Speech without a Speaker: Constitutional Coverage for Generative AI Output?

Marco Bassini*

*Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, Tilburg, Netherlands, email: m.bassini@tilburguniversity.edu

Generative AI systems' output as speech – Constitutional coverage for AI speech in the absence of a (human) speaker – Right of individuals to receive information as a perspective for framing constitutional coverage of generative AI output – Implications of constitutional coverage for content policing and content moderation by private platforms – Trends in the interpretation of existing content moderation regimes and their applicability to generative AI systems.

Introduction

Is the output produced by systems like ChatGPT constitutionally protected speech? If so, does the existence of coverage impact the ability of lawmakers to engage in content regulation? To understand and contextualise these apparently pointless, yet intriguing questions, it is necessary to take a step back to consider them in a broader context.

The large-scale advent of generative artificial intelligence[1] systems has intensified the ongoing debate about the need for future-proof AI regulation in Europe, which recently culminated in the entry into force of the EU AI Act[2]

[1]For an overview, *see* B. van der Sloot, *Regulating the Synthetic Society. Generative AI, Legal Questions, and Societal Challenges* (Bloomsbury 2023).

[2]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

European Constitutional Law Review, 21: 375–411, 2025
© The Author(s), 2025. Published by Cambridge University Press on behalf of University of Amsterdam. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.
doi:10.1017/S1574019625100771

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

and the adoption of the Council of Europe Framework Convention on Artificial Intelligence.[3]

The AI Act is the first comprehensive piece of legislation to regulate AI systems at a global level. With its adoption, the EU aims to capitalise on its 'first-mover' advantage over more technologically advanced superpowers, such as the US and China, in an effort to pave the way for a 'Brussels effect' to materialise.[4] However, recent developments in US politics and the emergence of DeepSeek have put the rights-based EU regulatory model under pressure.[5]

Months ago, the 'ChatGPT revolution' had a remarkable and – to a certain extent – troublesome impact on the legislative process for the AI Act, thereby illustrating the difficulty for regulation to keep pace with technological progress.[6] Generative AI did not specifically fall within the scope of the Commission's AI Act proposal, and its rise sparked debate among the co-legislators over whether *ad hoc* provisions should be included to cover it. The texts initially voted upon by the European Parliament (in June 2023) and the Council (in December 2022) prior to the trilogue negotiations were an attempt to fill this gap with *ad hoc* rules, thereby reflecting the concerns posed by such an unprecedented and disruptive technology.[7]

One of the most significant concerns associated with the proliferation of generative AI is its potential to spread inaccurate content and disinformation, which is often linked to political propaganda. Over the last decade, this issue has become a major concern for lawmakers, and it has garnered increasing attention, in particular, in the EU's policy agenda.[8] In fact, the use of large language models,

[3]Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Vilnius, 5.IX.2024. For a comparison between the approach of the EU and that of the Council of Europe, *see* F.P. Levantino and F. Paolucci, 'Advancing the Protection of Fundamental Rights through AI Regulation: How the EU and the Council of Europe are Shaping the Future', in P. Czech et al. (eds.), *European Yearbook on Human Rights 2024* (Brill 2025) p. 3.

[4]A. Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020).

[5]For an in-depth overview of the different regulatory approaches in the context of technology, *see* A. Bradford, *Digital Empires* (Oxford University Press 2023), who compares a right-based European approach to a market-driven US model and to a state-driven Chinese model.

[6]For a commentary on the Commission proposal, *see* M. Veale and F. Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act – Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach', 22(4) *Computer Law Review International* (2021) p. 97.

[7]*See* P. Hacker et al., 'Regulating ChatGPT and Other Large Generative AI Models', FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability and Transparency, June 2023, p. 1112.

[8]For an in-depth comparative overview of the most significant legislative and policy actions in this field, not limited to Europe, *see* O. Pollicino, *Freedom of Speech and the Regulation of Fake News* (Intersentia 2023). As regards the EU, the first steps to develop a strategy to counter disinformation date back to the Communication from the Commission to the European Parliament, the Council,

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

such as ChatGPT, can result in the propagation of information that despite being erroneous, appears to be accurate (so-called 'hallucinations'). While disinformation does not necessarily constitute illegal content, the output created by generative AI may also affect legal interests, such as in the case of defamatory statements that harm an individual's reputation.

As the example of disinformation illustrates, digital artifacts – much like content disseminated on social media – can have a significant impact on the public sphere and, by extension, on democracy. Interestingly, scholars and regulators appear more concerned with AI's potential to fuel disinformation[9] than with its ability to produce illegal content.

The similar ability of generative AI and social media to facilitate the spread of both illegal and harmful content highlights the need for an investigation into existing content moderation regimes. For example, this would require answers to questions such as: Who is liable for inaccurate content generated by large language models such as ChatGPT? Can these systems actually engage in defamatory conduct? Can their providers or deployers or perhaps even users face consequences under these circumstances? However, there is a fundamental, preliminary question that a constitutional law analysis should address: What is the constitutional status of the content created by AI systems?

The response to this question bears substantial implications for both the ability of legislators to regulate generative AI and that of providers and/or deployers to engage in content moderation, including moderation of illegal as well as merely harmful content.

The goal of this essay is to explore the relationship between generative AI and freedom of expression, with a view to determining whether output such as texts, audios, videos or images is eligible for constitutional coverage and protection. While disinformation is not a subject of investigation in this essay, it serves as a

the European Economic and Social Committee and the Committee of the Regions, 'Tackling online disinformation: a European Approach', COM/2018/236 final, April 2018. This communication was preceded by the Report of the Independent High-level Group on fake news and online disinformation, 'A Multi-dimensional Approach to Disinformation', released in March 2018, which provided a definition of disinformation as 'false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit'. *See also* A. Koltay, 'Freedom of Expression and the Regulation of Disinformation in the European Union', in R.J. Krotoszynski et al. (eds.), *Disinformation, Misinformation, and Democracy* (Cambridge University Press 2025) p. 133; A. Peukert, 'The Regulation of Disinformation in the EU – Overview and Open Questions', Research Paper of the Faculty of Law of Goethe University Frankfurt/M No. 2023. For an American perspective, *see* L.G. Jacobs, 'Freedom of Speech and Regulation of Fake News', 70 *The American Journal of Comparative Law* (2022) p. i279.

[9]For an overview, *see* the white paper by K. Bontcheva (ed.), 'Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities', published by the European Digital Media Observatory and co-funded by the EU projects TITAN, AI4Media, AI4Trust and vera.ai (2024).

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

paradigmatic example of the challenges inherent in moderating content that is not (necessarily) illegal and that may thus be able to claim constitutional protection.

This feature has key repercussions for both state authorities seeking to enforce existing restrictions on speech (without resulting in censorship) and service providers whose relationship with users is primarily governed by private terms of service.[10]

An understanding of whether digital artifacts are constitutionally protected also plays a key role in the applicable content policing regime and, ultimately, in determining the 'arms' to which the EU and other countries can resort to in the fight against illegal and harmful content. However, this article does not explore the role of users of AI systems, who may equally influence the content produced by generative AI when creating prompts, due to their more limited ability to shape the public sphere.

Exploring the relationship between generative AI and freedom of expression requires consideration of the diverging understanding of this constitutional right in Europe and the US, where it dramatically emerged in cyberspace. So far, it is primarily within US scholarship that the constitutional coverage of 'AI speech' has been explored, through the lens of the First Amendment.[11] This prompts another question: Could this remarkable difference in the essence of the right, which results in different constitutional coverage, also have an impact in the context of AI[12] and thus lead to diverging answers?

[10]As a result of the growing importance of online platforms as fora for the exchange of ideas and opinions, attempts have been made to enforce freedom of expression with horizontal effects in the relationship between service providers and their users, most notably in cases in which the latter had been subject to bans or suspended or limited in their ability to interact with other users. In the US, courts have considered the application of the public forum doctrine to social networks under specific circumstances, whereas they have excluded these platforms from qualification as state actors performing traditional, exclusive public functions. In Europe, there is likewise no consistent line of reasoning in the case law of national courts on the applicability of freedom of expression with horizontal effects in the digital sphere between service providers and users, despite the influence of the Drittwirkung doctrine. However, Art. 14(4) of the Digital Services Act (Regulation (EU) 2022/2065) has now 'operationalised' respect for fundamental rights and, particularly, of freedom of expression, in this relationship, requiring service providers to consider the impact on the rights 'of all parties involved' while imposing any restriction resulting from the enforcement of their terms and conditions. For a comparative overview of the case law, *see* M. Bassini, 'Social Networks as New Public Forums? Enforcing the Rule of Law in the Digital Environment', 1(2) *The Italian Review of International and Comparative Law* (2022) p. 311.

[11]More recently, *see* M. Kaminski and M.L. Jones, 'Constructing AI Speech', 133(1) *The Yale Law Journal Forum* (2024) p. 1212; M. Austin and M. Levy, 'Speech Certainty: Algorithmic Speech and the Limits of the First Amendment', 77 *Stanford Law Review* (2025) p. 1.

[12]For an overview of the different constitutional protection provided to free speech rights in the US and in Europe, *see* M. Rosenfeld and A. Sajó, 'Spreading Liberal Constitutionalism: An Inquiry into the Fate of Free Speech Rights in New Democracies', in S. Choudhry (ed.), *The Migration of Constitutional Ideas* (Cambridge University Press 2009) p. 142.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

This article will begin with a comparative overview of the coverage of freedom of expression in the US and in Europe in order to understand whether and how the respective legal systems can protect AI content as speech. It will then explore another domain that has key implications in this respect, namely that of content moderation and liability for illegal content. To complement its analysis of the 'arms' available to counter illegal and harmful content in the context of generative AI, the essay will question whether existing content moderation regimes for online platforms may be extended to AI operators (providers and/or deployers) and discuss whether such an option would prove beneficial overall and in alignment with the current policies of the US and Europe.

Generative AI output and constitutional coverage

'This must be the place': the role of freedom of expression in the debate on AI regulation

In the ongoing global debate on AI regulation,[13] freedom of expression stands out as a 'stone guest' of no small importance to the elephant in the room. There are various well-known reasons why speech may deserve protection:[14] its contribution to individual self-fulfilment, its key role in democracy or its function in advancing the search for truth. These correspond to the traditional positive theories that advocate for speech protection. As will be illustrated below, none of these justifications inherently exclude the viability of 'AI speakers'. However, a fourth, negative justification – rooted in the suspicion of government interference – best explains the importance of freedom of expression in an AI-driven society in which the public sphere is increasingly populated by digital artifacts. While the three positive theories may have limited potential when it comes to what constitutes 'machine speech', the negative justification based on suspicion of government control remains fully relevant regardless of whether the source of content is human or artificial.

Some of the issues that have gained prominence in the discussions around AI and its legal implications are strongly intertwined with freedom of expression. Determining the constitutional status of AI-generated content may not only help to decipher the impact on content policing but also solve questions regarding the

[13]To understand the roots of the debate on AI regulation, *see* N. Smuha, 'From a "Race to AI" to a "Race to AI Regulation": Regulatory Competition for Artificial Intelligence', 13(1) *Law, Innovation and Technology* (2021) p. 57. A more recent overview is provided by M. Hildebrandt, 'Artificial Intelligence Law', in J.M. Smits et al. (eds.), *Elgar Encyclopedia of Comparative Law* (Edward Elgar 2023) p. 139.

