German Law Journal (2025), pp. 1–22
doi:10.1017/glj.2025.24

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The European Constitutional Way to Address Disinformation in the Age of Artificial Intelligence

Giovanni De Gregorio[1] and Oreste Pollicino[2]

[1]Católica Global School of Law, Lisbon, Portugal and [2]Bocconi University School of Law, Milan, Italy
Corresponding author: Giovanni De Gregorio; Email: gdegregorio@ucp.pt

(Received 03 September 2024; accepted 05 December 2024)

Abstract
The spread of disinformation, such as false and fabricated content, as amplified by the expansion of artificial intelligence systems, has captured the attention of policymakers on a global scale. However, addressing disinformation leads constitutional democracies towards questions about the scope of freedom of expression as the living core of a democratic society. If, on the one hand, this constitutional right has been considered a barrier to public authorities' interferences to limit the circulation of disinformation, on the other hand, the spread of fabricated content and manipulative techniques, including deepfakes, has increasingly questioned liberal views. This constitutional challenge is further enriched by the role of online platforms which, by mediating speech in their online spaces, are essential tiles of a mosaic picturing the potential regulatory strategies and the limit of public enforcement to tackle disinformation. Within this framework, this work argues that the European constitutional approach to tackle disinformation has defined a unique model on a global scale. The European Union has developed a strategy that combines procedural safeguards, risk regulation, and co-regulation, as demonstrated by initiatives such as the Digital Services Act, the Strengthened Code of Practice on Disinformation, and the Artificial Intelligence Act. Positioned between liberal and illiberal models, the European approach proposes an alternative constitutional vision to address disinformation based on risk mitigation and the collaboration between public and private actors.

Keywords: Freedom of expression; disinformation; online platforms; artificial intelligence; fundamental rights

A. Introduction

Growing concerns about disinformation have been spreading in the last years.[1] The 2024 elections have been surrounded by fears about the creation and dissemination of false, fabricated, and misleading content. These concerns have been fueled by different cases, including the case of "disinformation for hire" about vaccinations,[2] information warfare strategies as underlined by the Ukrainian conflict,[3] political propaganda and populist narratives,[4] and the use of generative artificial intelligence (AI) applications such as Sora.[5]

[1]*See generally* C. R. Sunstein, Liars: Falsehoods and Free Speech in an Age of Deception (2021).

[2]Max Fisher, *Disinformation for Hire, a Shadow Industry, Is Quietly Booming: Back-ally Firms meddle in Elections and Promote Falsehoods on Behalf of Clients who Can Claim Deniability, Esclating Our Era of Unreality*, N. Y. Times (Jul. 25, 2021), https://www.nytimes.com/2021/07/25/world/europe/disinformation-social-media.html.

[3]*See* David Klepper, *Word War: In Russia-Ukraine War, Information Became a Weapon*, Associated Press (Feb. 23, 2023), https://apnews.com/article/russia-ukraine-technology-politics-782d23450e93b667afd7b57e0bba365f.

[4]*See* Paul Blokker, *Populism as a Constitutional Project*, 17 Int'l J. Const. L. 535, 550 (2019).

[5]*See* Lonni Besançon & Vahid Pooryousef, *What Is Sora? A New Generative AI Tool Could Transform Video Production and Amplify Disinformation Risks*, The Conversation (Feb. 19, 2024), http://theconversation.com/what-is-sora-a-new-generative-ai-tool-could-transform-video-production-and-amplify-disinformation-risks-223850.

The resulting fears about the ties between disinformation and democratic discourse have pushed leading constitutional democracies to wonder how to deal with the spread of online disinformation, particularly after the Brexit referendum and the U.S. presidential elections in 2016, which look nowadays like distant memories.[6] Although different regulatory measures and strategies have been adopted in different areas of the world,[7] as in the case of France and Brazil,[8] and judicial reactions have followed as in the case of the elections in Romania,[9] other constitutional systems have merely stepped aside, particularly the U.S., even after the spread of generative AI applications.[10]

This fragmented framework of reactions primarily results from the different views about the constitutional relevance of disinformation. Even when AI technologies are involved in the production and dissemination of fabricated content, addressing the spread of disinformation is primarily a matter of understanding the role of freedom of expression in a democratic society, which, indeed, does not always enjoy the same degree of protection across constitutional systems.[11] Therefore, different approaches to this fundamental right have led to liberal, democratic, or even repressive answers against the spread of online disinformation, thus showing a profound disagreement about the conceptualization of a democratic society.

These different constitutional points of view are also reflected in the answers to the transformation and privatization of the marketplace of ideas,[12] which looks anything but free in the digital age. This situation is indeed even more compelling for democracy when one considers the power of transnational private actors, primarily online platforms, to make decisions about online content. By relying on automated systems in content moderation, these actors primarily govern digital spaces by making decisions on online content, including disinformation.[13] As a result, the emergence of powerful private actors, and the transformation of how power itself is administered between public and private actors,[14] coupled with the increasing implementation of AI systems, have amplified constitutional concerns related to the spread of online disinformation and the potential strategies to address this issue. As a matter of fact, as Maduro and de Abreu Duarte have pointed out, "fostering a large community—similar to a public sphere—is key to the

[6]*See* Judit Bayer, Bernd Holzanagel, Katarzyna Lubianiec, Adela Pintea, Josephine B. Schmitt, Judit Szakács & Erik Uszkiewicz, Disinformation and Propaganda: Impact on the Functioning of the Rule of Law and Democratic Processes in the EU and Its Member States (2021), https://www.europarl.europa.eu/thinktank/en/document/EXPO_STU(2021)653633 (presenting a study completed by the authors at the request of the INGE committee of the European Parliament).

[7]*See* Daniel Funke & Daniela Flamini, *A Guide to Anti-misinformation Actions Around the World*, Poynter (Aug. 13, 2025), https://www.poynter.org/ifcn/anti-misinformation-actions/.

[8]*See* Loi n. 2018-1201 du 22 décembre 2018 de Lutte Contre la manipulation de l'information [Law 2018-1201 of December 22, 2018 on the Fight Against Information Manipulation], Journal Officiel de la République Française [J.O.] [Official Gazette of France], Dec. 23, 2018 ; Decreto No. 2.630 de 25 Abril de 2023, Diário Oficial da União [D.O.U.] (Braz.).

[9]Decizia privind anularea procesului electoral cu privire la alegerea Președintelui României din anul 2024 [Decision on the cancellation of the electoral process regarding the election of the President of Romania in 2024] Curtea Constituțională a României [Constitutional Court of Romania] No. 32.

[10]*See* Dick Durbin, U.S. Senate Majority Whip (D-IL), U.S. Senate Committee on the Judiciary, Opening Statement During Fudiciary Subcommittee Hearing on Oversight of Artificial Intelligence (May 16, 2023) (transcript, video, and audio available at https://www.judiciary.senate.gov/press/dem/releases/durbin-delivers-opening-statement-during-judiciary-subcommittee-hearing-on-oversight-of-artificial-intelligence).

[11]*See generally* Eric Barendt, Freedom of Speech (2007).

[12]*See* Daniel E. Ho & Frederick Schauer, *Testing the Marketplace of Ideas*, 90 N.Y.U. L. Rev. 1160, (2015); Eugene Volokh, *In Defense of the Marketplace of Ideas / Search for Truth as a Theory of Free Speech Protection*, 97 Va. L. Rev. 595, (2011); Alvin I. Goldman & James C. Cox, *Speech, Truth, and the Free Market for Ideas*, 2 Legal Theory 1, (1996).

[13]*See* Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. Rev. 27, 29 (2019).

[14]*See* Oreste Pollicino, *The Quadrangular Shape of the Geometry of Digital Power(s) and the Move towards a Procedural Digital Constitutionalism*, 29 Eur. L. J. 1, 16 (2023); Martin Moore & Damian Tambini, Digital Dominance: The Power of Google, Amazon, Facebook, and Apple (2018).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

business model" of major platforms.[15] The powerful role of online platforms underlines the reasons for a growing reliance of constitutional democracies on private actors in the algorithmic society,[16] which is characterized by the intimate connection between the public and the private sphere.

The European Union has demonstrated to be mindful of this intertwined scenario made of manipulated content and private governance as particularly underlined by the adoption of the Digital Services Act (DSA),[17] the Strengthened Code of Practice on Disinformation,[18] and the Artificial Intelligence Act (AI Act).[19] Together with other pieces of the strategy such as the Regulation on Transparency of Political Advertising (PAR),[20] and the European Media Freedom Act (EMFA),[21] these tools aim to tackle disinformation not by regulating speech but by targeting the dynamics affecting its circulation,[22] primarily looking at online platforms, and strengthening the ties between public and private actors. Even if Member States keep their competence in terms of content regulation, the European strategy on disinformation is a landmark example of how constitutional democracies could provide a different regulatory answer to the spread of disinformation.

Within this framework, this Article aims to examine how the European approach to online disinformation leads to a unique constitutional strategy. The Article argues that the Union is providing a model to address disinformation which does not focus on content regulation but on dealing with the dynamics characterizing the spread of disinformation, mainly online platforms and AI systems. Rather than arguing for a self-regulatory or illiberal approach, the Union proposes a hybrid strategy based on a regulatory mix based on procedural safeguards, risk regulation and co-regulation which involves public and private collaboration to address disinformation.

The first part of this Article focuses on the relevance of online disinformation in the age of AI from a European constitutional perspective. The second part examines the predominance of online platforms in governing digital spaces and underlines the reasons for the limited effectiveness of the strategies implemented by constitutional democracies to tackle online disinformation. The third part analyses the regulatory mix characterizing the European approach to fight online disinformation, as emphasized by the DSA, the Strengthened Code of Practice on Disinformation, and the AI Act. The fourth part decodes the features of the European constitutional way to tackle online disinformation.

