

Artificial Intelligence and Media^{*}

*Lidia Dutkiewicz, Noémie Krack, Aleksandra Kuczerawy,
and Peggy Valcke*

14.1 INTRODUCTION

Media companies can benefit from artificial intelligence (AI)¹ technologies to increase productivity and explore new possibilities for producing, distributing, and reusing content. This chapter demonstrates the potential of the use of AI in media.² It takes a selective approach to showcase a variety of applications in the following areas: Can ChatGPT write news articles? How can media organizations use AI to recommend public interest content? Can AI spot disinformation and instead promote *trustworthy* news? These are just a few opportunities offered by AI at the different stages of news content production, distribution, and reuse (Section 14.2). However, the use of AI in media also brings societal and ethical risks, as well as legal challenges. The right to freedom of expression, media pluralism and media freedom, the right to nondiscrimination, and the right to data protection are among the affected rights. This chapter will therefore also show how the EU legal framework (e.g., the Digital Services Act,³ the AI Act,⁴ and the European Media Freedom

^{*} This chapter received funding from EU Horizon 2020 programme grants: n° 951962 MediaFutures and n° 951911 AI4Media and from FWO grants: nr. 1214321N and ALGEPI (FWOAL1088).

¹ For the definition of AI, see Chapter 1 of this book.

² This chapter takes a narrower understanding of media, focusing on traditional mass media outlets such as news media, public service media, as well as media archives. However, because of the impact which social media algorithmic content moderation practices have on media content distribution and editorial decision-making, they will also be covered in this chapter. For a broad understanding of the use of AI in the audiovisual sector, see, for example, Rehm, “The Use of Artificial Intelligence in the Audiovisual Sector: Concomitant Expertise for INI Report: Research for CULT Committee” (European Parliament, Directorate-General for Internal Policies of the Union), <https://data.europa.eu/doi/10.2861/294829>.

³ Regulation (EU) 2022/2065 of the European Parliament and of the Council of October 19, 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) 2022 (OJ L 277/1).

⁴ 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 (Artificial Intelligence Act) 2024 (OJ L 2024/1689).

Act⁵) tries to mitigate some of the risks to fundamental rights posed by the development and the use of AI in media (Section 14.3). Section 14.4 offers conclusions.

14.2 OPPORTUNITIES OF AI APPLICATIONS IN MEDIA⁶

14.2.1 *AI in Media Content Gathering and Production*

Beckett's survey of journalism and AI presents an impressive list of possible AI uses in day-to-day journalistic practice.⁷ At the beginning of the news-creating process, AI can help gather material, sift through social media, recognize genders and ages in images, or automatically add tags for newspaper articles with topics or keywords.⁸

AI is also used in story discovery to identify trends or spot stories that could otherwise be hard to grasp by the human eye and to discover new angles, voices, and content. To illustrate, already in 2014, Reuters News Tracer project used natural language processing techniques to decide which topics are *newsworthy*.⁹ It detected the bombing of hospitals in Aleppo and the terror attacks in Nice and Brussels before they were reported by other media.¹⁰ Another tool, the Topics Compass, developed under EU-funded Horizon 2020 ReTV project, allows an editorial team to track media discourse about a given topic coming from news agencies, blogs, and social media platforms and to visualize its popularity.¹¹

AI has also been proven useful in investigative journalism to assist journalists in tasks that could not be done by humans alone or would have taken a considerable amount of time. To illustrate, in a cross-border Panama Papers investigation, the International Consortium of Investigative Journalists used an open-source data mining tool to sift through 11.5 million whistleblowers' documents.¹²

Once journalists have gathered information on potential stories, they can use AI for the production of news items: text, images, and videos. Media companies such

⁵ Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act) 2024 (OJ L 2024/1083).

⁶ For a broad overview, see, for example, Filareti Tsalakanidou, "AI technologies and applications in media: State of play, foresight, and research directions" (2022) *AI4Media*, www.ai4media.eu/wp-content/uploads/2022/03/AI4Media_D2.3_Roadmap_final.pdf.

