threats.
41. European Council. (2015). European Council Meeting, Conclusions, 19th and 20th of March, 2015, p. 15. Available at: <https://www.consilium.europa.eu/media/21888/european-council-conclusions-19-20-march-2015-en.pdf>.
42. European Council. (2019). Rapid Alert System, Strengthening coordinated and joint responses to disinformation. Available at: https://eeas.europa.eu/sites/eeas/files/ras_factsheet_march_2019_0.pdf.
43. Ey&Disinfo. (2020). Rapid Alert System, Strengthening coordinated and joint responses to disinformation, Available at: https://eeas.europa.eu/sites/eeas/files/ras_factsheet_march_2019_0.pdf.
44. Fang, X., Che S., Mao, M., Zhang H.& Zhao M. (2024). Bias of AI-Generated Content: An Examination of New Produced by Large Language Models. Cornell University, pp. 11-14, <https://doi.org/10.48550/arXiv.2309.09825>.
45. Federal Register. (2023). Executive Office of the President. Available on: <https://www.federalregister.gov/agencies/executive-office-of-the-president>.
46. Federal Register. (2023). Executive Order 14110—Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence October 30, 2023. Available at: <https://www.govinfo.gov/content/pkg/DCPD-202300949/pdf/DCPD-202300949.pdf>.
47. Floridi, L. (2023). AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models, *Philosophy & Technology*, 36:15. <https://doi.org/10.1007/s13347-023-00621-y>.
48. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30, pp. 681–694.
49. Foundation for Individual Rights in Education. (2024). Artificial intelligence, free speech, and the First Amendment. Available at: <https://www.thefire.org/research-learn/artificial-intelligence-free-speech-and-first-amendment>.
50. Foundation for Individual Rights and Expression. (2024). Artificial intelligence, free speech, and the First Amendment. Available at: <https://www.thefire.org/research-learn/artificial-intelligence-free-speech-and-first-amendment>.
51. Gasser, U., & Almeida, V. A. F. (2017). A Layered Model for AI Governance. *IEEE Internet Computing*, 21(6), pp. 58–62.
52. Gorwa, R. (2019). The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1404>.
53. Hakami, A.& Tazel, R. (2024). The Ethics of AI in Content Moderation: Balancing Privacy, Free Speech, and Algorithmic Control. DOI:10.13140/RG.2.2.19529.97121.
54. Helberger, N.; Pierson, J. & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility, *The Information Society*, Routledge Taylor&Francis Group Vol. 34, No. 1, pp. 1–14. <https://doi.org/10.1080/01972243.2017.1391913>.
55. Ignatidou, Sophia. (2019). EU–US Cooperation on Tackling Disinformation, Chatham House. The Royal Institute of International Affairs, p. 34.
56. Kamande, J. (2024). Redefining Free Speech: The Impact of AI-Driven Content Moderation on Privacy and Expression. <https://doi.org/10.13140/RG.2.2.32322.59848>.

57. Kazaryan, A. (2025). Is Artificial Intelligence Protected by the First Amendment? *Freedom Forum*. Available at: <https://www.freedomforum.org/artificial-intelligence-first-amendment/>.
58. Klonick, K. (2022). AI-Generated Speech and the First Amendment. *Stanford Law Review*, 74(3), pp. 987–1020.
59. Lathan&Watkins. (2024). EU AI Act: Navigating a Brave New World the Act establishes the world's first comprehensive regulatory framework for AI, and is expected to shape the future of AI regulation and governance both within and beyond the EU, July 2024. Available at: <https://www.lw.com/admin/upload/SiteAttachments/EU-AI-Act-Navigating-a-Brave-New-World.pdf>.
60. López-Borrull, A. (2025). Mapping the impact of generative AI on disinformation. *Journal of Media Studies*, 13(3), pp. 33–50. Available at: <https://www.mdpi.com/2304-6775/13/3/33>.
61. Lu, N. (2023). Large language models can be guided to evade AI detection. *OpenReview.net*. Available at: <https://openreview.net/forum?id=ILE0mWzUrr>.