[14]*See* E. Barendt, *Freedom of Speech* (Oxford University Press 2005) p. 1.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

enforcement of the right to data protection vis-à-vis, for example, large language models, or the relationship of AI-generated content with copyright.

Copyright is perhaps the main subject of the current discussions revolving around generative AI regulation.[15] In the copyright domain, the main problems posed by AI concern both the use of copyrighted material in the input phase and the creation of potentially infringing content in the output phase.[16] Not by coincidence, some resounding lawsuits[17] concern the lawfulness of the use of content published in the exercise of the freedom of information by the press. Looking at these cases, one can clearly see that new business models 'competing' with traditional editorial actors have emerged.[18] This may raise concerns about the potential impact on the quality of professional information, which is essential for maintaining the democratic character of society. If individuals 'trust' ChatGPT or Gemini more than a professional source of information, this can have undesirable consequences (such as the spread of misinformation[19]) and ultimately challenge, or otherwise weaken, the well-established role of the press as a public watchdog.[20]

The common understanding of copyright protection faces substantial challenges in the generative AI era as it has historically been rooted in a human authorship paradigm. These challenges stand out as useful terms of comparison to

[15]For an overview, *see* A. Gaon, *The Future of Copyright in the Age of Artificial Intelligence* (Edward Elgar 2021); C. Geiger, 'Elaborating a Human Rights-Friendly Copyright Framework for Generative AI', 7 *International Review for Intellectual Property and Competition Law* (2024) p. 1129; J.P. Quintais, 'Generative AI, Copyright and the AI Act', 56 *Computer Law & Security Review* (2025).

[16]A. Guadamuz, 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs', 73(2) *GRUR International* (2024) p. 111.

[17]*See* 'The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work', *The New York Times*, 27 December 2023, https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html/, visited 3 June 2025.

[18]For an in-depth overview of these challenges, *see* L. Dutkiewicz et al., 'Artificial Intelligence and Media', in N. Smuha (ed.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (Cambridge University Press 2025) p. 283.

[19]Misinformation is commonly understood as the unintentional spread of inaccurate information, which is not necessarily supported by intention as in the case of disinformation. According to the European Commission Communication on the European Democracy Action Plan (i.e. 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 (3 December 2020) COM(2020)790 final), 'misinformation is false or misleading content shared without harmful intent though the effects can be still harmful, e.g. when people share false information with friends and family in good faith'.

[20]*See* ECtHR 25 March 1985, No. 8734/79, *Barthold* v *Germany*, para. 58; ECtHR 26 November 1991, No. 13166/87, *The Sunday Times* v *UK (No. 2)*, para. 50; ECtHR 27 March 1996, No. 17488/90, *Goodwin* v *UK*, para. 39; ECtHR 20 May 1999, No. 21980/93, *Bladet Tromsø and Stensaas* v *Norway*, para. 59; ECtHR 10 December 2007, No. 69698/01, *Stoll* v *Switzerland*, para. 154; ECtHR, 7 February 2012, No. 39954/08, *Axel Springer* v *Germany*, para. 79.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

freedom of expression: while in the domain of copyright it is necessary to look for a human who qualifies as an author to properly speak of authorship, protected speech can exist even in the absence of a (human) speaker. If the lack of a person to qualify as a speaker were a determinative factor in excluding constitutional coverage, the framing of the problem would most likely be incorrect. The key point that this article aims to explore is why AI-generated content can amount to speech, regardless of the presence of a 'speaker'.

Long before the advent of generative AI, scholars and courts were grappling with similar issues in the context of less-advanced technological developments. These efforts were focused on some forms of 'algorithmic speech',[21] such as search engine results in the US[22] and search engine autocomplete suggestions in Europe.[23]

However, it is only in the US that scholars have specifically addressed algorithmic speech. This may depend on the peculiar understanding of free speech rights inherent in the 'exceptionalism' of the First Amendment,[24] which is reflected in its wording.

In the US, an investigation into Google's search practices launched by the Federal Trade Commission in 2011[25] triggered a discussion around search engine results as protected speech. The investigation sought to determine the existence of

[21] *See* A.M. Sears, 'Algorithmic Speech and Freedom of Expression', 53(4) *Vanderbilt Journal of Transnational Law* (2020) p. 1327. The author distinguished four categories of algorithmic speech, namely: curated production (news stories and search engine results); interactive/responsive production (chat bots); semi-autonomous production (search engine autocomplete functions); and fully autonomous production. At the time, no existing example of fully autonomous production could be provided, as the category was understood to cover 'the scenario in which an algorithm produces speech fully independent of human intervention or input'. Whether AI-generated content is truly independent of human intervention or input is debatable, but Sears' categorisation of this type of content can alternatively fall within the 'semi-autonomous production' of speech, the key aspect of which lies in the ability to collect input from external sources ('to learn', one might say) and then produce output that is also unexpected from what the programmers intended.

[22] Whereas the US Supreme Court did not have the chance to develop a free speech scrutiny in the *Twitter, Inc.* v *Taamneh*, 598 US 471 and *Gonzalez* v *Google LLC*, 598 US 617 (2023) judgments, which only concerned the actual scope of the immunity provided by Section 230 CDA for third-party content recommended by service providers such as YouTube and X in their capacity as social networking platforms.

[23] As reported by Sears, *supra* n. 21, p. 1332, the only remarkable precedents on the matter include two diverging judgments from the French Supreme Court (decisions of 19 February 2013 and 19 June 2013) and a case by the German Federal Court of Justice (decision of 14 May 2013).

[24] *See* F. Schauer, 'The Exceptional First Amendment', in M. Ignatieff (ed.), *American Exceptionalism and Human Rights* (Princeton University Press 2005) p. 29.

[25] At an even earlier stage, two US federal courts had already considered free speech rights applicable to Google search engine service in the landmark *Search King* v *Google Technology, Inc.*, Case No. Civ-02-1457-M (W.D. Okla., Jan. 13, 2003) and *Landon* v *Google, Inc.*, 474 F. Supp. 2d 622 (D. Del. 2007) cases. For an in-depth comment on the FTC's investigation and these cases, *see* J. Grimmelmann, 'Speech Engines', 98 *Minnesota Law Review* (2014) p. 868; Sears, *supra* n. 21, p. 1339.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

alleged search engine bias due to Google's dominance in the search advertising market. In 2013, the Commission concluded its investigation and found no evidence of any manipulation of Google's algorithms to unfairly influence the display of its search results.[26]

Against this background, Volokh and Falk[27] have argued that search engines are akin to editors and, thus, are speakers that enjoy First Amendment protection,[28] despite the fact that they 'produce and deliver speech through a different technology than that traditionally used for newspapers and books'.[29]

Wu has opposed these views by appealing to a *de facto* 'functionality doctrine', which serves to differentiate (human) expressions covered by the First Amendment from merely functional machine communications.[30] Years before, Bracha and Pasquale had advanced a similar line of reasoning, arguing for a prevailing performative rather than propositional character in the communicative acts of search engines,[31] albeit 'having an undeniable expressive element';[32] quoting Robert Post, they noted that the speech of a search engine 'is not a form of social interaction that realizes First Amendment values'.[33] Between the 'naysayer' and the 'advocates for constitutional coverage', using Collins' and Skovers' terminology,[34] Grimmelmann has claimed that neither the editor nor the conduit model properly capture the actual role of a search engine, which is, in fact, that of an advisor.[35] Acting as advisors, search engines engage in socially valuable speech that receives First Amendment protection; but search results constitute protected speech because 'they are valuable instrumentally rather than expressively' and not because they should be categorically protected.[36] Only Benjamin went beyond the mere domain

[26]*See* Statement of the Federal Trade Commission Regarding Google's Search Practices In the Matter of Google Inc. FTC, File Number 111-0163 January 3, 2013.

[27]E. Volokh and D.M. Falk, 'Google First Amendment Protection for Search Engine Search Results', 8 *Journal of Law, Economics & Policy* (2012) p. 883.

[28]According to Volokh and Falk, this conclusion follows from three aspects inherent in the functioning of search engines: they sometimes provide information that they themselves have prepared or compiled; by directing users to third party websites on the basis of certain criteria, they report on the content of others, which itself constitutes protected speech; they select and sort the results in order to provide their users with the most helpful and useful information: ibid., p. 884.

[29]Ibid., p. 885.

[30]T. Wu, 'Machine Speech', 161 *University of Pennsylvania Law Review* (2013) p. 1495.

[31]O. Bracha and F. Pasquale, 'Federal Search Commission? Access, Fairness, and Accountability in the Law of Search', 93 *Cornell Law Review* (2008) p. 1149.

[32]Ibid., p. 1193.

[33]Ibid., p. 1194; *see* R. Post, 'Encryption. Source Code and the First Amendment', 15 *Berkeley Technology Law Journal* (2000) p. 713 at p. 716.

[34]R. Collins and D. Skover, *Robotics. Speech Rights & Artificial Intelligence* (Cambridge University Press 2018) p. 34-35.

[35]Grimmelmann, *supra* n. 25.

[36]Ibid., p. 912.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

of search engines to investigate First Amendment coverage for algorithm-based decisions. He noted that 'reliance on algorithms [never] transforms speech into non-speech'; by re-interpreting the *Spence* standard,[37] he indicated that the key condition for algorithmic output to qualify as speech is the transmission of a substantive message to a listener that can recognise it.[38]

The US Supreme Court only very recently had the chance to add its voice to the debate on the conferral of speech rights to social networking platforms while they are performing content moderation.[39] In *Moody* v *NetChoice* the Court found that when online platforms decide which third-party content their feeds will display or how the display will be ordered and organised, they engage in expressive choices and, therefore, receive First Amendment protection.[40] The judgment seems to imply that the exercise of editorial judgement that is part of platforms' free speech rights does not conflict with the immunity from liability for third-party content granted by Section 230 of the Communications Decency Act[41] to interactive service providers. However, the paradox inherent in conferring free speech rights to (business) actors that have historically claimed to be neutral conduits of third-party information is still visible, just as it was ten years ago.

This paradox is even more visible in Europe, where in 2001, the EU lawmakers[42] immunised service providers from liability for third-party content, on the assumption[43] that they engaged in the performance of an activity of 'a mere technical, automatic and passive nature, which implies that the information society service provider has neither knowledge of nor control over the information which is transmitted or stored'.[44] Therefore, the application of the safe harbours

[37] *Spence* v *Washington*, 418 US 405 (1974).

[38] S. Benjamin, 'Algorithms and Speech', 161 *University of Pennsylvania Law Review* (2013) p. 1445.

[39] *See Moody* v *NetChoice, LLC*, 603 US ___ (2024). For a comment, *see* E. Bietti, 'Online Speech at the US Supreme Court in Moody v. Netchoice', *Verfassungsblog*, 11 July 2024, https://verfassungsblog.de/online-speech-at-the-us-supreme-court-in-moody-v-netchoice/, visited 3 June 2025.

[40] *Moody* v *Netchoice, LLC*, *supra* n. 39, p. 26.

[41] 47 USC § 230. This provision prevents service providers from being treated as a publisher or a speaker with respect to any third-party content posted or transmitted via their services, thus immunising them from liability, including for engaging in content moderation in good faith. For a comprehensive analysis of Section 230 CDA and its historical roots and significance, *see* J. Kosseff, *The Twenty-Six Words That Created the Internet* (Cornell University Press 2019).

[42] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular, electronic commerce, in the Internal Market ('Directive on electronic commerce').

[43] *See* M. Bassini, 'Fundamental Rights and Private Enforcement in the Digital Age', 25 *European Law Journal* (2019) p. 182.