⁷ Beckett, "New powers, new responsibilities. A global survey of journalism and artificial intelligence" (*Blogs LSE*, November 18, 2019), <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>, accessed March 18, 2023.

⁸ *Ibid.*

⁹ Stray, "The age of the cyborg" (*Columbia Journalism Review*, November 30, 2016), www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php, accessed March 18, 2023.

¹⁰ *Ibid.*

¹¹ See, for example, ReTV, <https://retv-project.eu/news-discourse-monitoring/>.

¹² Guevara, "How Artificial Intelligence Can Help Us Crack More Panama Papers Stories" (*International Consortium of Investigative Journalists*, March 25, 2019), www.icij.org/inside-icij/2019/03/how-artificial-intelligence-can-help-us-crack-more-panama-papers-stories/, accessed March 18, 2023.

as the Associated Press, *Forbes*, and *The New York Times* have started to automate news content.¹³ Terms like *robot journalism*, *automated journalism*, and *algorithmic journalism* have been used interchangeably to describe this phenomenon.¹⁴ In addition, generative AI tools such as ChatGPT,¹⁵ Midjourney,¹⁶ or DALL-E¹⁷ are being used to illustrate news stories, simplify text for different audiences, summarize documents, or writing potential headlines.¹⁸

14.2.2 AI in Media Content Distribution¹⁹

Media organizations can also use AI for providing personalized recommendations. Simply put, “recommendation systems are tools designed to sift through the vast quantities of data available online and use algorithms to guide users toward a narrower selection of material, according to a set of criteria chosen by their developers.”²⁰

In recent years, online news media (e.g., online newspapers’ websites and apps) started engaging in *news recommendation* practices.²¹ Recommendation systems curate users’ news feed by automatically (de)prioritizing items to be displayed in user interfaces, thus deciding which ones are visible (to whom) and in what order.²²

The 2022 Ada Lovelace report²³ provides an informative in-depth snapshot of the BBC’s development and use of recommendation systems, which gives insights into the role of recommendations in public service media (PSM).²⁴ As pointed out by

¹³ Graefe, “Guide to Automated Journalism” (2016) *Columbia Journalism Review* www.cjr.org/tow_center_reports/guide_to_automated_journalism.php.

¹⁴ Although they do not have the same meaning. See, for example, Graefe, Guide to automated journalism, p. 3; Monti, “Automated journalism and freedom of information: Ethical and juridical problems related to AI in the press field” (2018) *Opinion Juris in Comparatione, Studies in Comparative and National law* 1; Dörr, “Mapping the field of algorithmic journalism” (2015) *Digital Journalism*.

¹⁵ See OpenAI, “Introducing ChatGPT,” <https://openai.com/blog/chatgpt>, accessed April 5, 2023.

¹⁶ “Midjourney,” www.midjourney.com/home/?callbackUrl=%2Fapp%2F, accessed April 5, 2023.

¹⁷ OpenAI, “DALL-E2,” <https://openai.com/product/dall-e-2>, accessed April 5, 2023.

¹⁸ See also Generative AI in the Newsroom, <https://generative-ai-newsroom.com/>, accessed April 5, 2023.

¹⁹ This section focuses on recommendation systems. Note that in this chapter, the terms “recommendation systems” and “recommender systems” are used interchangeably. For the broader discussion about AI and media content distribution, see, for example, Carlson, “Order versus access: news search engines and the challenge to traditional journalistic roles” (2007) *Media, Culture & Society*, 29(6): 1014–1030.

²⁰ Ada Lovelace Institute “Inform, educate, entertain ... and recommend? Exploring the use and ethics of recommendation systems in public service media” (2022), www.adalovelaceinstitute.org/report/inform-educate-entertain-recommend/.

²¹ Vermeulen “The Algorithmic State of Mind: A Human Rights Frame for Governing News Recommendation” (2022) (Ghent University, Faculty of Law and Criminology).