[15]*See* Miguel Maduro & Francisco de Abreu Duarte, *Regulating Big Tech Will Take Pluralism and Institutions*, EURONEWS (Oct. 7, 2021), https://www.euronews.com/2021/10/07/regulating-big-tech-will-take-pluralism-and-institutions-view.

[16]*See* Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 UCD L. REV. 1149, 1153 (2018).

[17]*See* Comission Regulation 2022/2065 of October 19, 2022, on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) [hereinafter DSA Directive].

[18]*See* EUR. COMM'N, 2022 STRENGTHENED CODE OF PRACTICE ON DISINFORMATION (2022); *Commission Communication on European Commission Guidance on Strengthening the Code of Practice on Disinformation*, COM (2021) 262 final (May 26, 2021).

[19]*See* Commission Regulation 2024/1689 of June 13, 2024, Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024 O.J. (L 1689) [hereinafter AI Act].

[20]*See* Commission Regulation 2024/900 of Mar. 13, 2024, on the Transparency and Targeting of Political Advertising, 2024 O.J. (L 900).

[21]*See* Commission Regulation 2024/1083 Apr. 11, 2024, Establishing a Common Framework for Media Services in the Internal Market and Amending Directive 2010/13/EU, 2024 O.J. (L 1083) [hereinafter EMFA].

[22]Martin Husovec, *The Digital Services Act's Red Line: What the Commission Can and Cannot Do About Disinformation*. 16(1) J. MEDIA L. 47 (2024).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

B. The Constitutional Relevance of Disinformation

The questions around addressing disinformation directly touch on issues and concepts—namely democracy and freedom above all—which are at the heart of constitutionalism and democracy. Just as an example, the proliferation of populist movements on social media has led to an increase in the dissemination of propaganda. At the same time, the use of information operations for external interference has become a significant phenomenon in the digital age,[23] thus raising questions beyond constitutional law.[24]

The role of AI systems in contributing to the spread of disinformation is increasingly relevant.[25] Although AI can play a critical role in countering the spread of disinformation, the creation of false and fabricated content through the emergence of countless systems, linked in particular to generative AI, foundational models, and large language models capable of creating highly realistic images, videos, and synthetic texts, has made the exchange of ideas increasingly artificial and exposed public debate to potential exploitation from external interferences through information operation aiming to deceive citizens.

In this context, the number of instances where AI systems have been used to create false content, particularly deepfakes, is a cause for significant concern. This is particularly evident in the current global situation, which has been further exacerbated by the outbreak of conflicts, including the Russian-Ukrainian and Israeli-Palestinian ones. For instance, there are deepfake videos depicting U.S. Democratic Representative Alexandra Ocasio-Cortez discussing, in a rambling manner, the request for a ceasefire in Gaza,[26] or Sadiq Khan declaring his intention to postpone Armistice Day, a day dedicated in the United Kingdom to remembering the tragic events of the First World War, to allow for the organization of a march in favor of Palestine.[27] These are two examples of the spread of deepfakes in the digital environment.

These examples underline how the production of speech in a democratic society is increasingly based on AI systems, which essentially lend themselves to a plurality of possible uses aimed at producing materials and contents of an eminently disinformative nature, including AI hallucinations. This risk particularly showed up, most recently, in the guidelines drawn up by the European Commission at the beginning of 2024 for European elections. Particularly, the Commission underlined that the recent technological developments in generative AI "may bring many new opportunities, [and] they may lead to specific systemic risks in the context of elections."[28] In essence, the Commission recognizes that generative AI offers important tools for producing information and, therefore, opens the doors to a significant enrichment of the information landscape itself. Still, it requires at the same time to address the new threats connected to the manipulation of democratic processes.

Furthermore, targeting techniques becoming more sophisticated as empowered by AI systems also amplify this risk for democracy. The personalization of electoral messages and the use of chatbots opens the possibility of an even greater targeting of citizens, including the potential

[23]*See generally* Noémi Bontridder & Yves Poullet, *The Role of Artificial Intelligence in Disinformation*, 3 DATA & POL'Y 1, (2021); Katarina Kertysova, *Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation Is Produced, Disseminated, and Can Be Countered*, 29 SEC. & HUM. RTS. 55, (2018).

[24]*See* Björnstjern Baade, *Fake News and International Law*, 29 EUR. J. INT'L L. 1357 (2018).

[25]*See* CHRIS MARSDEN & TRISHA MEYER, REGULATING DISINFORMATION WITH ARTIFICIAL INTELLIGENCE: EFFECTS OF DISINFORMATION INITIATIVES ON FREEDOM OF EXPRESSION AND MEDIA PLURALISM (Mar. 2019) (comprising a study completed through the European Parliamentary Research Service).

[26]Rob Lever, *Deepfake Mocks Alexandria Ocasio-Cortez's Call for Gaza Ceasefire*, AGENCE FRANCE-PRESSE (Apr. 29, 2024), https://factcheck.afp.com/doc.afp.com.34693MW.

[27]*See* Dan Sabbagh, *Faked Audio of Sadiq Khan Dismissing Armistice Day Shared Among Far-right Groups*, THE GUARDIAN (Nov. 10, 2023), https://www.theguardian.com/politics/2023/nov/10/faked-audio-sadiq-khan-armistice-day-shared-among-far-right.

[28]*Commission Communication for Providers of Very Large Online Platforms and Very Large Online Search Engines on the Mitigation of Systemic Risks for Electoral Processes Pursuant to Article 35(3) of Regulation (EU) 2022/2065*, COM (2024) 3014 final (Apr. 26, 2024).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

categorization of individuals in order to develop parallel electoral campaigns according to individual demographic groups.[29] Particularly, political micro-targeting techniques raises questions about the protection of the right to individual self-determination and the right not to be subjected to profiling lacking valid consent.[30]

These challenges lead to wondering about the boundaries of freedom of expression in the age of AI. To what extent can constitutional democracies tolerate the spread of false and fabricated content as a price to pay to maintain a free public debate is a matter that depends on how they understand freedom of expression. As underlined in the next sub-sections, the approaches followed by constitutional democracies to address disinformation are primarily connected to how constitutional democracies conceive the role of freedom of expression in a democratic society.

I. Freedom of Expression and Disinformation

The relevance of the right to freedom of expression and the crucial role of falsehood have been already underlined in the seventeenth century by Milton,[31] and, in the nineteenth century, by Mill,[32] supporting a liberal view considering that even falsehood could contribute to reaching the truth, especially by avoiding the risk of dogmatization of knowledge.[33] Milton, in his *Aeropagitica*, lashed out against censure of the press, citing the concept of truth and comparing knowledge to water and the truth to a gushing source.[34] What must be avoided, in this paradigm, is whatever can block the free flow of ideas that leads to progress towards truth. Censure could thus affect that process of approaching the truth by impeding or restricting the emergence of new ideas. According to Milton, truth prevails in a free and open context of ideas.[35] Therefore, those ideas cannot be subject to limitations ahead of time that can compete in the battle against dogmas, even if Milton accepted the role of law as a potential limit to the spread of falsehood.

These underpinning liberal ideas are still the core of the right to freedom of expression, as underlined in the twentieth century by Justice Holmes in his dissenting opinion in the U.S. Supreme Court decision, *Abrams v. United States*.[36] According to Justice Holmes, men try to support their positions by criticizing opposing ideas, but they must not be persuaded from the start that their opinions are certain.[37] Holmes writes that the best test of truth "is the power of the thought itself to get accepted in the competition of the market."[38] This suggests that "the free market of ideas" can help uncover what the truth actual is and the publication of false information can serve a truth seeking function. Constitutional democracies tend to tolerate the political exchange of views as a precondition of pluralism, or, to use a neo-liberal metaphor, of the free marketplace of ideas.[39] Although the spread of disinformation can produce serious consequences on public opinion, it has still been considered an opportunity for promoting the exchange of ideas or a price to pay to live in a democratic society.

[29] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions on the European democracy action plan*, COM (2020) 790 final (Dec. 3, 2020), 3

[30] *See generally* Cristina Blasi Casagran, Mathias Vermeulen, *Reflections on the murky legal practices of political micro-targeting from a GDPR perspective*, 11(4) Int'l Data Privacy L. 348 (2021).

[31] *See generally* John Milton, Areopagitica and Other Prose Works of John Milton (1927).

[32] *See generally* John Stuart Mill, On Liberty (1859).

[33] *Id.*

[34] *See generally* John Milton, *Areopagitica: A Speech for the Liberty of Unlicensed Printing to the Parliament of England* (1644).

[35] *Id.*

[36] *See Abrams v. United States*, 250 U.S. 616, 630 (1919) (J. Holmes, dissenting).

[37] *Id.* at 630.