[44] *See* Recital 42, Directive 2000/31/EC. The problem is more visible in the EU, as in the US, Section 230 CDA prevents service providers from being treated as speakers or publishers regardless of the existence of any degree of control or editorial judgement. Therefore, it might also be argued

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

was made conditional upon the absence of any editorial control, which would have turned a mere service provider into a content provider.[45] This is perhaps one of the reasons commentators and regulators in Europe have thus far paid limited attention to algorithmic speech:[46] The conceptualisation of service providers as neutral intermediaries has precluded any reasoning regarding the existence of algorithmic speech,[47] and the idea itself of an editorial activity (implied in the creation of speech) would have conflicted with this paradigm.

While US courts have had the opportunity to test this concept in relation to search engine results and, more recently, social media content feeds, European courts have paid comparatively less attention to this debate. It is only in the context of search engine autocomplete suggestions that a few domestic courts have (sometimes incidentally) questioned whether such output qualifies for constitutional coverage as speech. In this domain, a possible counterargument to the 'no liability, no speech' mantra could rely on the observation that whereas search engine results (like content feeds) return third-party content, autocomplete suggestions are search engines' own content, as noted by the German Federal Court of Justice in 2013.[48] In turn, the French Court of Cassation considered the status of autocomplete suggestions in two cases but failed to engage in a deeper analysis of their possible nature as speech.[49]

Another reason for the more limited focus of European courts on algorithmic speech may lie in the fact that both the European Convention on Human Rights

that in the US, the third-party content for which a service provider bears no liability is different from the speech consisting in the selection and compilation, resulting from content curation, which may be attributed to hosting providers such as social networks or search engines.

[45]For an overview of the most important challenges posed by the Internet service providers' liability regime in the EU, *see* B. Petkova and T. Ojanen (eds.), *Fundamental Rights Protection Online: The Future Regulation of Intermediaries* (Edward Elgar 2020). It is no coincidence that the national courts of some EU member states have interpreted these exemptions restrictively when they found indications of editorial control in the way service providers (mostly hosting providers, such as social networks) operated their services, thus qualifying them as 'active' service providers. In this respect, *see* Bassini, *supra* n. 43 and, in a broader perspective, E. Apa and O. Pollicino, *Modeling the Liability of Internet Service Providers: Google vs. Vivi Down* (Egea 2013).

[46]*See* Sears, *supra* n. 21, p. 1341.

[47]This is without prejudice to the fact that if service providers were found to exercise editorial control, they would likely be equated with content providers, at least for the purposes of the applicable liability framework concerning third-party illegal content.

[48]*See* German Federal Court of Justice, VI ZR 269/12 of 14 May 2013. As highlighted by Sears, *supra* n. 21, p. 1332, this judgment found that autocomplete suggestions can impart meaning, thus creating room for a possible, yet undelivered, acknowledgment of constitutional coverage.

[49]*See* French Court of Cassation, 1e civ., 19 February 2013, Bull. Civ. I, No. 19 (Fr.) and 1e civ., 19 June 2013, Bull. Civ. I, No. 625 (Fr.). *See also* S. Karapapa and M. Borghi, 'Search Engine Liability for Autocomplete Suggestions: Personality, Privacy and the Power of the Algorithm', 23(3) *International Journal of Law and Information Technology* (2015) p. 261.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

and the Charter of Fundamental Rights of the European Union, similarly to most of the national constitutions, confer free speech rights to individuals ('everyone')[50] in a more specific way that goes beyond the comprehensive prohibition on abridgment of speech at the heart of the First Amendment.[51] This structural difference in the wording of the relevant provisions[52] has inevitably influenced the degree of openness of the relevant jurisdictions to considering the existence of possible constitutional coverage for such an innovative concept as algorithmic speech.

The different framing of freedom of expression has also been reflected in the scholarship exploring this relationship. As previously noted, US scholars have extensively investigated whether algorithmic output can aspire to some constitutional coverage, perhaps driven by a 'First Amendment expansionism'[53] that has led some authors to question if free speech is now living a new *Lochner* era;[54] in a different vein, European commentators have primarily focused on the possible application of algorithmic technologies to the performance of tasks with an impact on freedom of expression, such as content moderation.[55] The latter perspective mirrors the understanding of AI as a technology that may pose new

[50]As noted by Sears, *supra* n. 21, the French Court of Cassation, in its judgment of 19 February 2013, declined to grant free speech rights to legal persons, interpreting Art. 10 of the Convention as conferring freedom of expression only to individuals.

[51]*See* A.T. Kenyon, 'Complicating Freedom: Investigating Positive Free Speech', in A.T. Kenyon and A. Scott (eds.), *Positive Free Speech. Rationales, Methods and Implications* (Hart Publishing 2021) p. 1. *See also* Barendt, *supra* n. 14, p. 100.

[52]*See* some national constitutions: Art. 5 of the German Basic Law, Art. 21 of the Italian Constitution; *see also* Art. 20 of the Spanish Constitution.

[53]*See* L. Kendrick, 'First Amendment Expansionism', 56(4) *William & Mary Law Review* (2015) p. 1199.

[54]*See* G. Lakier, 'The First Amendment's Real Lochner Problem', 87 *The University of Chicago Law Review* (2020) p. 1241; more recently, E. Douek and G. Lakier, 'Lochner.com?', 138 *Harvard Law Review* (2024) p. 100.

[55]*See*, for instance, G. De Gregorio and P. Dunn, 'Artificial Intelligence and Freedom of Expression', in A. Quintavalla and J. Temperman (eds.), *Artificial Intelligence and Human Rights* (Oxford University Press 2023) p. 76; E. Llansó et al., 'Artificial Intelligence, Content Moderation, and Freedom of Expression, in One', Working Papers from the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 26 February 2020; M. Brkan, 'Freedom of Expression and Artificial Intelligence: On Personalisation, Disinformation and (Lack of) Horizontal Effect of the Charter', Maastricht Faculty of Law Working Papers (2019); T. Dias Oliva, 'Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression', 20 *Human Rights Law Review* (2020) p. 607; B. Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation', 43 *Fordham International Law Journal* (2019) p. 939. *See also* E. Longo, 'The Risks of Social Media Platforms for Democracy: A Call for a New Regulation', in B. Custers and E. Fosch-Villaronga (eds.), *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice* (Springer 2019) p. 169. In general terms, from a US-based perspective, *see* J. Balkin, 'Free

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

threats to freedom of expression rather than as a technology that opens new avenues for that freedom.

Against this background, it is necessary to shift from the speaker's perspective to that of the listener to determine whether AI-generated content can aspire to constitutional coverage as speech. It should come as no surprise that, once again, the question first raised the interests of US scholars.[56]

Algorithmic speech and generative AI speech: the right to speak

Having examined the current state of knowledge regarding the constitutional coverage of the forms of algorithmic speech that preceded the advent of AI systems, it is now time to focus specifically on generative AI and to delve into the two perspectives under which AI-generated content could be considered speech; the investigation into the two distinct aspects of the right to speak and the right to receive information will facilitate the identification of the existing grounds for constitutional coverage of generative AI output.

At both global and local levels, freedom of expression is, first and foremost, the right of individuals to disseminate thoughts and opinions without constraints;[57] thus, it is a right of the *speaker*.

If the scope of freedom of expression were limited to the right to speak, generative AI output might be prevented from receiving constitutional protection. A 'killer argument' in this respect could lie in the impossibility to find a 'holder' of the right to whom the content can be attributed, that is, a speaker.

Yet, the (apparent) lack of a human speaker does not seem to place an insuperable obstacle to conceptualising a freedom of expression coverage for AI-generated content.

Massaro, Norton and Kaminski, in a far-sighted 2017 piece[58] exploring the (at the time) future challenges of strong AI (a broader category including generative AI), found that extending free speech rights to AI speakers would be

Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', 51 *University of California Davis Law Review* (2018) p. 1148.

[56]*See* E. Volokh et al., 'Freedom of Speech and AI Output', 3 *Journal of Free Speech Law* (2023) p. 651; M. Kaminski, 'Authorship, Disrupted: AI Authors in Copyright and First Amendment Law', 51 *University of California Davis Law Review* (2017) p. 589; C. Sunstein, 'Artificial Intelligence and the First Amendment', ssrn.com, 27 April 2023; E. Volokh, 'Large Libel Models? Liability for AI Output', 3 *Journal of Free Speech Law* (2023) p. 489; T. Massaro et al., 'SIRI-OUSLY 2.0: What Artificial Intelligence Reveals about the First Amendment', 101 *Minnesota Law Review* (2017) p. 2481. *See also* Collins and Skover, *supra* n. 34; M. Lamo and R. Calo, 'Regulating Bot Speech', 66 *UCLA Law Review* (2019) p. 988.

[57]*See* S. Gardbaum, 'The Structure of a Free Speech Right', in A. Stone and F. Schauer (eds.), *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) p. 213.

[58]Massaro et al., *supra* n. 56, p. 2487-2491.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

consistent with the traditional free speech 'positive' justifications, namely: democracy and self-governance, marketplace of ideas and autonomy.[59]

In this way, they showed that First Amendment protection is agnostic to the human nature of the speaker since it is the existence of speech (rather than, and regardless of, a speaker) that truly matters. In fact, US courts found no impediment to qualifying legal persons, such as corporations, as holders of free speech rights.[60] This conclusion was perhaps facilitated by the wording of the First Amendment, which is centred on the prohibition of any abridgment of speech by the government. The First Amendment does not identify individuals as the sole holders of free speech rights, whereas Article 10 of the European Convention on Human Rights and Article 11 of the Charter of Fundamental Rights of the European Union state that 'everyone' is entitled to freedom of expression. However, the case law of the European Court of Human Rights has found that legal persons, particularly in the media industry, also qualify for protection under Article 10.[61] Similarly, the European Court of Justice has extended some of the rights enshrined in the Charter to legal persons, although no precedent specifically addresses freedom of expression.[62]

A possible solution that is often discussed in the context of algorithmic speech is to attribute the output of generative AI to the programmers. As a sort of *fictio juris*, such a solution could potentially accommodate the need for a human speaker if deemed necessary, while at the same time being consistent with the attribution of speech rights to non-human speakers, such as corporations.

[59]This analysis had already been developed in T. Massaro and H. Norton, 'Siri-ously? Free Speech Rights and Artificial Intelligence', 110 *Northwestern University Law Review* (2016) p. 1169 at p. 1175-1182.

[60]*See First National Bank of Boston* v *Bellotti*, 435 US 765 (1978), where the Court articulated for the first time its human-agnostic approach, emphasising that the First Amendment aims to protect against abridgments of speech, no matter whether the relevant speaker is an individual or a corporation (ibid., p. 776). The Court confirmed this view in *Citizens United* v *Federal Election Commission*, 558 US 310 (2010); *see also* Scalia, concurring, p. 9: 'The Amendment is written in terms of "speech," not speakers. Its text offers no foothold for excluding any category of speaker, from single individuals to partnerships of individuals, to unincorporated associations of individuals, to incorporated associations of individuals'. More recently, *see also Burwell* v *Hobby Lobby Stores*, 573 US 682 (2014).

[61]*See*, e.g., ECtHR 22 May 1990, No. 12726/87, *Autronic AG* v *Switzerland*; ECtHR 26 April 1979, No. 6538/1974, *The Sunday Times* v *United Kingdom*; ECtHR 15 December 2009, No. 821/03, *Financial Times Ltd* v *UK*; ECtHR 7 February 2012, No. 39954/08, *Axel Springer AG* v *Germany*; ECtHR 7 June 2012, No. 38433/09, *Centro Europa 7 SRL* v *Italy*.