²² *Ibid.*

²³ Ada Lovelace Institute, “*Inform, educate, entertain ... and recommend?(...)*,” p. 4.

²⁴ See also PEACH, “Relevant content to the people, crafted by broadcasters for broadcasters. Personalisation and Recommendation Ecosystem for the digital transformation,” <https://peach.ebu.io/>, accessed April 5, 2023.

the authors, developing recommendation systems for PSM requires an interrogation of the organizations' role in democratic societies in the digital age, that is, how to translate the public service values²⁵ into the objectives for the use of recommendation systems that serve the public interest. The report concludes that the PSM had internalized a set of normative values around recommendation systems: Rather than maximizing engagement, they want to broaden their reach to a more diverse set of audiences.²⁶ This is a considerable difference between the public and private sectors. Many user-generated content platforms rank information based on how likely a user is to interact with a post (comment on it, like it, reshare it) or to spend more time using the service.²⁷

Research shows that social media platforms are using a mix of commercial criteria, but also vague public interest considerations in their content prioritization measures.²⁸ Importantly, prioritization of some content demotes other.²⁹ As a way of example, Facebook explicitly says it will not recommend *content that is associated with low-quality publishing*, including news that it is unclear about its provenance.³⁰ In fact, online platforms use a whole arsenal of techniques to (de)amplify the visibility or reach of some content.³¹ To illustrate, in an aftermath of Russian aggression on Ukraine, platforms announced they would *restrict access* to RT and Sputnik media outlets.³² Others have also been adding labels and started reducing the visibility of content from Russian state-affiliated media websites even before the EU-imposed sanctions.³³

²⁵ Public service media organizations are legally mandated to operate with a particular set of public interest values. The EBU has codified the public service mission into six core values: universality, independence, excellence, diversity, accountability, and innovation, and member organizations commit to strive to uphold these in practice. See EBU, "Empowering society, a declaration on the core values of public service media," www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf.

²⁶ Ada Lovelace Institute, "Inform, educate, entertain ... and recommend?(...)," p. 4.

²⁷ See, for example, Mosseri, "Shedding More Light on How Instagram Works" *AboutInstagram.com* (June 8, 2021), <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>, accessed March 22, 2023.

²⁸ CMPF-CiTiP-IViR-SMIT, Study on Media Plurality and Diversity Online, CNECT/2020/OP/0099, May 2022, <https://digital-strategy.ec.europa.eu/en/library/study-media-plurality-and-diversity-online>, accessed April 5.

²⁹ Keller uses the term "demote" to cover any form of deamplification, including decreasing content's algorithmic ranking or excluding it from features like recommendations. Keller, "Amplification and Its Discontents." *Knight First Amendment Institute at Columbia University* (June 8, 2021), <https://knightcolumbia.org/content/amplification-and-its-discontents>, accessed March 22, 2023.

³⁰ Facebook, "What are recommendations on Facebook?" *Facebook Help Center*, www.facebook.com/help/1257205004624246, accessed April 5, 2023.

³¹ See, for example, Goldman, "Content Moderation Remedies" (2021) *Mich. Tech. L. Rev.*, 28: 1.

³² Kayali, "Facebook's Parent Company Restricts EU Access to Russia's RT, Sputnik" *Politico* (February 28, 2022), www.politico.eu/article/facebook-parent-company-restricts-eu-access-to-russia-rt-sputnik/, accessed April 5, 2023.

³³ Culliford "Twitter Will Label, Reduce Visibility of Tweets Linking to Russian State Media" *Reuters* (February 28, 2022), www.reuters.com/technology/twitter-will-label-reduce-visibility-tweets-linking-russian-state-media-2022-02-28/, accessed January 17, 2023.

Overall, by selecting and (de)prioritizing news content and deciding on its visibility, online platforms take on some of the functions so far reserved to traditional media.³⁴ Ranking functions and optimization metrics in recommendation systems have become powerful determinants of access to media and news content.³⁵ This has consequences for both the fundamental right to freedom of expression and media freedom (see Section 14.3).