[38] *Id.*

[39] *See generally* Alvin. Goldman & James Cox, Speech, truth, and the free market for ideas (1996).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

Nonetheless, this liberal view based on the metaphor of the free marketplace of ideas, and generally the broader scope of the First Amendment in the U.S. Constitution, has not always been shared on a global scale, even before the massive spread of online disinformation. Constitutional democracies have provided different answers to identify the truth, and even before that, the criteria necessary to define it and to separate the truth from what can be catalogued as false.[40] Although constitutional democracies tend to agree on the relevance of freedom of expression for democracy, they still disagree on where to draw the line of free speech to protect other constitutional interests deserving protection such as dignity, legitimate public interests like security, and collective democratic values such as trust and transparency.[41]

Even in liberal systems, the spread of false content has been considered a threat to dignity as in the case of defamation, which is usually condemned by criminal law.[42] Likewise, this type of content is usually considered a challenge for consumers and limited as a misleading practice in advertising, and this view is shared not only in Europe due to the relevance of consumer law but also in the U.S.[43] However, the protection of free speech in the U.S. constitutional system is broad, thus limiting the possibility for public authorities and institutions to address the spread of online disinformation. Even the constitutional questions raised by AI systems do not seem to be enough to trigger a reaction overcoming the scope of protection granted to the U.S. First Amendment, especially considering how AI can generate content even if this point is controversial.[44] If, on the one hand, such a tolerance is critical for freedom of expression, on the other hand, it risks preempting public actors, including lawmakers, to intervene to safeguard the same freedom or other democratic values when they are threatened by interferences from public and private actors. As underlined by Popper,[45] when a society is tolerant without limits, its tolerance is diluted or loses its effect, thus looking at intolerance as a solution to intolerance. The paradox of intolerance could particularly affect constitutional democracies where tolerance is an essential piece of the system.

This approach differs in some Asian countries. During the pandemic, the collectivist approach of several countries, including Japan, Singapore and China, has made it easier to react to the challenges raised by the spread of the virus. Even before the pandemic, Singapore's Parliament passed the Protection from Online Falsehoods and Manipulation Act in 2019 which led to harsh criticism from civil society, academia and internet platforms for its far-reaching effects.[46] This legislation targets content that is "false or misleading, whether wholly or in part" and/or there are reasons to believe it affects public interest.[47] As underlined by Byung Chul Han,[48] Asian states share a different perspective of paternalism and authority, primarily driven by the Confucian

[40]*See generally* Oreste Pollicino, Freedom of Speech and the Regulation of Fake News (2023).

[41]*See, e.g.,* Decizia privind anularea procesului electoral cu privire la alegerea Președintelui României din anul 2024 [Decision on the cancellation of the electoral process regarding the election of the President of Romania in 2024] Curtea Constituțională a României [Constitutional Court of Romania] No. 32.

[42]*Id.* at 11. For istance, Article 595 of the Italian Criminal Code, according to which "Whoever, outside the cases indicated in the preceding Article, by communicating with several persons, offends against the reputation of others, shall be punished by imprisonment of up to one year or a fine of up to one thousand thirty-two euro." *Id.*

[43]*See* Leslie Gielow Jacobs, *Freedom of Speech and Regulation of Fake News*, 70 Am. J. Compar. L. 278, (2022).

[44]*See, e.g.*, Alec Peters, *Machine Manipulation: Why an AI Editor Does Not Serve First Amendment Values*, 95 U. Colo. L. Rev. 307 (2024); Margot Kaminski & Meg Jones, *Constructing AI Speech*, 133 Yale L. J. F. 1212 (2024); Cass R. Sunstein, *Artificial Intelligence and the First Amendment*, SSRN Database https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4431251; Eugene Volokh, Mark A. Lemley & Peter Henderson, *Freedom of Speech and AI Output*, 3 J. Free Speech L. 651 (2023); Stuart Minor Benjamin, *Algorithms And Speech*, 161 Univ. Pa. L. Rev. 1445 (2013).

[45]*See generally* Karl R. Popper, The Open Society and Its Enemies (1952).

[46]*See* Protection from Online Falsehoods and Manipulation Act 2019, (2020) § 2 (Sing.).

[47]*Id.*

[48]*See generally* Byung-Chul Han, Infocracy: Digitalization and the Crisis of Democracy (2022).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

legacy.[49] This tendency opens the doors towards policies of surveillance limiting the relevance of individuals' rights and freedoms.

In the European constitutional tradition, the central position occupied by freedom of expression does not exclude potential restrictions due to the need to prevent abuses or to balance its exercise with other rights, which equally deserve constitutional protection. From the time of its solemn affirmation, found in the Declaration of the Rights of Man and of the Citizen of 1789,[50] freedom of expression has had an intrinsically malleable nature, that can be inferred from its very formulation: "The free communication of ideas and opinions is one of the most precious of the rights of man. Every citizen may, accordingly, speak, write, and print with freedom, but shall be responsible for such abuses of this freedom as shall be defined by law."[51] Therefore, with respect to disinformation, the essential questions on its limitability encompass first whether and to what extent falsehood may be granted protection under freedom of expression and, second, whether dissemination of falsehood with an intent to harm identified targets or society at large may be restricted or subject to sanctions.

This view of free speech can be considered the trigger for European countries such as Germany and France to adopt legislation to limit the spread of hate speech and online disinformation. Particularly, the adoption of the Network Enforcement Act (NetzDG), and its amendments, in Germany has defined a new system of procedural safeguards in the process of content moderation, *de facto* anticipating the European strategies on online platforms.[52] Likewise, the law against disinformation in times of election in France,[53] which has introduced an urgency-based procedure requiring judicial authorities to act upon requests within forty-eight hours of their reporting to make a decision on the truthfulness of the content in question. Furthermore, the annulment of the national election in Romania by the constitutional court considering the spread of disinformation is another example of the reactive European approach to freedom of expression.[54] However, the same approach has not characterized the strategies of other Member States, primarily Scandinavian countries, which have relied more on media literacy and other soft law instruments to target disinformation.[55]

Despite national nuances, the protection of freedom of expression in Europe is subject to an express balancing with other fundamental rights and may be subjected to conflicting legitimate interests. Even if freedom of expression is a critical value of the Union, its scope of protection is limited to protect other constitutional values as underlined by the European Convention on Human Rights (ECHR),[56] and the Charter of Fundamental Rights of the European Union.[57] Particularly, interfering with fundamental rights in Europe should always look at the protection of

[49]*See generally The Confucian Legacy: World-Historical Writing at the Turn of the Twentieth Century*, *in* WORLD HISTORY AND NATIONAL IDENTITY IN CHINA: THE TWENTIETH CENTURY 16 (Xin Fan ed., 2021).

[50]*See* Décret de 26 août 1789 de La Déclaration Des Droits De L'Homme et du Citoyen [The Declaration of the Rights of Man and of the Citizen of August 26], Aug. 26, 1789, (Fr.).

[51]*See id.* at art. 11.

[52]*See* Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [NetzDG] [Network Enforcement Act], Sept. 1, 2017, BUNDESGESETZBLATT TEIL I [BGBL I] at 3352 (Ger.).

[53]*See* Loi 2018-1202 du 22 décembre 2018 de relative à la lutte contre la manipulation de l'information [Law 2018-1202 of December 22, 2018, on the Fight Against the Manipulation of Information], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANCAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Dec. 22, 2018.

[54]*See, e.g., When Elections are Undone: Democracy & Disinformation in the 2024 Romanian Vote*, EUROPEAN UNIVERSITY INSTITUTE (Mar. 11, 2025), https://www.eui.eu/news-hub?id=when-elections-are-undone-democracy-disinformation-in-the-2024-romanian-vote.

[55]*See* NORDIS, ASSESSING INFORMATION DISORDER IN THE DIGITAL MEDIA WELFARE STATE: A RIGHTS-BASED APPROACH 10 (2024) .

[56]*See* Convention for the Protection of Human Rights and Fundamental Freedoms, Apr. 11 1950, 2889 U.N.T.S. 213, art. 10(2) [hereinafter ECHR].

[57]*See* Charter of Fundamental Rights of the European Union, 2012 O.J. (C 326) art. 52 (Oct. 26, 2012) [hereinafter Charter].

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

the rights' essence which cannot be touched and sacrificed in the name of the protection of other rights and freedoms.[58]

The essence of fundamental rights is also protected by the abuse of rights clause which aims to avoid granting absolute protection to one right, thus undermining de facto the protection of the other constitutional rights, and, therefore, their constitutional relevance,[59] particularly underlined by the role of dignity in European constitutionalism.[60] The clause of abuse of rights in European constitutionalism indeed leads to a different approach which rejects a rigid axiology among constitutional rights and compels public powers to take into account the entire constitutional framework and the implications of the potential clash between different rights and freedoms.

This constitutional architecture not only underlines the centrality of balancing in European constitutionalism but also makes it possible to look at freedom of expression not as an absolute right but as a part of a broader constitutional framework. As a result, the expansion of disinformation policies in EU is primarily connected to the idea that freedom of expression is not an absolute right and is protected not only as a negative freedom but also as a positive right. The spread of false and fabricated content raises questions not only about the protection of free speech but also about trust and self-determination in democratic societies. Although digital technologies have significantly contributed to information pluralism,[61] promoting the exchange of information and opinions at unprecedented levels,[62] great quantities of online content, including manipulated and fabricated content, have broken the traditional checks and filters on the quality and reliability of sources, primarily operated by media outlets.

This fragmented framework underlines how the migration of constitutional ideas, such as the free marketplace of ideas metaphor, does not always find space for cross-fertilization. The constitutional underpinnings justifying the U.S. liberal approach lead to the risk of excessive tolerance and trust in a public discourse which has been driven by market dynamics. When looking at the dissemination of disinformation, the marketplace of ideas metaphor has been profoundly challenged and decontextualized given that the economic market, as the source domain from which the metaphor has been taken, is far from free.[63]

Metaphorical language fits well with legal reasoning, but it should be handled properly—and with care.[64] The free market of ideas metaphor carries over from the source domain of economic activity to the target domain of speech a systematic set of entailments that supersedes the limitations of the older free speech model. Holmes wrote in a period of *laissez-faire* capitalism, in which the liberal state and market competition were at their zenith. If he was skeptical about any external verification of the truth and removal of news proven to be false, the concept of a free market provided a meaningful alternative model for the notion that truth, just as economic well-being, could result from competition between—true and false—ideas and information. Likewise, when the U.S. Supreme Court borrowed the metaphor, referring to the internet as the "new marketplace of ideas,"[65] the economic market of the Internet was free and not in any way affected by the private governance of transnational corporations.