[62]*See* ECJ A.G. Bobek's opinion in C-194/16, paras. 41-51, and the judgments mentioned therein.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

However, what makes AI systems truly unique and disruptive is their ability to learn from input data how to return output, such as predictions, content, recommendations or decisions that can influence physical or virtual environments.[63] As clearly stated in Recital 12 of the EU AI Act, what differentiates AI systems from simpler, traditional software systems or programming approaches is their capability to infer. This also marks an important difference compared to systems based on rules defined solely by natural persons to automatically execute operations, like the most common software programmes. It is precisely because of the underlying machine learning and knowledge-based or logic-based approaches that AI systems can go beyond the performance of predetermined tasks. Therefore, even if one were to consider whether AI output could be attributed to system programmers as their speech, this approach might prove pointless; determining a definitive 'speaker' would be practically impossible as these systems can deviate from the original input provided by programmers and produce output with a degree of 'autonomy'.

The difficulties of identifying a 'speaker' (whether human or non-human) may, therefore, prove to be an argument to decline constitutional coverage for generative AI output.

But even if we were able to identify 'AI speakers' and to confer speech rights to entities other than humans, is the speaker the correct angle from which to look at the constitutional coverage of ChatGPT output?

Some of the scholars who have recently ventured into the status of generative AI output have emphasised the importance of a qualified connection between the algorithmically generated content at hand and the input of the speaker (i.e. algorithm owner).[64] Benjamin, for example, tried to adapt and recontextualise the *Spence* test developed by the US Supreme Court,[65] which requires a substantive message that is 'sendable and receivable' and that the speaker 'chooses to send'.[66] Blackman, in turn, put the '*nexus* that the algorithmic outputs have with human interaction' at the heart of the relevant constitutional inquiry.[67] Wu's mentioned assertion of a functionality doctrine also draws

[63] *See* Art. 3(1)(1) AI Act.

[64] M. Goswami, 'Algorithms and Freedom of Expression', in W. Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms* (Cambridge University Press 2020) p. 558 at p. 566.

[65] *See* Kaminski, *supra* n. 56, p. 609.

[66] *See* Benjamin, *supra* n. 38, p. 1461.

[67] J. Blackman, 'What Happens if Data is Speech?', 16 *The University of Pennsylvania Journal of Constitutional Law* (2014) p. 25 at p. 34. According to Blackman, 'The more the human interacts, the closer the communication will be something the human created herself, and something that warrants protection. In contrast, outputs that are created with isolated autonomy, and involve little personal involvement – save for the programmer's coding – departs further from the humanistic expression that warrants protection'.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

boundaries between what constitutes speech and what does not.[68] In their *Robotica*, Collins and Skover appealed to a 'new norm of utility' to grant coverage and protection to robotic speech that a receiver experiences as 'meaningful and potentially useful or valuable';[69] in a nutshell, they recognise intention-less free speech as 'at the interface of the robot and receiver'.[70]

This variety of views has so far exclusively emerged in US scholarship, and it reflects the challenges inherent in adopting the perspective of the speaker to develop a coherent conceptual framework for the constitutional coverage of generative AI output. Even if scholars like Kaminski have found the conferral of free speech rights to 'AI speakers' to be consistent with the traditional positive free speech theories, this essay advocates for another justification. This call for a different rationale depends, first, on the technical aspects of the most advanced generative AI, which would make the search for a speaker (whether human or not) rather uncomfortable and uneasy,[71] if not impossible, considering the *black box* effect.[72] But in addition and above all else, the call rests on the need for a better angle than that of the speaker to discuss the constitutional coverage of generative AI output.

This does not mean that the problem of determining the existence of a speaker and defining whether the speaker is human or not is unworthy of careful legal consideration; but, as noted by Kaminski, 'a government that censors a political novel written by an algorithm is as problematic from the perspective of a reader as a government that censors a political novel written by Tolstoy'.[73] This observation mirrors an essential switch from the angle of speakers to that of listeners, where 'what matters' is 'whether the work reads as speech to those who encounter it'.[74]

Generative AI and the individual's right to receive information

The right to speak is not the only component of freedom of expression, which also protects the right of everyone to receive information. This peculiarity is consistent with both the rationale that states should refrain as much as possible from

[68]Wu, *supra* n. 30.

[69]Collins and Skover, *supra* n. 34, p. 42.

[70]Ibid.

[71]Kaminski, *supra* n. 56, p. 609-610, also highlights another possible shortcoming of this perspective. While commenting on Benjamin's proposal to adapt the *Spence* test to algorithmic speech to qualify as speech a 'message that is sendable and receivable and that one actually chooses to send', she notes that, in contrast to the *Spence* traditional requirement of a message that a speaker intends to communicate and that is likely to be understood, AI speakers may not meet this standard, for example, because 'AI cannot have intent or does not adequately express the intent of its human programmers'.

[72]F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

[73]Ibid., p. 610.

[74]Ibid.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

interfering in the public sphere, no matter whether an individual is considered as a speaker or a listener, and with the existence of positive obligations on states to foster media pluralism, obligations that go beyond a purely negative understanding of free speech.[75]

The question, then, is whether the existence of a right to receive information implies that AI-generated content is also protected speech; and if so, does this shift to the listener's perspective present fewer conceptual difficulties than the speaker's perspective in framing constitutional coverage?

Lessons from cyberspace

To elaborate on this perspective, it is key to recall how the birth of cyberspace raised similar issues in the late 1990s. As cyberspace became popular, scholars began to discuss the impact of this technology, especially on freedom of expression. In its earlier days, the new digital world looked, on the one hand, like a 'promised land of freedom',[76] yet on the other hand, the unique nature of cyberspace could facilitate conduct such as violations of privacy, defamation and copyright infringement. In the US, the unexpected proliferation of cyberporn soon became a major problem, which prompted Congress to legislate to protect minors from harmful content. These measures were challenged by civil society organisations, which saw them as an attack on Internet freedom, and were very strictly reviewed by the US Supreme Court beginning with the landmark case of *Reno* v *ACLU*.[77]

In such circumstances, it was easy to understand that not every human activity occurring in cyberspace amounted to an exercise of freedom of expression, despite the initial 'illusion' of a realm where everything was 'freedom',[78] a myth fuelled by Barlow's famous Declaration of Independence of Cyberspace.[79] It became clear that there was no general and necessary equivalence between the medium (i.e. the Internet) and the most important freedom that could be exercised on that medium.[80] In terms of constitutional coverage and protection, it is the content

[75] *See* Kenyon, *supra* n. 51.

[76] *See*, e.g., D.R. Johnson and D. Post, 'Law and Borders: The Rise of Law in Cyberspace', 48(5) *Stanford Law Review* (1996) p. 1367.

[77] *See Reno* v *ACLU*, 521 US 844 (1997); *see also Ashcroft* v *American Civil Liberties Union*, 542 US 656 (2004).

[78] *See* J. Goldsmith and T. Wu, *Who Controls the Internet? Illusions of a Borderless World* (Oxford University Press 2006). *See also* J. Goldsmith, 'Against Cyberanarchy', 65(4) *University of Chicago Law Review* (1998) p. 1199.

[79] J.P. Barlow, 'A Declaration of the Independence of Cyberspace', 18 *Duke Law & Technology Review* (2019) p. 5.

[80] *See*, e.g., ECJ 24 November 2011, Case C-70/10, *Scarlet* v *SABAM* and ECJ 16 February 2021, Case C-360/10, *SABAM* v *Netlog*, where the ECJ had to review the compatibility with fundamental rights, such as privacy, data protection and freedom of information, of injunctions

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

that ultimately matters. The rise of cyberspace and the subsequent regulatory developments have made this point undisputed: despite the fact that the courts have featured different approaches to freedom of expression in the digital age[81] (sometimes emphasising the potential for the exercise of such a paramount right in the new digital realm,[82] sometimes focusing more cautiously on the risks its exercise poses to other fundamental rights and interests[83]), the mere use of the Internet does not necessarily amount to the exercise of a right to speak.

Now that a new technological revolution is taking place, AI can perhaps be seen in a similar vein as another medium through which constitutional freedoms can be exercised.

The US perspective

From the perspective of information recipients, a restriction on freedom of expression occurs when the government obstructs their ability to access certain content. Since freedom of expression is not an absolute right, such restrictions are not inherently unconstitutional, but they necessitate a justification and must be proportionate to the objective pursued. However, it is precisely in the assessment of the constitutionality of potential restrictions that differing standards are applied in the legal systems of the US and Europe.

In the US, the First Amendment does not explicitly mention the right of individuals to receive information. Instead, it generally focuses on the prohibition of any abridgment of speech. However, as noted by the scholars who have investigated the relationship between AI and freedom of speech,[84] the perspective of the recipient of information nonetheless plays a key role.[85]

issued by the Belgian authorities against service providers and aimed at enforcing copyright on peer-to-peer networks by preventing possible infringements *ex ante*.

[81]*See* O. Pollicino, *Judicial Protection of Fundamental Rights on the Internet: A Road Towards Digital Constitutionalism?* (Hart Publishing 2021).

[82]*See Reno* v *ACLU*, *supra* n. 77.

[83]*See Brown* v *Entertainment Merchants Association*, 564 US 786 (2011); ECtHR 5 May 2011, No. 33014/05, *Editorial Board of Pravoye Delo and Shtekel* v *Ukraine*.

[84]*See* Volokh et al., *supra* n. 56; Sunstein, *supra* n. 56.

[85]In Sunstein's view, ibid., this conclusion is also supported by the Supreme Court judgment in *Citizens United* v *Federal Election Commission*, 558 US 310 (2010): 'Speech is an essential mechanism of democracy, for it is the means to hold officials accountable to the people . . . . The right of citizens to inquire, to hear, to speak, and to use information to reach consensus is a precondition to enlightened self-government and a necessary means to protect it'. In this judgment, the Court did emphasise the role of speech as an 'essential mechanism of democracy', most notably in light of the passive side of this freedom, and stressed that the relevant protection applies no matter whether the author is a human or non-human agent (such as a private company).

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

The idea that individuals are afforded the widest possible range of perspectives or the greatest extent of pluralism of information is a particularly appealing concept in the US, where freedom of expression – especially in the digital age – is still predominantly shaped by the metaphor of the free marketplace of ideas.[86] While it is genuinely supportive of a pluralism of ideas and opinions, this view of freedom of expression is sceptical of any governmental interference. Although some commentators have challenged the relevance of this metaphor in the context of the digital sphere,[87] US courts have consistently upheld a broad interpretation of the scope of protection granted by the First Amendment,[88] which has resulted in the presumption of constitutional impermissibility of content-based regulation of speech.[89]

Given this background, the free flow and confrontation of ideas and opinions, which are ideally pursued by the free marketplace metaphor, could be undermined by provisions that, for example, would place restrictions on the ability of generative AI to produce certain output. The problem is not the restriction, however, but rather its scope and justification.

Prominent free speech scholar Cass Sunstein[90] has recently delved into the permissibility of such restrictions. As the Supreme Court made clear in its case law, content-based restrictions are more challenging for freedom of speech, as they likely (albeit not necessarily) result in viewpoint discrimination. Consequently, they are subject to strict scrutiny, whereas the constitutionality of content-neutral restrictions is generally assessed based on intermediate or mid-level scrutiny.[91]

[86] *Abrams* v *United States*, 250 US 616 (1919), Holmes dissenting.