14.2.3 AI in Fact-Checking

Another important AI potential in media is fact-checking. The main elements of automated fact-checking are: (1) identification of false or questionable claims circulating online; (2) verification of such claims, and (3) (real-time) correction (e.g., flagging).³⁶

To illustrate, platforms such as DALIL help fact-checkers spot questionable claims which then require subsequent verification.³⁷ Then, to verify the identified content, the AI(-enhanced) tools can perform a reverse image search, detect bot accounts and deep fakes, assess source credibility, check nonfactual statements (claims) made on social media or analyze the relationships between accounts.³⁸ WeVerify plug-in is a highly successful tool which offers a variety of verification and analysis features in one platform to fact-check and analyze images, video, and text.³⁹ Some advanced processing and analytics methods can also be used to analyze different types of content and apply a trustworthiness scoring to online articles.⁴⁰

The verified mis- or disinformation can then be flagged to the end user by adding warnings and providing more context to content rated by fact-checkers. Some platforms have also been labeling content containing *synthetic and manipulated media*.⁴¹

³⁴ Council of Europe, “Guidance Note on the Prioritisation of Public Interest Content Online adopted by the Steering Committee for Media and Information Society (CDMSI) at its 20th plenary meeting, December 1–3, 2021,” <https://rm.coe.int/cdmsi-2021-009-guidance-note-on-the-prioritisation-of-pi-content-e-ado/1680a524c4>, accessed April 5, 2023.

³⁵ Ibid.

³⁶ Graves, “Understanding the promise and limits of automated fact-checking” (*Reuters Institute for the Study of Journalism*, February 2018), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf, accessed April 4, 2023.

³⁷ EU Neighbours South, “AI-driven platform launched to accelerate Arabic language fact-checking” (January 2, 2023), <https://south.euneighbours.eu/news/ai-driven-platform-launched-to-accelerate-arabic-language-fact-checking/>, accessed April 4, 2023.

³⁸ DW Innovation, “AI for Content Verification I: Status Quo and Current Limitations” (*DW Innovation* October 24, 2022), <https://innovation.dw.com/articles/ai-for-content-verification-i-status-quo-and-current-limitations>, accessed April 4, 2023.

³⁹ See WeVerify, weverify.eu/verification-plugin/, accessed April 5, 2023.

⁴⁰ Nucci et al., “Artificial Intelligence against Disinformation: The FANDANGO Practical Case (short paper)” (International Forum on Digital and Democracy (IFDaD), Venice, Italy, 2020).

⁴¹ Twitter, “Synthetic and manipulated media policy,” *Twitter Help Centre*, <https://help.twitter.com/en/rules-and-policies/manipulated-media>, accessed April 5, 2023.

Countering disinformation with the use of AI is a growing research area. The future solutions based on natural language processing, machine learning, or knowledge representation are expected to deal with different content types (audio, video, images, and text) across different languages.⁴² Collaborative tools that enable users to work together to find, organize, and verify user-generated content are also on the rise.⁴³

14.2.4 AI in Content Moderation

AI in content moderation is a broad topic. Algorithmic (commercial) content moderation can be defined as “systems that classify user-generated content based on either matching or prediction, leading to a decision and governance outcome (e.g., removal, geoblocking, and account takedown).”⁴⁴ This section focuses on the instances where AI is used either by media organizations to moderate the discussion on their own sites (i.e., in the comments section) or by social media platforms to moderate posts of media organizations and journalists.

14.2.4.1 Comment Moderation

For both editorial and commercial reasons, many online news websites have a dedicated space under their articles (a comment section), which provides a forum for public discourse and aims to engage readers with the content. Empirical research shows that a significant proportion of online comments are *uncivil* (featuring a disrespectful tone, mean-spirited, disparaging remarks and profanity),⁴⁵ and encompass stereotypes, homophobic, racist, sexist, and xenophobic terms that may amount to hate speech.⁴⁶ The rise of incivility in online news comments negatively affects people’s perceptions of news article quality and increases hostility.⁴⁷ “Don’t read the comments” has become a mantra throughout the media.⁴⁸ The amount of hateful and racist comments, together with high costs – both economic and psychological – of human moderators, has prompted news sites to change their practices.