The multifaceted character of disinformation requires constitutional democracies to deal with challenges which are not exclusively related to the role of truth in a democratic society or the potential interferences of public authorities. Addressing online disinformation calls for thinking

[58]*See id.*

[59]*See id.* at art. 54; ECHR, *supra* note 56, at art. 17.

[60]*See* Catherine Dupr, The Age of Dignity: Human Rights and Constitutionalism in Europe (2015).

[61]*Ahmet Yıldırım v Turkey,* App No. 3111/10 (Mar. 18, 2013), https://hudoc.echr.coe.int/fre?i=001-115705.

[62]Case C-160/15 *GS Media BV v Sanoma Media Netherlands BV and Others* ECLI:EU:C:2016:644, para 45.

[63]*See* Morgan Weiland, *First Amendment Metaphors: The Death of the "Marketplace of Ideas" and the Rise of the Post-Truth "Free Flow of Information"*, 33 Yale J. L. & Humans. 366, 383–84 (2022).

[64]*See generally* Alessandro Morelli & Oreste Pollicino, *Metaphors and Judicial Enforcement of Fundamental Rights*, 68 Am. J. Compar. L. 616 (2020).

[65]*See Reno v. Am. Civ. Liberties Union*, 521 U.S. 844, 885 (1997).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

about how to deal with the exercise of powers, particularly when the threats to free speech do not come only from public intervention but also from private determinations driven by profit maximization and mediated by AI systems. As underlined in the next section, online platforms play a predominant role in digital spaces, enforcing private standards on freedom of expression online. The private governance of expressions in the digital age underlines how addressing disinformation is not only about understanding to what point constitutional democracies can tolerate falsehood but also dealing with the role of private actors in mediating speech on a global scale.

II. Online Disinformation and Private Governance

The expansion of the digital age indeed has been dominated by neoliberal narratives in the last twenty years that limited the intervention of public authorities or regulation.[66] At the end of the last century, the Internet was considered the engine for the exchange of opinions and expressions, as well as for pluralism, which, at that time, was concerned with scarcity of resources and market regulation of the media.[67] Although one of the priorities in the media sector is to protect the pluralism of information, the internet promises to break old barriers and foster media pluralism. If it is true that the First Amendment broadly protects falsehood, this is even truer in the digital age, thanks to the enhanced opportunity to express thoughts.

When in the translation from the world of atoms to that of bits, the Supreme Court defined the Internet as the new marketplace of ideas at the end of the 1990s; the Supreme Court actually referred to the founding moment for the birth and development of cyberspace as a truly free *market* at that time, which deceived the pioneers of the web into believing it could really be independent from the real world.[68] In other words, the dominant narrative has been based on the idea that public powers should not have any role in dealing with the ever-growing online disinformation, because users are—optimistically—supposed to have all the tools they need in order to select the best information and disregard false news. Within this context, the metaphor of the free marketplace of ideas and the proposed test for the truth—competition in the absence of any public control—made perfect sense.

This view explains why the spread of disinformation has not always been addressed as a problem from the perspective of constitutional law. It is based on complete trust in the capacity of the information market to self-correct and a sense of distrust towards the role of public actors, for instance, in ensuring media policy such as pluralism. Just as the economic market knows no test of product validity but allows demand to drive supply, relying on the market to distinguish between viable and shoddy products,[69] the best way of dealing with the phenomenon of disinformation is to secure the widest possible dissemination of all news, including news from contradictory and unreliable sources.

The rise of the internet and the consolidation of online platforms in the digital environment have challenged this vertical paradigm. As observed by Balkin, in the digital age, freedom of expression is like a triangle.[70] The regulation of speech no longer involves the relationship between the states and the speaker but also multiple players outside the control of the state, such as social media. Unlike traditional media outlets, social media can be considered governors of digital

[66]Oreste Pollicino, *The Judicial Bridges of Privacy and Speech in the Information Society*, *in* Judicial Protection of Fundamental Rights on the Internet A Road Towards Digital Constitutionalism 182 (2021).

[67]Michel Rosenfeld, *Hate Speech in Constitutional Jurisprudence: Comparative Analysis*, 24 Cardozo Law Review 1529 (2003).

[68]*See* David R. Johnson & David Post, *Law and Borders: The Rise of Law in Cyberspace*, 48 Stan. L. Rev. 1367, 1367 (1996).

[69]*See* Rosenfeld, *supra* note 67, at 1529–35.

[70]*See* Jack M. Balkin, *Free Speech Is a Triangle*, 118 Colum. L. Rev. 2011, (2018).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

spaces,[71] by performing content moderation activities implementing automated systems which can decide in a heartbeat whether to delete or demote online content and how to organize that information in social media spaces.[72]

The manner in which individuals express opinions and ideas online has changed in the last twenty years, and the digital environment has been a crucial vehicle to foster democratic values such as freedom of expression.[73] Particularly, in digital spaces, the predominant role of online platforms on free speech defines an imperative for constitutional democracies to engage with transnational actors to address the issue of disinformation. These entities have business and incentive models that challenge traditional media regulation approaches.[74]

Disinformation would not be such an issue if online platforms had not risen as gatekeepers,[75] or, from a constitutional perspective, private powers.[76] The digital liberal approach has led to developing new business models based on content moderation and users profiling to offer tailored advertising services. As observed by Gillespie, content moderation is not an ancillary activity.[77] Quite the opposite, it is essential for platforms in order to ensure a safe environment where users can share their content freely. As a result, the interest of platforms is not just focused on facilitating the spread of opinions and ideas across the globe, but establishing a digital environment where users feel free to share information and data that can feed commercial networks and channels and, especially, attract profits coming from advertising. This system encourages users to spend more time online and interact with content. The fact that virtually every internet user can become a creator or an editor to share—even false—information and the corresponding much greater potential impact of falsehoods on the internet are exponentially amplifying the questions raised by the post-truth era,[78] and are even more compelling with the expansion of generative AI systems.

The increasing role of AI technologies in content moderation has contributed to shaping the enforcement and protection of online speech. The information uploaded by users is processed by automated systems that define–or at least suggest to human moderators–content to remove or shadow ban in a bunch of seconds according to technical opaque standards and without providing users access to any remedy against a specific decision.[79] If, on the one hand, content moderation constitutes an important resource for online platforms, on the other hand, the use of technologies, primarily machine learning, for moderating content on a global scale shapes the scope of disinformation in the digital environment.

The organization and circulation of online content is primarily driven by logic that is far from constitutional narratives about fundamental rights and public interests, such as tackling the spread of disinformation. Notwithstanding several social media exploit rhetoric statements advocating to represent a global community enhancing free speech transnationally, online

[71] *See* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1599, 1601–02 (2018).

[72] *See* Tarleton Gillespie, *Content Moderation, AI, and the Question of Scale*, 7 BIG DATA & SOC'Y 1, 3 (2020).

[73] *See generally* YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM (2006).

[74] *See generally* PHILIP MICHAEL NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST: MEDIA REGULATION IN THE DISINFORMATION AGE (2019).

[75] *See* Emily B. Laidlaw, *A Framework for Identifying Internet Information Gatekeepers*, 24 INT'L REV. L. COMPUT. & TECH. 263 (2010); Jonathan Zittrain, *A History of Online Gatekeeping*, 19 HARV. J. L. & TECH. 253 (2006).

[76] *See* Giovanni De Gregorio, *Democratising Online Content Moderation: A Constitutional Framework*, 36 COMPUT. L. & SEC. REV. 1, 17 (2020).

[77] *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 21 (2018).

[78] *See* YOCHAI BENKLER, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS 37 (2018).

[79] *See* Sarah Myers West, *Censored, Suspended, Shadowbanned: User Interpretations of Content Moderation on Social Media Platforms*, 20 NEW MEDIA & SOC'Y 4366, 4374 (2018).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

platforms increasingly focus their attention on content moderation to avoid losing users' trust and, therefore, advertising revenues while also answering regulatory pressures.[80] These actors are increasingly called to monitor and remove online speech to limit any risk related to the spread of harmful content. The case of the genocide in Myanmar or the Cambridge Analytica scandal have increased the pressure on online platforms to behave responsibly.[81] This trend potentially fosters collateral censorship,[82] which occurs when private actors are entrusted to remove unlawful content when they become aware of its presence. Because online platforms are privately run, these actors would likely try to avoid the risks of being sanctioned for non-compliance. In other words, online platforms, as business actors, would likely focus on minimizing this economic risk rather than adopting a constitutional-based approach.

In the case of disinformation, social media, such as Meta and Twitter—now X—have proposed voluntary measures and policies to address disinformation, and they have been at the forefront in removing or signaling alleged false content.[83] This voluntary fight against disinformation led to adopting an Executive Order to react to Twitter's discretion in placing fact-checking labels on presidential tweets relating to mail-in ballots and election fraud.[84] Furthermore, platforms have also been involved in providing information in times of election, even if they met resistance as in the case of the Spanish Data Protection Authority (AEPD), which used its emergency power to ban Meta from launching a feature on Facebook and Instagram to collect data on voters in Spain.[85]

However, the pandemic has underlined how the implementation of AI in content moderation can contribute to the spread of disinformation without human oversight. The decision of Google and Meta to limit the process of human moderation has affected the entire process of content moderation with the result that different accounts and content have been suspended even if there was no reason to remove the content.[86] This situation has not only affected users' rights but also led to the spread of disinformation in a time where reliance over good health information has been critical.[87]

The role of online platforms in governing free speech, even during crises, provides another example of how the digital marketplace of ideas is not free or even neutral. Against this background, if the private governance of online speech is arguably one of the most significant and pervasive sources of failure in the marketplace of ideas, the possibility to intervene to address this challenge raises fewer constitutional concerns. In contrast to the U.S. Supreme Court's definition of the internet as the new free marketplace of ideas, the source domain of the digital relevant

[80] *See* GILLESPIE, *supra* note 77, at 72.