[87] *See* A. Morelli and O. Pollicino, 'Metaphors, Judicial Frames, and Fundamental Rights in Cyberspace', 68(3) *The American Journal of Comparative Law* (2020) p. 616, who illustrate, in particular, some concerns from the European standpoint; *see also* D. Nunziato, 'The Marketplace of Ideas Online', 94 *Notre Dame Law Review* (2019) p. 1519; with respect to disinformation, *see* A. Waldman, 'The Marketplace of Fake News', 20(4) *Journal of Constitutional Law* (2018) p. 845 and T. Wu, 'Disinformation in the Marketplace of Ideas', 51 *Seton Hall Law Review* (2020) p. 169.

[88] *See also* the recent case *Murthy* v *Missouri*, 603 US ____ (2024) and the comment by T. Wu, 'The First Amendment Is Out of Control', *The New York Times*, 2 July 2024, https://www.nytimes.com/2024/07/02/opinion/supreme-court-netchoice-free-speech.html/, visited 3 June 2025.

[89] *See* Massaro and Norton, *supra* n. 59, p. 1186, who highlight the 'broad protection of speech regardless of content (with all bets on the audience's ability to sort good speech from bad)' and see it as an argument to support strong AI speech 'regardless of nontraditional source or format'.

[90] Sunstein, *supra* n. 56.

[91] However, *see Turner Broad. Sys.* v *FCC*, 512 US 622, 642 (1994), where the Supreme Court highlighted that 'Deciding whether a particular regulation is content-based or content-neutral is not always a simple task. We have said that the "principal inquiry in determining content-neutrality ... is whether the government has adopted a regulation of speech because of [agreement or] disagreement with the message it conveys" ... But while a content-based purpose may be sufficient in certain circumstances to show that a regulation is content-based, it is not necessary to such a showing in all cases ... Nor will the mere assertion of a content-neutral purpose be enough to save a law which, on its face, discriminates based on content'.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

In Sunstein's view, as with any other content-based restrictions, measures that interfere with generative AI's ability to produce certain output (e.g. content critical of political majorities) and that result in viewpoint discrimination should be presumed to be contrary to the First Amendment.[92] If not, lawmakers would have a truly unique opportunity to interfere with the exercise of free speech: for example, by limiting what large language models can 'say', governments would be able to influence what individuals can find, see and read on ChatGPT, that is, the information they can receive.[93]

This approach is consistent with the reasoning of the Supreme Court in *Reno* v *ACLU*,[94] its first cyberlaw precedent. The Court highlighted that 'as a matter of constitutional tradition, in the absence of evidence to the contrary, we presume that governmental regulation of the content of speech is more likely to interfere with the free exchange of ideas than to encourage it'.[95]

The solution endorsed by Sunstein does not exclude important consequences. First, the constitutional coverage of generative AI output under the First Amendment as part of the individual's right to access information does not preclude the adoption of legitimate restrictions on the use of these systems that are inherent to the medium. Such limitations might depend on circumstances rather than on the content. Among possible content-neutral limitations, Sunstein includes the example of a ban on the use of technology imposed by academic institutions in the context of examinations.[96] These content-neutral restrictions would be dependent on the protection of a significant public interest.

Moreover, and consistent with the latest finding, the fact that the content created through generative AI enjoys constitutional coverage does not mean that every output will necessarily receive protection.[97]

[92]On a parallel but connected domain, namely copyright, *see* Kaminski, *supra* n. 56.

[93]Sunstein, *supra* n. 56, p. 10, quotes the Supreme Court case *Kleindienst* v *Mandel*, 408 US 753 (1972), in which the speaking subject could not be considered the holder of free speech rights as he lacked the status of US citizen, whereas this right, in the passive projection of the right to receive information, could certainly be considered relevant due to the listeners of the speaker (in this case, those who had extended an invitation to deliver a speech at a conference in the US).

[94]521 US 844 (1997), *supra* n. 77.

[95]Ibid., p. 885.

[96]Sunstein, *supra* n. 56, p. 14.

[97]On the difference between coverage and protection, *see* M. Tushnet, 'The Coverage/Protection Distinction in the Law of Freedom of Speech – An Essay on Meta-Doctrine in Constitutional Law', 25 *William & Mary Bill of Rights Journal* (2017) p. 1073; *see also* F. Schauer, 'What Is Speech? The Question of Coverage', in Stone and Schauer (eds.), *supra* n. 57, p. 158. For a comparative law analysis, *see* Goswami, *supra* n. 64, p. 564, who explores the Canadian legal system and section 2(b) of the Charter of Rights and Freedoms, which requires an activity that 'conveys or attempts to convey a meaning' for freedom of expression protection to be claimed.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

Looking at the problem from the listener's perspective with the 'support' of the predominantly negative view of the First Amendment adopted by the Supreme Court, it turns out that the US legal system provides solid grounds to treat AI-generated content as speech. This framing captures what is, in fact, the most common source of concerns in the digital age: that illiberal governments fulfil their desire to suppress speech by limiting the ability of individuals to access content, for example, under the guise of fighting disinformation. The negative view of freedom of speech is consistent with the widespread distrust of governments, which are seen as 'bad actors'.[98]

The adoption of this approach does not strip validity from the justification offered by positive free speech theories for 'AI speech'. But the negative perspective offers the advantage, in the words of Massaro, Norton and Kaminski, of a shift 'from asking whether a particular activity is speech to asking whether the government intends to target speech'.[99]

Finally, this perspective succeeds in extending free speech coverage to generative AI output without engaging in any complex and disputable attribution of speaker's rights. As Sunstein noted, when the Supreme Court found content other than the speech of individuals, such as video games, to be covered under the First Amendment, it did not imply that the latter had *per se* constitutional protection. Rather, such protection applies to the individuals who engage in video games. The rationale, as articulated by Volokh, Lemley and Henderson, is that generative AI content qualifies as speech 'because of the First Amendment rights of those who would receive the speech, whether or not AI companies' own free speech rights are implicated as well'.[100]

The European perspective

The European understanding of freedom of expression differs from that of the US legal system, among others, due to a more balanced relationship among the various fundamental rights protected in the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. This

[98]*See* Massaro et al., *supra* n. 56, p. 2493, who recall the adoption of this perspective by the Supreme Court, for example, in *Heffernan* v *City of Paterson*, 578 US 266, where it emphasised First Amendment's 'restraints on potentially dangerous governmental power rather than positive reasons for protecting speakers or speech'.

[99]Massaro et al., *supra* n. 56, p. 2494. The authors also note that adhesion to the negative theory 'offers no meaningful limiting principles that would permit governments to regulate speech under certain conditions. It also does not elide the "what is speech" question entirely, as no free speech problem arises if a government motive is to regulate pure conduct and the law is applied in a speech-neutral way'; ibid.

[100]Volokh et al., *supra* n. 56, p. 657.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

difference also reflects the value attached to freedom and dignity in the respective jurisdictions,[101] as evidenced by some early cyberlaw cases.[102]

While both the Charter and the Convention stand out as key provisions, it is through the lenses of Article 10 of the Convention that one can obtain a more insightful perspective on the European understanding of freedom of expression. For the purposes of this analysis, it is therefore adopted as the paradigm of this right in Europe – also because of its long-standing influence as a general principle of EU law and the consistency clause under Article 52(3) of the Charter, which fostered the alignment between the case law of the European Court of Human Rights and the European Court of Justice on freedom of expression.

Article 10 of the Convention (like Article 11 of the Charter) does not model freedom of expression as an absolute right, but it does reflect a less suspicious attitude towards governments than does the First Amendment – like various national constitutions, it primarily focuses on the rights conferred to individuals. These include the right to hold opinions and to receive and impart information and ideas without interference by public authorities and regardless of frontiers.

Therefore, in Europe freedom of expression is also afforded multifaceted coverage, extending not only to those who *speak* but also to those who *listen*. The question of whether the right to receive information implies a generalised right to 'proper information' (or 'not to be disinformed'), which is information that meets certain quality standards, remains open to debate, though. This construction of the right to receive information might not only validate measures aimed at countering disinformation but also impose positive obligations on the contracting states to actively intervene.[103] To date, the case law of the European Court of Human Rights has not directly addressed this point, except for a few cases concerning pluralism in public service broadcasting and access to government information.

The freedom to receive information has also played a significant role in the domestic case law of some constitutional courts on the concept of pluralism.[104] For instance, the Italian Constitutional Court[105] relied heavily on the right to

[101] *See*, e.g., M. Rosenfeld and A. Sajó, 'Spreading Liberal Constitutionalism: An Inquiry into the Fate of Free Speech Rights in New Democracies', in S. Choudry (ed.), *The Migration of Constitutional Ideas* (Cambridge University Press 2005) p. 142.

[102] *See* TGI Paris, 22 May 2000, *Licra et UEJF* v *Yahoo Inc and Yahoo France*; *Yahoo!, Inc.* v *La Ligue Contre Le Racisme*, 169 F Supp 2d 1181 (ND Cal 2001); *Yahoo! Inc.* v *La Ligue Contre Le Racisme et L'antisemitisme, UEJF*, 433 F.3d 1199 (9th Cir. 2006).

[103] G. Pitruzzella and O. Pollicino, *Disinformation and Hate Speech: A European Constitutional Perspective* (Bocconi University Press 2020).

[104] *See* E. Barendt, 'The Influence of the German and Italian Constitutional Courts on their National Broadcasting Systems', *Public Law* (1991) p. 93.

[105] *See*, e.g., Italian Constitutional Court rulings Nos. 202/1976, 148/1981, 153/1987, 826/1988, 112/1993.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

receive information, which is implicit in Article 21 of the Italian Constitution,[106] to require the legislature to implement adequate safeguards to ensure media pluralism; likewise, the never-ending series of judgments of the German Federal Constitutional Tribunal concerning the national public service broadcasting is deeply rooted in Article 5 of the Basic Law[107] in affording protection to viewers and listeners.[108]

Against this background, scholars such as de Vries[109] have pinpointed the 'passive side' of freedom of expression as the most suitable option to provide protection to generative AI output. As she pointed out,[110] the language of paragraph 1 of Article 10 of the Convention refers to the relevant rights (*human* rights, indeed) as applying to 'everyone', thereby implying that the fundamental right to freedom of expression can only be attributed to individuals. In her view,[111] although the European Court of Human Rights has broadly interpreted the term 'everyone', encompassing legal persons under exceptional circumstances,[112] extending it to AI systems would be at odds with Article 35(3) of the Convention, which stipulates that applicants must possess the status of victims of an infringement. Accordingly, the speaker's perspective would not be the most well-suited to provide coverage for AI output.[113] As noted in the analysis of the

[106]Whose para. 1 only focuses on the active side, providing that 'Everyone has the right to freely express their ideas through speech, in writing and by any other means of communication'. However, the second paragraph ('The press shall not be subjected to authorization or censorship') and the third paragraph (requiring a judicial order as a condition for the lawful seizure of press materials, permitted only in specific cases provided by law) were extensively interpreted by Italian courts to apply the relevant safeguards beyond the scope of the press.

[107]Whose para. 1 establishes that 'every person shall have the right freely ... to inform himself without hindrance from generally accessible sources', thus specifically covering the right of individuals to information.