⁴² See, for example, vera.ai, www.veraai.eu/home, accessed April 5, 2023.

⁴³ See Truly Media, www.truly.media, accessed April 5, 2023. See also AI4Media, “UC1: AI for Social Media and Against Disinformation,” *AI4Media*, www.ai4media.eu/uc1-ai-for-social-media-and-against-disinformation/, accessed April 4, 2023.

⁴⁴ Gorwa, Binns, and Katzenbach, “Algorithmic content moderation: Technical and political challenges in the automation of platform governance” (2020) *Big Data & Society*, 7.

⁴⁵ Coe, Kenski, and Rains “Online and uncivil? Patterns and determinants of incivility in newspaper website comments” (2014) *Journal of Communication*, 64: 658.

⁴⁶ Che, *Online Incivility and Public Debate: Nasty Talk* (Springer International Publishing AG, 2017).

⁴⁷ Kathleen Searles, Sophie Spencer, and Adaobi Duru, “Don’t read the comments: The effects of abusive comments on perceptions of women authors’ credibility” *Information, Communication & Society*, 23(7).

⁴⁸ Gardiner, “‘It’s a terrible way to go to work’: What 70 million readers’ comments on the Guardian revealed about hostility to women and minorities online” (2018) *Feminist Media Studies*, 18(4): 592–608.

Some introduced AI systems to support their moderation processes. To illustrate, both the New York Times⁴⁹ and the Washington Post⁵⁰ use machine learning to prioritize comments which are then evaluated by human moderators or to automatically approve or delete abusive comments. Similarly, STANDARD Community (part of the Austrian newspaper DerSTANDARD) has developed an automated system to prefilter problematic content, as well as a set of preemptive moderation techniques, including forum design changes to prevent problematic content from being posted in the first place.⁵¹

Others, like Reuters or CNN, have removed their comment sections completely.⁵² Apart from abusive and hateful language, the reason was that many users were increasingly commenting on media organizations' social media profiles (e.g., on Facebook), and not on media organizations' websites.⁵³ This, however, did not remove the problem of hateful speech. To the contrary, it amplified it.⁵⁴

14.2.4.2 Content Moderation

Online intermediary services (e.g., online platforms such as social media) can, and sometimes have to, moderate content which users post on their platforms. In the EU, to avoid liability for illegal content hosted on their platforms, online intermediaries must remove or disable access to such content when the illegal character of the content becomes known. Other content moderation decisions are performed by platforms voluntarily, based on platforms' community standards, that is, private rules drafted and enforced by the platforms (referred to as *private ordering*).⁵⁵ Platforms can therefore remove users' content which they do not want to host according to their terms and conditions, even if the content is not illegal. This includes legal editorial content of media organizations (see Section 14.3.4).

⁴⁹ Traub, "Why Humans, Not Machines, Make the Tough Calls on Comments" *The New York Times* (October 26, 2021), www.nytimes.com/2021/10/26/insider/why-humans-not-machines-make-the-tough-calls-on-comments.html, accessed April 5, 2023.

⁵⁰ WashPostPR, "The Washington Post leverages artificial intelligence in comment moderation" *The Washington Post* (June 22, 2017), www.washingtonpost.com/pr/wp/2017/06/22/the-washington-post-leverages-artificial-intelligence-in-comment-moderation/, accessed April 5, 2023.

⁵¹ Wagner, Kübler, Pirková, Gsenger, and Ferro "Reimagining content moderation and safeguarding fundamental rights. A study on community-led platforms" *The Greens/EFA in the European Parliament* (May 3, 2021), www.greens-efa.eu/files/assets/docs/alternative_content_web.pdf, accessed April 4, 2023.