[81] *See* Steve Stecklow, *Why Facebook is Losing the War on Hate Speech in Myanmar*, REUTERS (Aug. 15, 2018), https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/; Nicholas Confessore, *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*, N.Y. TIMES (Apr. 4, 2018), https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

[82] *See* Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 295–96 (2011); Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 UNIV. PA. L. REV. 11, 14 (2006).

[83] *See* Jay Peters, *Twitter Introducing New Labels for Tweets with Misleading COVID-19 Information*, VERGE (May 11, 2020), https://www.theverge.com/2020/5/11/21254733/twitter-covid-19-misleading-information-label-warnings-misinformation.

[84] Exec. Order No. 13,925, 85 Fed. Reg. 34,079 (June 2, 2020) (striving to prevent online censorship).

[85] *See The Agency Orders a Precautionary Measure That Prevents Meta from Implementing the Electoral Functionalities That it Plans to Launch in Spain*, AGENCIA ESPAÑOLA PROTECCIÓN DATOS (May 31, 2024), https://www.aepd.es/en/press-and-communication/press-releases/the-agency-orders-precautionary-measure-prevents-meta.

[86] *See* Elizabeth Dwoskin & Nitasha Tiku, *Facebook Sent Home Thousands of Human Moderators Due to the Coronavirus. Now the Algorithms Are in Charge*, WASH. POST (July 23, 2020), https://www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/.

[87] *See* Rosalie Gillett & Tobias R. Keller, *Why Is It so Hard to Stop COVID-19 Misinformation Spreading on Social Media?*, THE CONVERSATION (Apr. 13, 2020), http://theconversation.com/why-is-it-so-hard-to-stop-covid-19-misinformation-spreading-on-social-media-134396.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

market is anything but a free market, being characterized by economic concentration and the strength of—a few—private actors. Here, the point is not about the opportunity to set a public tribunal of the true or enhance the liability regime of online platforms but to recognize the limits of metaphors in legal imagination and define a constitutional strategy to address online disinformation.

However, even a reactive regulatory approach risks failing to address this situation. Relying exclusively on top-down solutions would not only affect freedom of expression but also face the limit of public enforcement in the digital age. Public actors do not only face potential lack of expertise and information asymmetry in certain cases, but also limited options to interfere with the flow of online information, which is governed by online platforms.[88] This imbalance has led to relying on internet shutdowns to fight disinformation in illiberal regimes,[89] and, as already underlined, to the criminalization of disinformation.

These examples demonstrate how law and regulation cannot be considered in isolation or as a means for governing, but it is also an achievement of governance.[90] Even if addressing online disinformation is insidious for constitutional democracies due to the risk of compromising freedom of expression and market freedoms, this challenge leads to developing a different strategy. The next section underlines how the European constitutional approach proposes a model for striking a balance between conflicting constitutional interests to address the spread of online disinformation.

C. The European Approach to Disinformation

The European strategy against disinformation has not been based on repressive narratives or the protection of absolute rights, but on striking a balance among conflicting constitutional rights and freedoms clashing in the disinformation arena. This approach has not resulted in the introduction of illiberal measures, such as the criminalization of fabricated content or the establishment of a neoliberal deadlock that allows online platforms to consolidate their private governance. Instead, it has led to the development of measures that seek to regulate the processes and dynamics that characterize the spread of disinformation.

Broadly, this result comes from the rise and consolidation of a renovated phase for European constitutionalism in the digital age.[91] In the case of disinformation, the central role of balancing in European constitutionalism has unveiled the compelling need to ensure the right to freedom of expression as a cornerstone of a democratic society, whose protection is not absolute rather limited by the need to protect other constitutional interests, including public legitimate interests, such as public health or security challenges. This approach has also led to stretch European competences to tackle disinformation,[92] moving the perspective from internal market goals to a constitutional-oriented strategy.

The European strategy also results from the limited competence of the Union in terms of content regulation, which preempted attempts to sanction disinformation, for instance, through criminal penalties.[93] Instead, this area is left in the hands of the Member States, as also underlined by the DSA,[94] which also contribute to shaping freedom of expression, and, more generally,

[88]Katharina Kausche & Moritz Weiss, *Platform Power and Regulatory Capture in Digital Governance*, Business and Politics 1 (2024).

[89]*See generally* Giovanni De Gregorio & Nicole Stremlau, *Internet Shutdowns and the Limits of Law*, 14 Int'l J. Commc'n 4224 (2020).

[90]*See generally* Lewis A. Kornhauser, *Law as an Achievement of Governance*, 47 J. Legal Phil. 1 (2022).

[91]*See generally* Giovanni De Gregorio, Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society (2022).

[92]*See generally* Judit Bayer, *The EU Policy on Disinformation: Aims and Legal Basis*, 16 J. Media L. 18 (2024).

[93]*See* Treaty on the Functioning of the European, 2012 O.J. (C 326) [hereinafter TFUE], at art. 83.

[94]*See* DSA Directive, *supra* note 17, at art. 3(h).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

fundamental rights, at the national level based on their constitutional identity.[95] The national dimension indeed is not only relevant for content regulation but also courts are primary actors in shaping the regulatory framework on disinformation, particularly considering the remedies introduced by the DSA.[96]

As underlined in the next sub-section, the constitutional strategy of the Union to address online disinformation has increasingly become sophisticated by mixing a hard and soft way. Indeed, the adoption of the DSA can be considered as a fundamental step to mitigate the risks for fundamental rights raised by online platforms to protect European democratic values.[97] It provides safeguards to increase transparency and accountability in the process of content moderation by, very large, online platforms based on risk regulation.[98] Likewise, the AI Act sets additional safeguards to increase the transparency of deepfakes and limit the use of AI systems for the purposes of manipulation.[99] This hard way of the European approach has been complemented by introducing additional regulatory solutions to cooperate with platforms in the fight against disinformation, as underlined by the adoption of the Strengthened Code.[100] This soft way to deal with the spread of disinformation introduces a system based on the creation of trust and collaboration among different stakeholders.

I. The Hard Way: Procedural Safeguards and Risk Regulation

The challenges brought by the private governance of online content and the spread of AI systems have led the Union to adopt legal instruments such as the DSA, the AI Act, the PAR and the EMFA. Despite not exclusively focusing on specific issues, such as false or fabricated content, these instruments play an important role in defining the European strategy against disinformation.

The DSA came with the goal of defining a new regulatory path for online platforms and content moderation in the digital age.[101] Among the main objectives, the DSA aims to ensure "a secure, predictable and trustworthy online environment,"[102] which inevitably involves tackling the spread of disinformation to achieve that goal. Although this European regulation maintains the rules of liability for online intermediaries, now established as the foundation of the digital economy and instrumental to the protection of fundamental rights, it aims to increase the level of transparency and accountability of online platforms to mitigate societal risks, including disinformation.[103] In other words, the DSA does not directly address illegal content but the actors dealing with that content, primarily online platforms. Indeed, the DSA does not define disinformation, not even illegal content.[104] That definition is left to Member States' competence, thus confirming the content-neutral spirit of the DSA, but not avoiding fragmentation among Member States.[105]

The DSA requires online platforms to take into account fundamental rights when enforcing their terms of service.[106] Particularly, it increases online platforms' responsibilities by requiring

[95] *See generally* Julian Scholtes, The Abuse of Constitutional Identity in the European Union (2023).

[96] *See* DSA Directive, *supra* note 17, at art. 54.

[97] *See generally* Giancarlo Frosio & Christophe Geiger, *Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime*, 29 Eur. L. J. 31 (2023).

[98] *See* DSA Directive, *supra* note 17, at arts. 34, 42.

[99] See AI Act, *supra* note 19, at art. 5.

[100] *See* 2022 Strengthened Code of Practice on Disinformation, *supra* note 18.

[101] *See* Martin Husovec, Principles of the Digital Services Act (2024); Caroline Cauffman & Catalina Goanta, *A New Order: The Digital Services Act and Consumer Protection*, 12 Eur. J. Risk Regul. 758 (2021).

[102] *See* DSA Directive, *supra* note 17, at art. 1(1).

[103] *See id.* at ¶ 9.

[104] *See id.* at art. 3(h).

[105] *See generally* Ronan Ó Fathaigh, Natali Helberger & Naomi Appelman, *The Perils of Legally Defining Disinformation*, 10 Internet Pol'y Rev. 2 (2021).