[108]*See* German Federal Constitutional Tribunal, rulings Nos. 12 BVerfGE 205 (1961); 31 BVerfGE 314 (1971); 57 BVerfGE 295 (1981); 73 BVerfGE 118 (1986); 74 BVerfGE 297 (1987); 83 BVerfGE 238 (1991); 90 BVerfGE 60 (1994); 97 BVerfGE 228 (1998); 119 BVerfGE 181 (2007).

[109]K. de Vries, 'Let the Robot Speak! AI-Generated Speech and Freedom of Expression', in S. Hindelang and A. Moberg (eds.), *YSEC Yearbook of Socio-Economic Constitutions 2022* (Springer 2021) p. 93.

[110]Ibid., p. 99.

[111]Ibid., p. 100.

[112]*Supra* n. 61.

[113]However, it is worth noting that the persistent adequacy of the traditional understanding of agency is currently debated in literature, most notably with respect to other fundamental rights, such as that of privacy, particularly in light of recent technological developments that are likely to bring large-scale human rights violations. *See*, for example, E. Kosta, 'Algorithmic State Surveillance: Challenging the Notion of Agency in Human Rights', 16(1) *Regulation & Governance* (2022) p. 212.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

right to speak, the lack of a human speaker may not be an unsurmountable obstacle to constitutional coverage; however, the perspective of the speaker may be misguiding, and the listener's standpoint may more accurately capture the sense of extending constitutional coverage to generative AI output.

The European Court of Human Rights has addressed violations of the right to receive information[114] on various occasions. The Court ruled on the compatibility with Article 10 of the Convention of measures taken by national authorities consisting of the blocking of certain websites. In *Cengiz* v *Turkey*,[115] the Court determined that the complete blocking of YouTube in Turkey, ordered in the context of a criminal trial to prevent access to ten webpages, was incompatible with Article 10. The applicants were Turkish law professors who complained about the overbroad effects of the blocking order, which, in their view, interfered with their right to receive (and impart) information and ideas. It is noteworthy that the Court did not dispute the status of the applicants as victims, even though they were not the target of the blocking order, which had been adopted without being prescribed by law. In its case law concerning the blocking of the Internet, the Court has placed particular emphasis on the impact of collateral effects on Internet users, which may include the impossibility of accessing significant sources of information.

Similarly, the Court has emphasised the pivotal role of audiovisual media (radio and television) as a conduit for the dissemination of information and ideas on matters of public interest, a function that is arguably as influential as that of the press.

In considering the role of the press in society, the Court has pointed out that its responsibility is particularly significant in light of the right of individuals to receive the information it imparts.[116]

The Court also clarified that the right to receive information is not confined solely to matters of public concern; under certain conditions, it may extend to cultural expressions and pure entertainment.[117]

Furthermore, the Court underscored that in the context of the public broadcasting service, contracting states bear a responsibility to guarantee 'that the public has access through television and radio to impartial and accurate information and a range of opinion and comment, reflecting inter alia the

[114]For an in-depth and exhaustive overview of the distinctive 'perspectives' emerging in the case law of the ECtHR concerning the right to receive information, *see* S. Eskens et al., 'Challenged by News Personalisation: Five Perspectives on the Right to Receive Information', 9(2) *Journal of Media Law* (2017) p. 259.

[115]ECtHR 1 December 2015, Nos. 48226/10 and 14027/10, *Cengiz* v *Turkey*.

[116]ECtHR 7 December 1976, No. 5493/72, *Handyside* v *UK*, para. 49; ECtHR 8 July 1986, No. 9815/82, *Lingens* v *Austria*, paras. 41-42.

[117]ECtHR 16 March 2009, No. 23883/06, *Khurshid Mustafa and Tarzibachi* v *Sweden*.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

diversity of political outlook within the country'.[118] Of course, it remains debatable whether this could give rise to positive obligations for contracting states to prevent or counter disinformation, including in the context of generative AI output. In fact, the Court made these remarks to stress the role of the state as 'the ultimate guarantor of pluralism',[119] for example, in the shaping of the media system or public service broadcasting. Similarly, the Court recently[120] emphasised the key role of accuracy and reliability of the information provided by public authorities for the effectiveness of the individuals' right of access to government information.[121] But the only general claim of a 'right of the public to be *properly* informed' came in the *Sunday Times* v *UK* case back in 1979; as noted by Pentney, 'the qualifier disappeared in subsequent cases'.[122] So far, then, a right not to be disinformed has been conceptualised only as part of the right of access to information from public authorities.

Given the specific structure of Article 10 of the Convention, the Court's case law has focused more on the specific conditions that could on a case-by-case basis justify the relevant limitations. Thus, any measure seeking to limit the ability of generative AI to 'say something', thereby impacting the right of individuals to receive such information, could theoretically be reviewed by the Court under the Article 10-based test. In the end, this would be no different from the review engaged in by the Court for restrictions specifically concerning, for example, online speech.

What matters is speech, not the medium

A comparative analysis of the First Amendment of the US Constitution and Article 10 of the Convention reveals that despite different understandings of freedom of speech, there is no apparent reason to exclude AI-generated content from constitutional coverage. However, this does not imply that every use of generative AI systems should amount to speech.

Given that coverage does not necessarily imply protection,[123] the existence of constitutional coverage would not prevent states from legislating on this matter, for example, by requiring that generative AI output not consist of illegal content,

[118]ECtHR 17 December 2009, No. 13936/02, *Manole and Others* v *Moldova*, para. 100.

[119]Ibid., para. 107.

[120]ECtHR 1 July 2021, Nos. 56176/18, 56189/18, 56232/18, 56236/18, 56241/18 and 56247/18, *Association Burestop and Others* v *France*, para. 108.

[121]For an in-depth comment on the case, *see* K. Pentney, 'The Right of Access to "Reliable" Information under Article 10 ECHR: From Meagre Beginnings to New Frontiers', 5 *European Convention on Human Rights Law Review* (2024) p. 230.

[122]Ibid., p. 250 (specifically at n. 91).

[123]*See* Schauer, *supra* n. 97, p. 158.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

such as defamatory statements, or by banning the use of large language models in specific contexts, such as school examinations.

However, in both the US and European legal systems, such measures must pass a First Amendment or Article 10 of the European Convention of Human Rights scrutiny based on their impact on the right to receive information from the public. This conclusion should not be affected by the absence of a speaker as a natural or legal person to whom the exercise of freedom of expression can be attributed.

Content moderation and liability for generative AI output

A tale of two paradigms

If there is room to claim constitutional coverage for AI artifacts, an intertwined question concerns the applicable liability regime for illegal content produced by such systems. This point is worth exploring to understand the possible implications on freedom of expression. As noted, in pre-AI forms of 'algorithmic speech', such as search engine results, the fact that the operators of the relevant services (e.g. search engine providers) were framed as neutral conduits without any editorial control was seen as an argument depriving their output of any relevance as free speech.[124] However, in a recent judgment concerning another type of algorithmic speech (that of social networking platforms), the US Supreme Court seemed to imply that holding free speech rights does not *per se* conflict with the role of online intermediaries.[125] The point is thus worth further investigation with respect to generative AI output.

Legal systems distinguish between two possible actors in the digital sphere: service providers and content providers.[126] What differentiates the latter from the former is generally the existence of an editorial role, which service providers lack. This is the reason jurisdictions such as the US and the EU have established liability exemptions for third-party illegal content that service providers host or transmit, although to a (very) different extent. As the case law, for example, of the European Court of Human Rights illustrates, the application and enforcement of the norms

[124]*See* Sears, *supra* n. 21.

[125]*Moody* v *NetChoice, LLC*, *supra* n. 39.

[126]According to Section 230 of the US Communications Decency Act, an 'information content provider' is 'any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service', while the service provider is the provider of an 'interactive computer service', i.e. any 'information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions'.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

on moderation and liability for third-party content against intermediaries bear key implications for freedom of expression.[127]

A key question for understanding the impact of generative AI on the public sphere concerns the applicability of these rules to digital artifacts and their 'creators'. The fact that generative AI output receives constitutional coverage may have key implications for the ability of lawmakers to require a certain type of content moderation from the providers or deployers of these systems. Once again, disinformation provides an example of the challenges that AI operators may encounter in content policing, most notably when it comes to content that, although harmful, is not necessarily illegal.

Both the US and the EU legal systems have regulated the role and responsibilities of online platforms, which reflects the relevant understanding of freedom of expression. However, at least *prima facie*, neither providers nor deployers of generative AI systems seem to fit into the category of service providers given that some characteristics suggest their editorial role. Unlike service providers, AI systems do not merely process third-party content. On the contrary, they contribute to produce content that has never been created by third parties.[128] One notable difference, then, is that generative AI systems do not return merely user-generated content as output. The following sections will explore the existing content moderation frameworks in the EU and in the US with a view to understanding whether their application to generative AI output would be desirable from the perspective of freedom of expression.

Online content moderation in EU law

Under EU law, online content moderation is now governed by the Digital Services Act,[129] a regulation that entered into force in 2022, replacing the pre-existing E-Commerce Directive, which dated back to 2001. While the Digital Services Act has preserved the two key pillars of the framework once enshrined in the E-Commerce Directive – namely the absence of a general monitoring obligation

[127] *See* ECtHR 16 June 2015, No. 64569/09, *Delfi* v *Estonia*. *See also* ECtHR 2 February 2016, No. 22947/13, *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt* v *Hungary*; ECtHR 9 March 2017, No. 74742/14, *Phil* v *Sweden*; ECtHR 15 May 2023, No. 45581/15, *Sanchez* v *France*. For an overview, *see* R. Spano, 'Intermediary Liability for Online User Comments under the European Convention on Human Rights', 17(4) *Human Rights Law Review* (2017) p. 665.

[128] Volokh, *supra* n. 56, p. 495.

[129] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), also known as 'DSA'. For an overview of the relationship between the DSA and freedom of expression in the EU, *see* among others C. Corrado, 'Towards the Institutions of Freedom: The European Public Discourse in the Digital Era', 26(1) *German Law Journal* (2025) p. 114.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

and the safe harbour liability regime for online intermediaries in relation to third-party content – it has shifted away from the one-size-fits-all approach behind it, acknowledging the inherent and increased complexity of digital services.[130] The rationale behind the Digital Services Act is that online intermediaries may take different shapes, depending on the services they provide, and therefore pose different levels of risk to fundamental rights and public interests. The regulation thus operationalises a risk-based approach,[131] by differentiating four categories of services – namely, intermediary services, hosting services, online platforms, and very large online platforms as well as very large search engines – and imposing obligations that match the role, size and impact of digital services.[132] Hosting providers – a category that includes the most impactful online platforms, such as social networks – do not bear liability for third-party content unless they fail to comply with a notice-and-action mechanism that largely reflects the notice-and-take down procedure established at the time by the E-Commerce Directive.[133] Coming into being in 2022, the Digital Services Act captured a profoundly changed shape of cyberspace compared to its early stages, which were the main reason behind such a minimal regulation as the E-Commerce Directive.[134] Such transformation was mainly driven by the rise of online platforms, which more and more act as 'digital oligarchs'. This evolved background also affected the goals of the Digital Services Act, which – to be effective – not only targets illegal content but also focuses on harmful content. In particular, the Digital Services Act acknowledges that 'very large online platforms and very large online search engines can be used in a way that strongly influences safety online, the shaping of public opinion and discourse, as well as online trade'.[135] Accordingly, it subjects

[130]Before the Digital Services Act came into force, some national courts struggled to interpret the E-Commerce Directive to determine whether the more sophisticated online platforms, such as social networks, were eligible for the liability exemptions established for hosting service providers. Recital 42 of the E-Commerce Directive had a crucial influence on case law, requiring courts to assess whether the relevant service providers had gone beyond the boundaries of a purely technical, automatic and passive role. *See* in this regard, Apa and Pollicino, *supra* n. 45. *See also* M. Bassini, 'Mambo Italiano: The Perilous Italian Way to ISP Liability', in Petkova and Ojanen (eds.), *supra* n. 45, at p. 84.