62. Mackintosh, E. (2020). European Union task force holds its first summit on fighting Russian disinformation. Available at: <https://edition.cnn.com/2020/01/30/europe/european-union-task-force-russian-disinformation-intl/index.html>.
63. Masrek, M. N.; Baharuddin, M.F. & Altaf, A. Perceived misinformation, disinformation and malinformation experience and the relationship with information overload, *Record and Library Journal*, Vol. 10, No. 2. <https://doi.org/10.20473/rjl.V10-I2.2024.212-234>.
64. Morosoli, S.; Naudts, L.; Cools, H. et al. (2025). Public accountability and regulatory expectations for AI in journalism: qualitative evidence from focus groups with Dutch citizens. *AI & Soc.* <https://doi.org/10.1007/s00146-025-02591-5>.
65. National Constitution Centre. (2025). National Constitution Centre. Available at: <https://constitution-center.org/blog/lawsuit-analyzes-first-amendment-protection-for-ai-chatbots-in-civil-case>.
66. Nicoleta Annemarie Munteanu (2022), Combating Misinformation at European Union Level, *The European Union in the Age of (In)Security: From Theory to Practice*, Cambridge Scholars Publishing, Lady Stephenson Library, Newcastle upon Tyne, UK, 2022, pp. 103-133.
67. Nicoleta Annemarie Munteanu. (2029). Cyber Security Threats. Combating Fake News, the European Commission Vision, *Studia Securitatis Journal*, Vol. XIII (1), 2019. pp. 55-70.
68. Office of the United States Trade Representative. 2023. U.S.-EU Joint Statement of the Trade and Technology Council. Available at: <https://ustr.gov/about-us/policy-offices/press-office/press-releases/2023/may/us-eu-joint-statement-trade-and-technology-council?utm>.
69. Park, S. (2025). Generative AI and misinformation: A scoping review of the dual role of large language models. *AI & Ethics*, 10(2), pp. 123–145. Available at: <https://link.springer.com/article/10.1007/s00146-025-02620-3>.
70. Pelc, P. (2024). Cybersecurity Issue in the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence from October 30, Cybersecurity and Law No.1 (11), p. 55.
71. Pennycook, G.; McPhetres, J.; Bago, B., & Rand, D. G. (2021). Attitudes about COVID-19 in Canada, the United Kingdom, and the United States: A comparative study. *Psychol. Sci.* 32, pp.2023–2036. DOI: 10.1177/0956797620939054.
- a. Pinge, A. (2024). First Formal Proceedings Launched by Commission under DSA: X Under Investigation, p. 245. Available at: <https://eucrim.eu/news/first-formal-proceedings-launched-by-commission-under-dsa-x-under-investigation/>.
72. Puczyńska, J. (2024). AI in disinformation detection. *Journal of Cybersecurity and Information Warfare*, 20(2), pp. 200–220. Available at: <https://www.acijournal.com/AI-in-Disinformation-Detection%2C200200%2C0%2C2.html>.
73. Romanishyn, A. (2025). AI-driven disinformation: Policy recommendations for mitigating the impact of generative models. *Journal of Information Integrity*, 12(3), pp. 45–62. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/>.

74. Romanishyn, A.; Malytska, O. & Goncharuk, V. (2025). AI-driven disinformation: policy recommendations for democratic resilience, *Frontiers in Artificial Intelligence*. <https://doi.org/10.3389/frai.2025.1569115>.
75. Salib, N. P. (2024). AI Outputs Are Not Protected Speech, *Washington University Law Review*, 102:83, Available at: https://wustllawreview.org/2024/11/05/ai-outputs-are-not-protected-speech/?utm_source=chatgpt.com#post-23164-footnote-383.
76. Souverain, T. (2025). Watermarking Large Language Models in Europe: Interpreting the Ai Act in Light of Technology. Available at: <https://arxiv.org/abs/2511.03641?utm>.
77. Supreme Court of the United States. (2019). Donald J. Trump, President Of The United States, Et Al., Petitioners V. Knight First Amendment Institute At Columbia University, Et Al. Available at: https://www.supremecourt.gov/DocketPDF/20/20-197/150726/20200820102824291_Knight%20First%20Amendment%20Inst.pdf?utm.