[106] *See* DSA Directive, *supra* note 17, at art. 14.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

them to act in a diligent, objective, and proportionate manner in the implementation of their terms of services, which includes the rights and legitimate interests of all parties involved, and is not limited to freedom of expression or freedom and pluralism of the media, but also other fundamental rights and freedoms as enshrined in the Charter.[107] On the one hand, this approach limits the discretion of online platforms in enforcing their terms of services. On the other hand, the DSA indirectly recognizes a critical role of these actors in enforcing their standards to moderate content, which includes platforms' rules to define and moderate disinformation.[108]

A variety of the DSA provisions precisely march in the direction of limiting the discretion of platforms in governing their services by introducing substantive and procedural safeguards.[109] For instance, the DSA proceduralizes the process of notice-and-takedown—now notice-and-action[110]—while also requiring platforms to provide a statement of reason when removing content,[111] thus providing more context to understand how disinformation is moderated in their online spaces. Furthermore, it requires online platforms to take the necessary technical and organizational measures to ensure that notices submitted by trusted flaggers are processed and decided upon with priority and without delay.[112] This system facilitates greater involvement by fact-checkers and other civil society organisations in the process of content moderation and the reporting of online disinformation. The implementation of such technological due process would also have implications for substantive rights, as it should preserve,[113] values such as accuracy; the appearance of fairness; equality of inputs; predictability, transparency, and rationality; participation; revelation; and privacy-dignity.[114]

This approach also extends to other areas that are critical for disinformation, including targeted advertising. The DSA, recognizes that advertising systems used by very large online platforms pose particular risks, for instance, relating to the spread of disinformation which could impact on public health, public security, civil discourse, political participation and equality.[115] As a result, the DSA requires very large online platforms to compile and make publicly available in a specific section of their online interface a repository containing various information, including the nature of the advertiser, and keep this information online for at least one year after the advertisement was presented for the last time on their online interfaces.[116]

Likewise, the DSA proceduralizes the process of crisis in cases of extraordinary circumstances affecting public security or public health.[117] In these cases, the Commission has the power to rely on crisis protocols to coordinate a rapid, collective and cross-border response, especially when online platforms are misused for the rapid spread of illegal content or disinformation or where the need arises for rapid dissemination of reliable information.[118] In these cases, very large online platforms are required to adopt these protocols, although these are to be applied only temporarily and should not lead platforms to a general monitoring obligation of online content.

Procedural safeguards do not exhaust the obligations introduced by the DSA. With respect to very large online platforms these actors are required to conduct a risk assessment at least once a

[107]*See* João Pedro Quintais, Naomi Appelman & Ronan Ó Fathaigh, *Using Terms and Conditions to Apply Fundamental Rights to Content Moderation*, 24 GERMAN L. J. 881, 894 (2023).

[108]*See* DSA Directive, *supra* note 17, at art. 3.

[109]*See* Martin Husovec, *Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation of Global Internet Rules*, 38 BERKLEY TECH. L. J. 883, 917–19 (2023).

[110]*See* DSA Directive, *supra* note 17, at art. 14.

[111]*Id.* at art. 15.

[112]*Id.* at art. 19.

[113]*See generally* Danielle Keats Citron, *Technological Due Process*, 85 WASH. UNIV. L. REV. 1249, (2008).

[114]*See* Martin H. Redish & Lawrence C. Marshall, *Adjudicatory Independence and the Values of Procedural Due Process*, 95 YALE L. J. 455, 483–91 (1986).

[115]*See* DSA Directive, *supra* note 17, at para. 81.

[116]*Id.* at art. 39.

[117]*Id.* at art. 36.

[118]*See id.* at art. 37.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

year about any significant systemic risks, including disinformation, stemming from the functioning and use made of their services in the EU,[119] while putting in place reasonable, proportionate, and effective mitigation measures.[120] Likewise, very large online platforms have to include in their terms and conditions the parameters used by recommender systems in a clear, accessible, and easily comprehensible manner.[121]

The move to risk regulation leads the DSA, and, broadly the European strategy on disinformation, towards a more flexible system of enforcement. In Europe, risk regulation has broadly expanded from environmental law and food safety,[122] to digital policies, as also underlined not only by the DSA but also by the AI Act.[123] As an attempt to deal with "risk society,"[124] through a rational and technocratic approach that fosters more efficient, objective, and fair governance,[125] risk regulation tends to reject "over-regulation, legalistic and prescriptive rules, and the high costs of regulation."[126] Risk-based regulation prioritizes and targets enforcement action considering the actual hazard, thus leading to contextual enforcement of the law based on concrete risk scores.[127] As a result, considering disinformation as a risk requires online platforms to act properly in order to mitigate the spread of false content.

Nonetheless, the AI Act adopts a different model of risk regulation.[128] Unlike the DSA, the AI Act expressly defines different thresholds of risk without limited discretion for public and private actors. It subjects the uses of AI systems to increasingly stringent obligations, if not banning certain uses.[129] The AI Act bans the possibility of using AI systems that aim to manipulate individuals and also introduces obligations for AI systems which are considered "high-risk." This includes "systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda,"[130] and excludes the systems used to "organise, optimise or structure political campaigns from an administrative or logistical point of view."[131]

Additionally, deepfakes are also subject to transparency obligations.[132] Providers of AI systems, including generative AI models capable of generating synthetic audio, image, video or textual content, have to ensure that such content is marked up in a machine-readable format and detectable as artificially generated or manipulated.[133] Furthermore, those deploying of AI systems capable of producing audio or video constituting deepfakes must disclose that such content has been artificially generated or manipulated,[134] including when the purpose is of informing the public on matters of public interest.

[119] *See* DSA Directive, *supra* note 17, at art. 34.
[120] *Id.* at art. 35.
[121] *Id.* at art. 38.
[122] *See generally* Hans-Wolfgang Micklitz & Takis Tridimas, Risk and EU Law (2015).
[123] *See generally* AI Act, *supra* note 19.
[124] *See generally* Ulrich Beck, Risk Society: Towards a New Modernity (1992).
[125] *See generally* Bridget Hutter, *A Risk Regulation Perspective on Regulatory Excellence*, *in* Achieving Regulatory Excellence 101 (C. Coglianese ed., 2016).
[126] *See* Milda Maceinate, *The "Riskification" of European Data Protection Law through a Two-Fold Shift*, 8 Eur. J. Risk Regul. 506, 509 (2017); Julia Black, *The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom*, 10 Pub. L. 510 (2005).
[127] *See* Claudia Quelle, *Enhancing Compliance Under the General Data Protection Regulation: The Risky Upshot of the Accountability- and Risk-Based Approach*, 9 Eur. J. Risk Regul. 502, 510 (2018).
[128] *See* Giovanni De Gregorio & Pietro Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, 59 Common Mkt. L. Rev. 473 (2022).
[129] *See* DSA Directive, *supra* note 17, at art. 5.
[130] *See* AI Act, *supra* note 19, at annex III, at. (8)(b).
[131] *See* AI Act, *supra* note 19, at annex III, at. (8)(b).
[132] *Id.* at art. 50.
[133] *See* AI Act, *supra* note 19, at art. 50.
[134] *See* AI Act, *supra* note 19, at art. 49 (5).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

This risk-based approach also requires deployers, which are public actors or private entities providing public services, to perform an assessment of the impact on fundamental rights that the use of such a system may produce.[135] This assessment, named Fundamental Rights Impact Assessment (FRIA), defines another critical step in assessing the risks coming from the use of AI systems which could be used to spread disinformation. Although this obligation does not apply to private actors, including online platforms, which are subject to a risk assessment based on the DSA, it defines another example of the European approach towards the accountability of the actors using AI systems.

Furthermore, the Union has also focused on the connection between political advertising and disinformation.[136] As underlined in the PAR Recitals, "[p]olitical advertising can be a vector of disinformation, in particular where the advertising does not disclose its political nature, comes from sponsors outside of the Union or is subject to targeting techniques or ad-delivery techniques" and "[a] high level of transparency is necessary inter alia to support open and fair political debate and political campaigns, and free and fair elections or referendums, and to counter information manipulation and interference, as well as unlawful interference, including from third countries."[137]

Among the different safeguards, the PAR introduces specific requirements related to targeting and ad-delivery techniques for online political advertising.[138] Primarily, political targeted advertising is restricted to personal data collected directly by the data controller, it requires data subjects' explicit consent for this purpose and cannot involve profiling as defined in the General Data Protection Regulation (GDPR).[139] Additionally, the PAR mandates additional transparency obligations,[140] such as keeping the records on the use of such techniques, the relevant mechanisms, and parameters used, and providing additional information necessary to allow individuals to understand the logic and parameters of targeting techniques, including whether AI has been used.

Likewise, the EMFA can be considered part of this European strategy.[141] Particularly, the Union has considered media pluralism as a way to tackle disinformation, while recognising the role of online platforms as "gateways to media content, with business models that tend to disintermediate access to media services and amplify polarising content and disinformation."[142] The goal of the EMFA is also to advance and gives priority to media content in order to ensure that quality information plays a role even in digital spaces.

In this case, the EMFA introduces a specific regime for content of media service providers on very large online platforms as defined in the DSA. Particularly, it restrict the possibility for online platforms to moderate content from media service providers which have been self-certified based on the system established by the EMFA. Indeed, the suspension or the limitation of the visibility in relation to content from a media service provider due to a violation of its terms and conditions triggers an obligation to inform the provider of the reasons for this decision, and the media service provider must then be given a chance to respond within 24 hours or, in urgent situations, within a shorter timeframe.

[135]*Id.* at art. 27.

[136]See *generally* Max Zeno van Drunen, Natalie Helberger & Ronan Ó Fathaigh, *The beginning of EU political advertising law: unifying democratic visions through the internal market*, 30(2) Int'l J. L. & Info. Tech. 181 (2022).

[137]Parliament and Council Regulation 2024/900 of March 13, 2024, Transparency and Targeting of Political Advertising 2024 O.J. (L 2024/900), recital 4 [hereinafter PAR].

[138]*Id.* at art. 18.

[139]Parliament and Council Regulation 2016/679 of April 27, 2016, Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) 2016 O.J. (L 119), art. 4(4).

[140]PAR, *supra* note 137, at art. 19.

[141]*See generally* Elda Brogi et al. for the European Parliament's Committee on Civil Liberties, Justice and Home Affairs, The European Media Freedom Act: Media freedom, Freedom of Expression and Pluralism (2023); Mark D. Cole & C. Etteldorf, *The European Media Freedom Act Unpacked*, European Audiovisual Observatory (2024).