⁵² Liu and McLeod, "Pathways to news commenting and the removal of the comment system on news websites" (2021) *Journalism*, 22(4): 867–881.

⁵³ Ibid.

⁵⁴ United Nations, "Hate Speech: Turning the tide" *UN News, Global perspective Human stories* (January 30, 2023), <https://news.un.org/en/story/2023/01/1132617>, accessed April 5, 2023; Munn, "Angry by design: Toxic communication and technical architectures." (2020) *Humanities and Social Sciences Communications*, 7: 53.

⁵⁵ Belli and Venturini, "Private ordering and the rise of terms of service as cyber-regulation" (2016) *Internet Policy Review*, 5(4).

Given the amounts of content uploaded on the Internet every day, it has become impossible to identify and remove illegal or unwanted content using only traditional human moderation.⁵⁶ Many platforms have therefore turned to AI-based content moderation. Such automation can be used as proactive detection of potentially problematic content prior to its publication or as a reactive moderation after it has been flagged by other users or automated processes.⁵⁷ Besides deleting content and suspending users, platforms use a whole arsenal of tools to reduce the visibility or reach of some content, such as age barriers, geo-blocking, labeling content as fact-checked or adding a *graphic content* label to problematic content before or as users encounter it.⁵⁸

Algorithmic moderation systems help classify user-generated content based on either matching or prediction techniques.⁵⁹ These techniques present a number of technical limitations.⁶⁰ Moreover, speech evaluation is highly context dependent, requiring an understanding of cultural, linguistic, and political nuances as well underlying facts. As a result, AI is frequently inaccurate; there is growing empirical evidence of platforms' over-removal of content coming from individuals and media organizations (see Section 14.3.4).⁶¹

14.3 LEGAL AND ETHICAL CHALLENGES OF AI APPLICATIONS IN MEDIA

This section identifies the legal and ethical challenges of AI in media across various stages of the media value chain described earlier. The section also shows how these challenges may be mitigated by the EU legal framework.⁶²

⁵⁶ Llansó et al., "Artificial Intelligence, Content Moderation, and Freedom of Expression" (Working Papers from the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, 2020), www.ivir.nl/publicaties/download/AI-Llanso-Van-Hobken-Feb-2020.pdf, accessed April 5, 2023.

⁵⁷ Cambridge Consultants, "Use of AI in online content moderation" (Report produced on behalf of Ofcom 2019), www.cambridgeconsultants.com/sites/default/files/uploaded-pdfs/Use%20of%20AI%20in%20online%20content%20moderation.pdf, accessed April 5, 2023.

⁵⁸ Goldman, "Content moderation remedies" (2021) *Michigan Technology Law Review*, 28(1): 1–59.

⁵⁹ Gorwa, Binns, and Katzenbach, "Algorithmic content moderation: Technical and political challenges in the automation of platform governance," p. 6.

⁶⁰ Llansó et al., *Artificial Intelligence, Content Moderation, and Freedom of Expression*.

⁶¹ Keller and Leerssen, "Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation" in Persily and Tucker (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform* (SSRC *Anxieties of Democracy* (Cambridge University Press), 220–251.

⁶² The issues of attribution of responsibility for automated content between the journalist, editor, media organization, and AI system providers, as well as liability regarding AI systems, fall outside of the scope of this chapter. See Chapter 6 AI and Responsibility and Chapter 8 AI and Liability Law for more information. The challenges related to how to assign authorship or copyright to an automated article are also left out. See Chapter 12 AI and IP Law.

14.3.1 *Lack of Data Availability*

Lack of data availability is a cross-cutting theme, with serious consequences for the media sector. Datasets are often inaccessible or expensive to gather and data journalists rely on private actors, such as data brokers which have already collected such data.⁶³ This concentrated control over the data influences how editorial decision-making is automated (see Section 14.3.6).