[131]*See* G. De Gregorio and P. Dunn, 'The European Risk-based Approaches: Connecting Constitutional Dots in the Digital Age', 59(2) *Common Market Law Review* (2022) p. 473 at p. 483.

[132]The shift in the approach from the E-Commerce Directive to the DSA is well-captured in M.C. Buiten, 'The Digital Services Act: From Intermediary Liability to Platform Regulation', 12(5) *Journal of Intellectual Property, Information Technology and E-Commerce Law* (2021) p. 361.

[133]Interestingly, the notice-and-take-down mechanism was modelled on the provisions of the Digital Millennium Copyright Act, passed by Congress in 1998 to provide a specific framework for copyright infringements (*see* 17 USC § 512).

[134]So-called 'digital liberalism' in the words of G. De Gregorio, *Digital Constitutionalism in Europe* (Cambridge University Press 2022).

[135]Recital 79.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

the providers of very large online platforms and very large online search engines to specific obligations to assess the systemic risk 'stemming from the design, functioning and use of their services, as well as from potential misuses by the recipients of the service'[136] and to take appropriate risk mitigation measures.[137] The Digital Services Act identifies four categories of systemic risks that providers of very large online platforms or very large online search engines have to assess, including the risks concerning 'any actual or foreseeable negative effects on democratic processes, civil discourse and electoral processes, as well as public security'[138] and 'any actual or foreseeable negative effects in relation to ... the protection of public health and minors and serious negative consequences to the person's physical and mental well-being'.[139] Therefore, disinformation is a key concern for EU lawmakers and the Digital Services Act has provided some responses against the dissemination of content that, albeit not illegal, is considered harmful.[140] This trait consistently echoes the efforts of the European Commission on the Strengthened Code of Practice on Disinformation, recently integrated into the framework of the Digital Services Act, which endorses codes of practice among the applicable risk mitigation measures.[141]

Therefore, while the Digital Services Act aims to promote intermediaries' accountability and increase transparency in content moderation, it does not mean to incentivise censorship. It also does not establish any general monitoring or active fact-finding obligation[142] and exonerates hosting providers from liability unless they have actual knowledge or awareness of illegal content and fail to expeditiously act to remove it.[143]

The combination of measures provided by the Digital Services Act is intended to fully equip the EU legal system in the fight against both illegal content and harmful content, such as disinformation.[144] However, this important achievement could now be undermined by the emergence of a truly disruptive

[136]Ibid.

[137]*See* Art. 34 and Art. 35 DSA.

[138]*See* Art. 34(1)(c) DSA.

[139]*See* Art. 34(1)(d) DSA.

[140]*See* M. Husovec, 'The Digital Services Act's Red Line: What the Commission Can and Cannot do about Disinformation' 16(1) *Journal of Media Law* (2024) p. 47.

[141]European Commission, 'Commission endorses the integration of the voluntary Code of Practice on Disinformation into the Digital Services Act', 13 February 2025, https://digital-strategy.ec.europa.eu/en/news/commission-endorses-integration-voluntary-code-practice-disinformation-digital-services-act, visited 3 June 2025.

[142]*See* Art. 8 DSA (and, prior to the DSA, Art. 15 Directive 2000/31/EC).

[143]*See* Art. 6 and Art. 16 DSA (and, prior to the DSA, the similar provisions in Art. 14 Directive 2000/31/EC).

[144]As noted by A. Kenyon, 'Democratic Freedom of Expression and Disinformation', in Krotoszynski et al. (eds.), *supra* n. 8, p. 68: 'Direct state action against disinformation is not being

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

technology such as generative AI, something the legislators could not have imagined at the time of drafting the Digital Services Act. It is indisputable that the new rules were not meant to govern the creation of content by generative AI, as these systems do not fall within its scope of application.

However, while acknowledging that the Digital Services Act was not intended to cover AI systems, Botero Arcila[145] has proposed an interpretation that would extend the applicability of some of its provisions to at least large language models.

This conclusion is supported by a broad construction[146] of the notion of online search engines established by the Digital Services Act:

> an intermediary service that allows the user to formulate queries in order to search, in principle, all websites, or all websites in a particular language, on the basis of a query on any topic in the form of a keyword, voice prompt, phrase or other input, and that returns results in any format in which information related to the requested content can be found.[147]

By extensively interpreting this notion, some similarities can be found between the category of online search engines and that of large language models. The operation of both services depends on the initial query entered by their users and is aimed at retrieving information based on the input keywords.[148] Both services can return results *in any format in which* the information related to the requested content can be found. This feature may render it irrelevant that large language models do not provide any third-party content, but rather, contribute to its creation.

According to Botero Arcila, the fact that the definition of online search engines is agnostic to the forms in which a system delivers results may pave the way for directly applying some of the provisions of the Digital Services Act to large language models. This extension could have substantial effects in the case of large language models with a significant presence on the market, which could qualify as very large online search engines.

taken here. Rather, influential private entities need to assess and mitigate systemic risks posed by their systems. The requirements go beyond illegal content'.

[145]B. Botero Arcila, 'Is It a Platform? Is It a Search Engine? It's Chat GPT! The European Liability Regime for Large Language Models', 3(2) *Journal of Free Speech Law* (2023) p. 455.

[146]Ibid, p. 462.

[147]Art. 3(1)(j) DSA.

[148]According to Botero Arcila, *supra* n. 145, the systems in question would not present any editorial control, nor would they play an active role capable of founding effective content knowledge, relying merely on algorithms capable of performing predictive operations on the association between different words.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

Of course, there are many ontological differences between search engines and large language models: it is not by chance that large language models can be integrated into search engine services.[149] Large language models are less transparent than search engines due to their inherent opacity; nonetheless, they can perform better than search engines, which merely retrieve information from the Internet.

Thus, even though the Digital Services Act defines online search engines very broadly, theoretically offering some coverage for large language models, this was certainly not the *voluntas legislatoris*. After all, generative AI systems were almost unknown at the time the legislative process for the AI Act began, providing more evidence that they were not considered in the scope of the Digital Services Act. Having established that there is no solid basis for this interpretation, it is possible to discuss whether such an outcome would in any case be desirable and beneficial for freedom of expression.

As Botero Arcila has noted, treating large language models as online search engines for the purposes of the Digital Services Act would have two main practical consequences. I see these effects as a 'carrot' and a 'stick', whose combination may lead to interesting results.

The 'carrot': large language models would benefit from the application of the Digital Services Act by virtue of the extension of the liability exemption for illegal content under Chapter II for intermediary services.[150]

The 'stick': the extension of the Digital Services Act to large language models[151] would require the implementation of the risk assessment mechanisms under Article 34, if large language models with a significant presence in the market were considered very large online search engines.

With respect to the first consequence, various scholars have called for a selective extension of the Digital Services Act to generative AI systems. Hacker, Engel and Mauer[152] have emphasised the importance of applying the provisions on content moderation, such as the notice-and-action mechanism, to generative AI. The extension of these measures beyond the original scope of the Digital Services Act would be very impactful with respect to services that are provided as standalone solutions.[153] To the extent that generative AI systems are integrated into downstream services (such as caching and hosting services) that fall within the scope of the Digital Services Act, the practical consequences are likely to be

[149]For some similarities from a technical perspective, *see* N. Ziems et al., 'Large Language Models are Built-in Autoregressive Search Engines', Findings of the Association for Computational Linguistics: ACL 2023, 9-14 July 2023, p. 2666.

[150]*See* Art. 5.

[151]Botero Arcila, *supra* n. 145, p. 486-487.

[152]Hacker et al., *supra* n. 7.

[153]Ibid., p. 1120.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

quite limited, as such services are already subject to the liability exemptions and, in the case of hosting services, to the notice-and-action mechanism.[154]

Thinking about the impact of applying these rules to generative AI, it is precisely the reporting of content by users and trusted flaggers according to a notice-and-action mechanism that would make the adoption of the necessary corrective measures (by the large language model provider) possible. This mechanism would circumscribe to a pre-litigation phase the settlement of disputes depending not only on intentional manipulation but also on the inherent inaccuracy and fallacy of generative AI. Society also has an interest in taking advantage of AI technologies according to high quality standards, and such a goal demands adequate 'training' of AI systems,[155] even in their deployment and use.

This conclusion may appear paradoxical if one considers generative AI as engaged in a genuinely expressive activity or having an editorial responsibility for the output they produce.

However, this assumption should be reconsidered given the technical features that characterise generative AI systems, in which the processing of a given output, while closely related to the input received in terms of both training data sets and user queries, can hardly be the result of a proper editorial activity. Like hosting and caching providers, large language models have neither knowledge of nor control over the content they make available, which, however, is not just a selection of third-party input, but rather something more like their own product.

If it is then difficult to equate large language models with online search engines or other service providers, it is equally hard to qualify them as purely content providers. Perhaps the fact that, as these systems themselves make clear in their disclaimers, generative AI may produce inaccurate results, the correction of which would require some technical steps, could make the notice-and-action regime under the Digital Services Act an appropriate solution to minimise mistakes. The notice-and-action mechanism could be a more flexible and consistent with the state-of-the-art alternative to direct liability for illegal content 'made up' by generative AI. Indeed, the reporting of illegal content allows the provider or operator of a generative AI system to take necessary action and improve the

[154]In other terms, these provisions should apply equally to user-generated content and AI-generated content: ibid.

[155]Ibid., the authors observe that the enforcement of content moderation rules by generative AI systems could effectively rely on the combination of two components, one of centralised control and another of decentralised control, within the framework of the notice-and-action procedure (Art. 16 DSA). The first component would require the reporting of problematic content by users and by trusted flaggers. The second component would contemplate the direct involvement of AI providers and developers, supported by technical experts: it is precisely this phase that plays a vital role as it requires that, in view of the content reported by users and trusted flaggers, technical experts, such as engineers, act to modify the content generated by AI systems.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

performance of the models. In the end, attaching legal relevance to the reporting of the produced output as inaccurate or illegal could facilitate legal certainty, marking what under the Digital Services Act is the actual knowledge or awareness standard, a condition that triggers some legal obligations for the service provider.[156]

Regarding the second practical consequence of the extension of the Digital Service Act to generative AI, that is, the application of the risk management obligations under Article 34, this requirement could reciprocate the application of the favourable content moderation regime. However, the AI Act, while silent on the application of the content moderation framework, provides some clarification (mostly through some recitals) on the intersection between the new regulation and the Digital Services Act. The AI Act recognises that AI systems and models can be integrated into downstream services, which include very large online platforms and very large online search engines (Recitals 118 and 119). Given this possible scenario, the AI Act acknowledges that the relevant providers of very large online platforms or search engines are likely to have already implemented the risk management framework for the purposes of compliance with the AI Act in the fulfilment of their obligations set out in the Digital Services Act.[157] Recital 119 also adds that the Digital Services Act must be interpreted in a technology-neutral manner, further justifying that AI systems covered by the AI Act may be provided as intermediary services or parts thereof. This clarification is quite significant, as it may open an interpretive avenue for considering at least some generative AI systems to fall within the categories of intermediary services even when operated as standalone solutions.