[142]EMFA, *supra* note 21, at recital 4.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

This extensive hard way to tackle disinformation also leaves space for adopting codes of conduct to complement its structure.[143] By following the model adopted by the DSA, as further discussed in the next sub-section in the case of the Strengthened Code, the Union introduces the possibility of developing co-regulatory instruments which could also introduce additional rules to address the spread of disinformation.

II. The Soft Way: Co-Regulation and Trust

The European strategy to tackle disinformation has not always been based on an extensive set of norms and regulations. The development of the hard way has been mostly the result of the failure of the self-regulatory approach adopted by the Commission as reflected in the Code of Practice on Disinformation adopted in 2018.[144]

The European decision to follow the path of self-regulation in 2018, still primarily rooted in the metaphor of the free marketplace of ideas, led the European Commission to delegate the writing of the first Code of Practice on Disinformation to online platforms. When, at that time, the European Commission decided to adopt a strategy to deal with disinformation, the prevailing discourse was precisely oriented towards the importation of the metaphor of the internet as the new marketplace of ideas from the humus of U.S. constitutionalism as also demonstrated by the internal market focus of the e-Commerce Directive.[145] Therefore, the European approach was guided by blind faith in the self-corrective capacity of the market to bring out the truth through a free competition of ideas and opinions—even false ones—or in any case, to isolate disinformation without the need for any intervention from public institutions. This ideology was translated into a self-regulatory mechanism as the only viable policy option available, thus making the internal market the primary point of reference.

This first attempt, which also represented a unicum worldwide as a model of voluntary commitment by online platforms to adopt a whole series of measures that contained the phenomenon, was disappointing in terms of the vagueness of the obligations assumed by the platforms themselves. There was also an almost complete absence of criteria for verifiability and measurability of the commitments, as underlined by the Sounding Board on the Multistakeholder Forum on Disinformation.[146] In particular, the 2018 version did not provide the basic conditions for making the Code of Practice an effective tool to combat disinformation, particularly considering the absence of objectives and guidelines defined by the Commission and tools for assessing the measures agreed by the signatories.

This approach was anything but in harmony with European constitutional traditions and with the level of competitiveness of the pluralism of public debate that has characterized the continental model. The metaphor of the free digital marketplace of ideas fits uneasily with the European value system both with regards to the role played by freedom of expression, and with reference to the concept of abuse of right and to attention to the passive profile of the right to be informed if not in a truthful way. When, in 2018, the metaphor was forcibly imported by the European Commission into the old continent in order to develop the first European strategy against disinformation, the economic structure that characterized the digital environment was and is still far from free due to

[143] *Id.* at art. 50(7).

[144] Commission Code of Practice on Disinformation (2018), https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation.

[145] *See* Directive 2000/31, of the European Parliament and of the Council of June 8, 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (Directive on electronic commerce), 2000 O.J. (L 178).

[146] *See Sounding Board of Forum on Disinformation Online Issues Unanimous Opinion on So-Called Code of Practice*, EUR. BROAD. UNION (Sept. 26, 2018), https://www.ebu.ch/news/2018/09/sounding-board-of-forum-on-disinformation-online-issues-unanimous-opinion-on-so-called-code-of-practice.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

the consolidation of private governance in digital spaces.[147] Nonetheless, the constitutional traditions of the Member States were not immune to the import of the free marketplace of ideas metaphor. As already stressed, some proactive national approaches, including Germany and France, have anticipated the European regulatory measures to address the spread of disinformation, thus rejecting a self-regulatory approach.

The increasing challenges raised by online disinformation and the risk of fragmentation at the national level have then encouraged the European Commission to issue new guidelines and call for the writing of a new code that could fill the gaps of the first attempt and provide a more effective tool to counter disinformation,[148] also in coordination with the DSA. The Union has not abandoned a soft way to deal with online disinformation. Compared to 2018, the Strengthened Code Preamble refers to the European constitutional traditions and to the fundamental role of the Charter of Fundamental Rights, thus giving center stage to fundamental rights, and their balancing, in contrast to disinformation.[149] This approach is aligned with the expansion of constitutional values in European digital policies and underlines how the Union is putting the accent on the protection of rights and freedoms as an overarching goal in the digital age.[150]

The Strengthened Code aims to enhance transparency to combat online disinformation, particularly focusing on content monetization mechanisms within social media business models, which can incentivize the spread of disinformation.[151] Despite the limited focus on AI, it emphasizes the fight against disinformation professionals to reduce the financial incentives for polluting public discourse.[152] Regarding political advertising, the Strengthened Code calls for clear identification and labelling of such content on the web, especially during sensitive times, like elections.[153] It also places a greater emphasis on improving security and countering hidden disinformation tactics and procedures, while empowering users by providing tools and risk mitigation measures to combat disinformation, complementing the DSA. Additionally, the Strengthened Code emphasizes granting researchers access to data for in-depth studies on online disinformation while complying with GDPR regulations, addressing the previous limitations in data access from online platforms.[154]

The enlarged content of the Strengthened Code already shows a different commitment of all the stakeholders to this policy objective. This result has been possible in part due to considering the diversity of the signatories, which extended not only to online platforms but also to other stakeholders: Representatives of civil society; the community of fact-checkers and advertising companies; the European Regulators Group for Audiovisual Media Services (ERGA); and the European Digital Media Observatory (EDMO).[155] The Commission focused on defining a response to disinformation based on European constitutional values which resulted in a more balanced and dialogue-based mechanism of collaboration and trust between signatories and the Commission.[156] The Digital Services Act contributes to make codes of conduct, including potentially the Strengthened Code, tools of co-regulation and risk-assessment, thus an instrument to go beyond the dilemma between self-regulation and hard law.

[147]Pollicino, *supra* note 66.

[148]*See Commission Communication Guidance on Strengthening the Code of Practice on Disinformation*, COM (2021) 262 final (May 26, 2021).

[149]*See* 2022 Strengthened Code of Practice on Disinformation, *supra* note 18, at 1.

[150]*See Commission Declaration on Digital Rights and Principles for the Digital Decade*, COM (2022) 28 final (Jan. 26, 2022).

[151]*See generally* Matteo Monti, *Lo Strengthened Code of Practice on Disinformation: Un'altra Pietra Della Nuova Fortezza Digitale Europea?*, 2 Rivista di diritto dei media (2022).

[152]*See generally* Gregorio & Dunn, *supra* note 128.

[153]Monti, *supra* note 151.

[154]*Id.*

[155]*See* 2022 Strengthened Code of Practice on Disinformation, *supra* note 18, at 1.

[156]*Id.* at 29.

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

The guidelines of the Commission set the red line defining a limit to safeguard the protection of rights and indicating proposals for measuring and monitoring the objectives of the Strengthened Code. One added value consists of a much more massive and detailed presence of performance indicators relating to the effectiveness–and therefore measurability–of the commitments agreed by the signatories, almost completely absent in 2018, thus defining critical steps to verify whether and how the commitments undertaken by the signatories are then translated into concrete actions.[157]

This new architecture to fight disinformation has also been fostered by establishing a Transparency Centre and a Task Force which are essential bodies reflecting the dynamic identity of the Strengthened Code as a work in progress.[158] They aim to support the effectiveness of the commitments taken by online platforms and their general implementation. The Task Force also contributes to the definition of Structural Indicators as diagonal measures that allow the general measurement of the objectives, as also supported by the EDMO and the ERGA. This part plays a fundamental role in ensuring that the Strengthened Code is a living instrument and can be adapted to the challenges of disinformation and efficiently contributes to countering it.[159]

The Strengthened Code represents the soft way of the European strategy to disinformation that aims to regulate the ecosystem driving the spread of disinformation. This emphasis on the dynamics of disinformation, as opposed to the content itself, gives rise to an enforcement system based on the collaboration of public and private actors.

D. Towards a New Policy Framework

The Union has advanced a regulatory model to fight disinformation which is unique on a global scale. Rather than approaching disinformation through self-regulation or oppressive measures, the European strategy tends to balance the need to ensure speech in a democratic society, on the one hand, and provide an answer to the risks for democratic values deriving from the mix of fabricated content and private ordering of online content, on the other hand. This approach indeed breaks the deadlock by rejecting the liberal narratives that have driven self-regulatory solutions, and by regulating the dynamics of disinformation rather than its content.

Broadly, this strategy defines a new approach to address online disinformation. It does not represent a simple shift from self-regulation to hard regulation but underlines a different balance between public and private actors. When looking at its shape, the new European regulatory framework to address disinformation does not only consider the protection of conflicting constitutional interests, primarily rights, and freedoms, but also the connections between public and private actors. The Union has indeed recognized the critical role of private actors in overcoming the limitation of a system which is exclusively based on public enforcement.[160] The need to develop flexible approaches and collaborative structures to address the spread of false content comes from a mix of limited resources of public actors to effectively tackle disinformation, and the constitutional limits of their involvement in digital spaces.

The capacity of online platforms to implement measures to tackle disinformation has made private actors a critical fragment of a broader constitutional mosaic,[161] which does not only see public actors as sources of power. As already underlined, online platforms have also been proactive in developing policies and implementing actions to tackle the spread of disinformation. Considering their role in governing online spaces, online platforms have been in a privileged

[157] *Id.* at 30.

[158] *See Commission Declaration on Digital Rights and Principles for the Digital Decade*, COM (2022) 28 final (Jan. 26, 2022).

[159] *See* Iva Nenadic, Elda Brogi & Konrad Bleyer-Simon, *Structural Indicators to Assess Effectiveness of the EU's Code of Practice on Disinformation* 7 (Eur. Univ. Inst., Working Paper No. 34, 2023).