78. Supreme Court of the United States. (2023). Reynaldo Gonzalez, Et Al., Petitioners V. Google Llc. Available at: https://www.supremecourt.gov/opinions/22pdf/21-1333_6j7a.pdf?utm
79. The White House. (2025). Presidential Action. Removing Barriers to American Leadership In Artificial Intelligence. Available at: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.
80. U.S. Congress, H.R.6216. (2020). National Artificial Intelligence Initiative Act of 2020. Available at: <https://www.congress.gov/bill/116th-congress/house-bill/6216>.
81. Ummer, K., Asif, B., Ghouse, K. (2025). Disinformation Security at the Nexus of Cybersecurity and AI: Defending Digital Ecosystems Against Automated Deception. *World Journal of Advanced Engineering Technology and Sciences*, 15(02), pp. 2618–2625. DOI: <https://doi.org/10.30574/wjaets.2025.15.2.0813>
82. Ummer, K.; Asif, B. & Ghouse K. (2025). Disinformation Security at the Nexus of Cybersecurity and AI: Defending digital ecosystems against automated deception. *World Journal of Advanced Engineering Technology and Sciences*. 15(02), pp. 2618–2625. <https://doi.org/10.30574/wjaets.2025.15.2.0813>.
83. United State District Court for the Middle District of Florid. (2025). Garcia v. Character Technologies, Inc., No. 6:24-cv-01903 (M.D. Fla. May 21, 2025). Available at: <https://www.thefire.org/research-learn/order-motion-dismiss-garcia-v-character-technologies-inc?utm> .
84. United States Court of Appeals for the Eleven Circuit. (2024). Mark Walters v. OpenAI, L.L.C., No. 23-13843. Available at: <https://www.courtlistener.com/docket/68122080/mark-walters-v-openai-llc/?utm>.
85. United States Court of Appeals for the Fourth Circuit. (2022). White Coat Waste Project V. Greater Richmond Transit Company. Available at: <https://law.justia.com/cases/federal/appellate-courts/ca4/20-1710/20-1710-2022-05-20.html?utm>.
86. United States Court of Appeals for the Ninth Circuit. (2018). Tamara Fields, on behalf of herself, as a representative of the Estate of Lloyd Fields, Jr.; Heather Creach, on behalf of herself and as a representative of the Estate of James Damon Creach; J.C. (1), a minor; J.C. (2), a minor, v. Twitter, Inc. Available at: <https://cdn.ca9.uscourts.gov/datastore/opinions/2018/01/31/16-17165.pdf?utm> .
87. United States District Court for the District of New Mexico. (2018). Hunt v. Board of Regents of the University of New Mexico. Available at: <https://law.justia.com/cases/federal/district-courts/new-mexico/nmdce/1:2016cv00272/340322/24/>.
88. Vestager, Margrethe (2024). Commission sends preliminary findings to X for breach of the Digital Services Act Available at: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_24_3761/IP_24_3761_EN.pdf.
89. Vinay, R.; Spitale, G.; Biller, A. N. & Germani, F. (2025). Emotional prompting amplifies disinformation generation in AI large language models. *Frontiers in Artificial Intelligence*, 8, 1543603. Available at: <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1543603/full>.
90. Wagner, B. (2021). Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping? *Being Profiled*, pp. 253–272.

91. Wardle, C., & Derakhshan, H. (2017). Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking. Council of Europe. pp. 5-7. Available at: <https://rm.coe.int/information-disorder-report-november-2017/1680764666>.
92. White House. (2025). Removing Barriers to American Leadership in Artificial Intelligence. Available at: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>.
93. White House. (2023). Presidential Documents. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Available at: <https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf>.
94. Zhou, J. (2023). Understanding AI-generated misinformation and its implications. *ACM Computing Surveys*, 56(8), pp. 1318–1345. <https://dl.acm.org/doi/fullHtml/10.1145/3544548.3581318>.