Online content moderation in the US

Even in the US, defining the applicable legal regime for generative AI output seems to be far from an easy task. These difficulties are unsurprising given that Congress passed the Communications Decency Act, the first piece of legislation governing cyberspace, including the role and responsibilities of online platforms, in 1996.

Reflecting the paramount importance of freedom of speech in the US legal order, this provision aimed to relieve service providers from any consequence for engaging in content moderation, establishing a broad immunity from liability for

[156]Ibid., p. 486; the convergence on the point is also highlighted, albeit from a different perspective, by Volokh, *supra* n. 56, p. 514.

[157]However, this is without prejudice to any other systemic risk that may arise from the use of the relevant AI system or model and that is not covered by the DSA, which will require intermediaries to take action.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

third-party offensive material[158] and preventing them from being treated as speakers or publishers of third-party content.

However, this provision was passed when cyberspace was populated by a multitude of small virtual communities, and big tech companies were not yet on the scene. The current scheme of Section 230 of the Communications Decency Act does not provide any other option than the two in the alternative between information service providers and information content providers. So, generative AI systems will most likely fall into the scope of the latter category:[159] in such a scenario, their providers would not enjoy the broad immunity from liability enshrined in Section 230.

For some time now, there has been a major debate in the US about whether Section 230 should be revisited. This provision, which has received both criticism and praise from scholars, still reflects a nascent understanding of cyberspace and its actors, which faced overwhelming developments over the following two decades. Courts have adapted their enforcement of the Section 230 immunity to the evolving digital environment, but it seems unlikely that this provision can be extended beyond its original scope to encompass generative AI systems. Scholars like Volokh have noted that 'Congress didn't make the choice to immunize companies that deploy software which itself creates messages that had never been expressed by third parties';[160] likewise, Section 230 does not immunise companies that have materially contributed to the alleged unlawfulness, as in the case of statements harming someone's reputation that are made up by a software.[161] Against this view, it is argued that the output produced by systems such as ChatGPT is 'entirely composed of third-party information scraped from the web',[162] so that there is no creative activity that would transform the service provider into a content provider. However, the difficulties of applying Section 230 to something that in 1996 was unimaginable are quite visible.[163]

Amid the various challenges faced by Section 230 over nearly three decades is the question: Can the rise of generative AI pave the way for a reconsideration of

[158]Namely for 'material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected'. *See* 47 USC § 230(c)(2)(A).

[159]*See* M. Perault, 'Section 230 Won't Protect ChatGPT', 3 *Journal of Free Speech Law* (2023) p. 364 at p. 366-367. *Contra*, *see* J. Miers, 'Yes, Section 230 Should Protect ChatGPT and Other Generative AI Tools', *Techdirt*, 17 March 2023, https://www.techdirt.com/2023/03/17/yes-section-230-should-protect-chatgpt-and-others-generative-ai-tools/, visited 3 June 2025.

[160]*See* Volokh, *supra* n. 56, p. 495.

[161]Ibid.

[162]Miers, *supra* n. 159.

[163]*See also* P. Henderson et al., 'Where's the Liability in Harmful AI Speech?', 3 *Journal of Free Speech Law* (2023) p. 589 at p. 644.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

the provision's scope?[164] And even if generative AI systems were conceptualised as information content providers, would there be room to discuss the possible consequences in terms of attribution of liability, for example, in the case of defamatory content?[165]

To answer these questions, it is worth recalling that when Congress passed Section 230 of the Communications Decency Act in 1996, it did so to strategically protect information service providers from stricter liability standards (for example, that standard applicable to publishers which was applied in some pre-Section 230 case law[166]) in view of their vital role for the exercise of free speech rights[167] by individuals. Section 230 prevented service providers from being treated as speakers of information that they merely hosted or transmitted to eliminate the 'specter of tort liability in an area of such prolific speech',[168] which would have obvious chilling effects. However, the purpose of this provision was not to discourage content moderation, but to promote it. As highlighted by Henderson, Hashimoto and Lemley, 'one of the original purposes of Section 230 was to encourage proactive efforts to filter content by changing legal rules that got you in trouble if you intervened'.[169] This legal paradigm was simply found to be more consistent with their role as 'enablers' of free speech.

Similar considerations could now come into play in assessing which legal regime is more appropriate with respect to illegal content produced by generative AI: Henderson, Hashimoto and Lemley have emphasised that AI companies could now face a similar paradox, where those 'who put their head in the sand may be immune, but those who intervene to make things better may lose immunity'.[170] Should we expect a Section 230-like provision for generative AI systems?

The way forward

In the absence of crystal-clear legislative answers, the applicability of existing content moderation and liability rules in the context of generative AI will likely

[164]Despite the 'resilience' of Section 230 CDA in recent cases before the Supreme Court, where the degree of control and liability over the contested content implicated by the use of complex algorithmic procedures was at issue: *see Twitter, Inc.* v *Taamneh* and *Gonzalez* v *Google LLC*, *supra* n. 22.

[165]For an in-depth discussion of the ability of generative AI systems to 'commit' defamation in the US legal system, *see* Volokh, *supra* n. 56.

[166]*See* the '*Prodigy*' case, *Stratton Oakmont, Inc.* v *Prodigy Services Co.*, 23 Media L. Rep. 1794 (N.Y. Sup. Ct. 1995).

[167]The result, however, was immediately achieved, as the Zeran case very well proved: *Zeran* v *America Online, Inc.*, 129 F.3d 327 (4th Cir. 1997).

[168]Ibid.

[169]Henderson et al., *supra* n. 163, p. 645.

[170]Ibid.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

depend on judicial activism in potential future litigation. Although some scholars have suggested that the Digital Services Act and Section 230 of the Communications Decency Act could be broadly interpreted in a way that would allow the extension of the relevant provisions to generative AI systems (or to at least some of them), the current framing of these provisions does not make them very well suited to such innovative services. However, these outcomes may be partially desirable, most notably to the extent that they help strike a balance between the need to govern the early stages of this technology without hampering it and the need to promote a development of technology consistent with the protection of fundamental rights.

Some of these mechanisms could also be beneficial from a law and economics perspective and indirectly facilitate more competition in the market.[171] Currently, the generative AI market is predominantly populated by leading tech companies that are well positioned to bear the costs of potential litigation and the negative consequences resulting from the operation of their services. Strict liability rules, for example, may discourage startups from entering this market due to a high risk of litigation, which could constitute a significant barrier. In some ways, the rationale may be similar to the legislative option Congress took in 1996 when it enacted Section 230 of the Communications Decency Act: to remove a source of risk (that of being held directly liable as a publisher) that could ultimately discourage service providers from entertaining a business with key implications for fundamental rights. This not-so-unintentional lack of regulation may have seemed like a strategic option at the time, but it came at a significant cost in the medium term. Could a similar rationale guide lawmakers now, at such a crucial time for future developments in AI? One possible objection is that the AI market is already dominated by large technology companies, whereas at the time of the Communications Decency Act, none of these players had yet emerged.

However, given the still limited margin of accuracy of AI output, holding providers and/or deployers directly liable for any illegal content without any safe harbour regime could have chilling effects and, ultimately, undermine innovation in the relevant market. This is a problem that the AI Act generally addresses only to a certain extent: but will this be enough?

Finally, turning back to Europe, it should not be underestimated that, under EU law, the application of the content moderation regime enshrined in the Digital Services Act (i.e. liability exemptions combined with the notice-and-action mechanism) would come with the obligation to comply with accountability measures and other requirements, most notably in the case of

[171]For a specific and comprehensive analysis of the competition law perspective on generative AI, *see* F. Bostoen and A. van der Veer, 'Regulating Competition in Generative AI: A Matter of Trajectory, Timing and Tools', 2 *Concurrences* (2024) p. 27.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

services with a significant market presence that could qualify as very large online platforms or very large online search engines.[172] These risk mitigation measures, established by Article 34 and Article 35 of the Digital Services Act, may have a significant impact, among others, on the fight against disinformation to the extent its spread would amount to a systemic risk likely to affect the EU's fundamental rights and values. Therefore, the application of a more lenient liability regime for the output generated would not necessarily make generative AI operators less accountable and undermine the contrast between illegal and harmful content, which is at the heart of the Digital Services Act; on the contrary, it could prove to a certain extent to be desirable and beneficial.

Conclusion

Claiming constitutional coverage for AI-generated content under the umbrella of free speech may look like a provocation. It may even appear audacious given the huge amount of inaccurate information and fabricated content that is available by virtue of the use of these systems.

However, when it comes to the digital sphere, there is a common tendency to think of 'digital constitutionalism' primarily in terms of challenges to fundamental rights posed by private actors. Although the problem of horizontal effects of fundamental rights vis-à-vis online platforms is of key importance in the current debates, constitutional freedoms and rights first claim protection from governmental interferences. This is why looking at whether generative AI content should be constitutionally protected as speech provides some perspectives on how governments can limit the ability of individuals to receive information in the AI-driven society. As elaborated in this essay, extending constitutional coverage to generative AI content does not prevent state authorities from prohibiting certain types of speech that harm other legal interests (such as individuals' privacy or reputation), but such restrictions would not differ from any other limits on free speech that are constitutionally permissible in the offline world (and also online, via other media). As courts have made clear, particularly in the US, content-based restrictions should be strictly reviewed to prevent them from resulting in viewpoint discrimination.

Problems may arise when addressing content that is not illegal but may still qualify as harmful, as in the case of disinformation. However, these difficulties are

[172]Pursuant to Art. 14(4) DSA, intermediaries must act 'diligently, objectively and proportionately' in the context of content moderation 'with due regard for the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service, such as freedom of expression, freedom and pluralism of the media and other fundamental rights and freedoms enshrined in the Charter'.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press

not unique to generative AI; similar challenges exist in other domains, such as cyberspace. While generative AI and large language models increase the likelihood of individuals encountering false or inaccurate information (such as hallucinations), the core issue remains the same. Therefore, assessing whether mechanisms such as those introduced by the Digital Services Act for very large online platforms can be extended to generative AI systems is crucial in determining whether legal systems are adequately equipped to face its risk, including the spread of disinformation. This will be one of the key tests for the recently adopted AI Act, which does not offer specific solutions to the issues discussed in this article.

The AI Act only acknowledges that generative AI systems can be integrated into intermediary services covered by the Digital Services Act, so that there is alignment as to risk management measures required of the relevant operators under the two regulations. However, it seems that there is still a gap between the respective pieces of legislation regarding the liability exemptions for illegal content, which may turn out as a safeguard with the same value that Section 230 of the Communications Decency Act had at the rise of cyberspace. And this gap may have implications on the way content moderation is performed and, ultimately, the right of individuals to receive this speech, even without a speaker.

Marco Bassini is an Assistant Professor of Fundamental Rights and Artificial Intelligence at the Tilburg Institute for Law, Technology, and Society (TILT) – Tilburg University. The author is grateful to Ronald Leenes, Chiara Marchisotti, Giuseppe Martinico, Sunimal Mendis, Oreste Pollicino, Anuj Puri, Maurice Schellekens, Bart van der Sloot, Giulio Enea Vigevani, and Jiawei Zhang for their insightful comments on earlier drafts of this article. The usual disclaimers apply. This essay forms part of the research carried out under the 'RetrAIn – Enforcing Constitutional Rights in the Age of Generative Artificial Intelligence' Dutch government-funded starter grant, for which the author serves as principal investigator.

https://doi.org/10.1017/S1574019625100771 Published online by Cambridge University Press