[160] Pollicino, *supra* note 66, at 255.

[161] *See generally* Gunther Teubner, Constitutional Fragments: Societal Constitutionalism and Globalization (2012).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

position of proxies to enforce public policy.[162] This role has been reinforced by the regulatory approach of the Union which recognizes a broader involvement of online platforms in the enforcement of public policies on disinformation while making these actors more accountable, as particularly underlined by the DSA.

The interdependency between public and private actors leads to a transformation of power relationships in digital policy. What the European approach to disinformation has made particularly relevant is the relationship of trust between regulators and stakeholders, which is not only based on a formal compliance mechanism but on accountability, collaboration and trust. Being aware of the need to rely on private actors and the perils of disproportionate regulatory measures, the Union has increasingly resorted to regulatory strategies meant to increase flexibility in European digital policy.[163] The European approach has indeed not followed a rigid command and control system but a flexible performance-based approaches,[164] as well as it moved from a rights-based approach to risk-based regulation.[165]

The risk-based approach introduced by the DSA can be considered a different way of regulating disinformation. Rather than imposing strict obligations, the Union aims to make platforms more accountable by delegating the risk assessment and the consequent risk mitigation measures to private actors while keeping control over their assessment.[166] This approach requires collaboration and trust between public and private actors. Although the implementation of risk mitigation measures to tackle disinformation is left to private actors, this process is still subject to the scrutiny of the Commission.[167] As a result, enforcing the measures to tackle disinformation leads to striking a balance between market freedoms, the protection of individual rights, and democratic values. Likewise, the AI Act contributes to increasing the responsibility of online platforms, as providers and deployers, for certain types of AI systems, as in the case of deep fakes, while keeping the enforcement in the hands of public actors.[168] Moreover, the AI Act expressly refers to the DSA to underline how it does not interfere with the obligations of risk assessment which apply to very large online platforms.[169]

However, the risk-based approach is only one part of the European way to address online disinformation. Both the DSA and the AI Act underline the role of codes of conduct in laying down a system of dialogue between public and private actors. Particularly, the Strengthened Code complements the approach of the Union followed by supporting a cooperative regulatory regime.[170] This approach symbolizes a first attempt towards a co-regulation model that supports the European soft way against disinformation. It broadly highlights the limits of top-down enforcement strategies to deal with the predominance of self-regulatory solutions to address the spread of false content in the digital age.

The DSA contributes to this transition by shedding light on the still voluntary nature of codes of conduct, but underlining the role of co-regulation as a way to define measures to address harmful content such as disinformation. In this case, codes of conduct aim to play an important role in tackling the amplification of false news through bots and fake accounts and may be

[162]*See generally* Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power (2019).

[163]*See generally* Raphaël Gellert, The Risk-Based Approach to Data Protection (2020); Zohar Efroni, *The Digital Services Act: Risk-Based Regulation of Online Platforms*, Internet Pol'y Rev. (Nov. 16, 2021), https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606.

[164]*See* Cary Coglianese, *The Limits of Performance-Based Regulation*, Univ. Mich. J. L. Reform 525, 527 (2017).

[165]*See generally* Julia Black & Robert Baldwin, *When Risk-Based Regulation Aims Low: Approaches and Challenges*, 6 Regul. & Governance 2 (2012).

[166]*See* DSA Directive, *supra* note 17, at 37, 42.

[167]*See* DSA Directive, *supra* note 17, at 66, 67.

[168]*See* AI Act, *supra* note 18, at art. 52.

[169]*Id.* at art. 2.

[170]*See generally* Christopher T. Marsden, Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace (2011).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

considered an appropriate risk mitigation measure by very large online platforms.[171] The DSA recognizes to the Commission, and the European Board for Digital Services, the role of encouraging and facilitating the drawing up of voluntary codes of conduct, taking into account in particular the specific challenges of tackling different types of illegal content and systemic risks.[172] Furthermore, according to the DSA, the refusal to participate in this process without proper explanations by a very large, online platform can be taken into account by the Commission when assessing whether online platforms have infringed the obligations introduced by the DSA.[173] Even if participation in the Strengthened Code does not ensure compliance, this system increases the accountability of online platforms in addressing disinformation and reduces their discretion in content moderation. Consequently, codes of conduct are not merely self-regulatory instruments. Rather, they are co-regulatory tools that derive from the agreement between public and private actors.

However, the scope of Strengthened Code could be challenged by the expansion of European policies addressing online platforms and content moderation.[174] Some parts of the Strengthened Code tend to overlap with legal obligations which have been introduced by European legislation after its adoption. For instance, access to data owned by online platforms for research purposes in the Strengthened Code overlaps with the legal framework introduced by the DSA.[175] Likewise, the rules introduced by the PAR are likely to meet the obligation which will be introduced by the regulation on transparency of political advertising.[176]

Despite the overlaps, codes of conduct provide another example of how the Union is designing a third way to tackle disinformation. Rather than relying on top-down rigid strategies, such a mechanism promotes policymaking and enforcement processes where private actors actively participate in conversation with public actors. These processes contribute to making public actors closer to their goal of enforcing their policy in digital spaces and, particularly, mitigating the spread of disinformation. Furthermore, the participation and agreement of rules between public and private actors also increase the reactiveness of private actors to implement measures to address disinformation and the acceptance of potential sanctions. Indeed, more dialogue with regulators in the enforcement phase would help to mitigate reactive, and potentially disproportionate, measures as underlined, for instance, by the temporary suspension of ChatGPT by the Italian Data Protection Authority.[177]

However, this approach also raises constitutional challenges. The extensive reliance on risk regulation and co-regulation to address disinformation can lead to questions of accountability and transparency in decision-making and enforcement. Indeed, the broad and often vague nature of risk-based obligations may lead to creating a governance model where private entities wield regulatory power while facing legal uncertainty. Likewise, the collaborative nature of co-regulatory arrangements may blur the lines between public and private actors, thus leading to collaboration and regulatory capture limiting accountability. Therefore, the success of the European approach to disinformation also depends on keeping this regulatory framework open to different stakeholders and ensuring accountability.

[171] *See* Council Directive 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), 2022 O.J. (L 277) ¶ 104.

[172] *See id.* at art. 45.

[173] *Id.*

[174] *See* Elda Brogi & Giovanni De Gregorio, *From the Code of Practice to the Code of Conduct? Navigating the Future Challenges of Disinformation Regulation*, 16 J. MEDIA L. 38, 44 (2024).

[175] *See* Council Directive 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), 2022 O.J. (L 277), art. 40.

[176] *See Proposal for a Regulation of the European Parliament and of the Council on the Transparency and Targeting of Political Advertising*, COM (2021) 731 final (Nov. 25, 2021).

[177] *See* Garante Per La Protezione Dei Dati Personali [Italian Data Protection Authority] 30 marzo 2023, web doc. n. 9870832 (It).

https://doi.org/10.1017/glj.2025.24 Published online by Cambridge University Press

These challenges underline the complexity of the European strategy, showing the Union's intention to tackle disinformation without disproportionately regulating its content or stretching the EU's legal bases to achieve this purpose. This approach defined by the hard and soft way leads to the definition of a new balance between public and private actors to achieve public policy objectives. Rather than implementing restrictive measures or supporting self-regulation, the European approach defines a different way to develop to limit the risks coming from online disinformation.

E. Conclusions

The spread of online disinformation has raised constitutional questions. The strategies to address the increasing flowing of false and fabricated content pushed by the expansion of AI systems have revealed how disinformation is not only a matter of protecting freedom of expression but leads to focus on a broader set of conflicting constitutional values. Across constitutional systems, the answers to the spread of disinformation are diverse and show different sensitivity to the role of freedom of expression in a democratic society. Regardless of either repressive measures or a liberal approach to digital spaces, online disinformation has been considered a risk and, at the same time, a necessary part of the public discourse, even if the threats raised by disinformation driven by AI technologies have been shaping regulatory conversations.

The different answers to disinformation have also been driven by the predominance of online platforms in moderating online content and, broadly, governing digital spaces. The role of online platforms has indeed raised questions about the marketplace of ideas metaphor. Although the notion of the internet as a new free marketplace of ideas persists, the reality is far more complicated, characterized by economic concentration and the influence of a few powerful private actors governing online speech.

This situation has profoundly impacted the European policy on disinformation. The introduction of the DSA has been a landmark example of the European hard way which does not only aim to limit the discretion of online platforms in making decisions on freedom of expression, but also to address disinformation. Likewise, the Strengthened Code has encouraged more dialogue between public and private actors to define additional measures for the same purpose, thus striking a balance among conflicting constitutional interests. This approach leads to a model where the enforcement of public policies does not only come from private determinations or regulatory interventions but relies on regulatory instruments of collaboration between public and private actors, which, however, can also lead to other constitutional concerns, particularly related to transparency and accountability.

The primary challenge to address online disinformation is to find a balanced outcome among conflicting constitutional interests. Between leaving the market free to shape disinformation and relying on oppressive measures to tackle the spread of disinformation, the European approach underlines that the fight against disinformation cannot be based solely on self-regulation or a top-down strategy, but requires the establishment of a relationship of trust and cooperation between public and private actors in the context of regulatory measures. Although this approach requires further safeguards to limit the risks of collaborative arrangements, potentially escaping accountability and transparency, it defines a different constitutional way to tackle disinformation.

Competing Interests. The authors declare none.

Funding Statement. This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

Cite this article: De Gregorio G, Pollicino O (2025). The European Constitutional Way to Address Disinformation in the Age of Artificial Intelligence. *German Law Journal*, 1–22. https://doi.org/10.1017/glj.2025.24