Data availability is also of paramount importance for news verification and fact-checking activities. Access to social media data is vital to analyze and mitigate the harms resulting from disinformation, political microtargeting, or the effect of social media on elections or children's well-being.⁶⁴ This is because it enables journalists and researchers to hold platforms accountable for the working of their AI systems. Equally, access to, for example, social media data is important for media organizations that are developing their own AI solutions – particularly in countries where it can be difficult to gain access to large quantities of data in the local language.⁶⁵

Access to platforms' data for researchers is currently mainly governed by contractual agreements, platforms' own terms of service, and public application programming interfaces (APIs). Application programming interfaces access can be restricted or eliminated at any time and for any reason.⁶⁶ The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression stressed a lack of transparency and access to data as “the major failings of companies across almost all the concerns in relation to disinformation and misinformation.”⁶⁷

A key challenge for research access frameworks is to comply with the General Data Protection Regulation (GDPR).⁶⁸ Despite a specific derogation for scientific

⁶³ van Drunen and Fechner, “Safeguarding Editorial Independence in an Automated Media System: The Relationship between Law and Journalistic Perspectives” (2022) *Digital Journalism*.

⁶⁴ Pasquetto et al., “Tackling misinformation: What researchers could do with social media data” (2020) *Harvard Kennedy School Misinformation Review*, 1(8): 01–14; Ausloos, Leerssen, and ten Thije, “Operationalizing Research Access in Platform Governance: What to Learn from Other Industries?” *Algorithm Watch* (June 25, 2020), www.ivir.nl/publicaties/download/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf, accessed March 23, 2023.

⁶⁵ Bocyte, Krack, Dutkiewicz, Schjøtt Hansen, ‘Blog series: More policies and initiatives need to support responsible AI practices in the media’ Medium (July 29, 2024), medium.com/ai-media-observatory/blog-series-more-policies-and-initiatives-need-to-support-responsible-ai-practices-in-the-media-2a42d271d1e1, accessed July 29, 2024.

⁶⁶ See, for example, “We research misinformation on Facebook. It just disabled our accounts” *The New York Times* (August 10, 2021), www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html?referringSource=articleShare, accessed April 5, 2023; Nicolas Kayser-Bril, “AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook” *Algorithm Watch* (August 13, 2021), <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>, accessed April 5, 2023.

⁶⁷ Khan, “Disinformation and freedom of opinion and expression: report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression” (April 13, 2021).

⁶⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of natural persons with regard to the processing of personal data and on the free movement

research purposes (art. 89), the GDPR lacks clarity regarding how platforms might share data with researchers (e.g., on what legal grounds).⁶⁹ To mitigate this uncertainty, various policy and regulatory initiatives aim to clarify how platforms may provide access to data to researchers in a GDPR-compliant manner.⁷⁰ In addition, there have been calls for a legally binding mechanism that provides independent researchers with access to different types of platform data.⁷¹

The Digital Services Act (DSA) requires providers of very large online platforms (VLOPs) and very large online search engines (VLOSEs) to grant *vetted researchers* access to data, subject to certain conditions.⁷² Data can be provided “for the sole purpose” of conducting research that contributes to the detection, identification and understanding of systemic risks and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures (art. 40(4)). Vetted researchers must meet certain criteria and procedural requirements in the application process. Importantly, they must be affiliated to a research organization or a not-for-profit body, organization or association (art. 40(12)). Arguably, this excludes unaffiliated media practitioners, such as freelance journalists or bloggers. Many details about researchers’ access to data through the DSA will be decided in delegated acts that have yet to be adopted (art. 14(13)).

Moreover, under the Digital Markets Act,⁷³ the so-called *gatekeepers* will have to provide advertisers and publishers with access to the advertising data and allow business users to access the data generated in the context of the use of the core platform service (art. 6(1) and art. 6(8)).

Furthermore, the European strategy for data⁷⁴ aims at creating a single market for data by establishing common European data spaces to make more data available

of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119. See also Chapter 9 AI and Privacy Law.

⁶⁹ Dutkiewicz “From the DSA to Media Data Space: the possible solutions for the access to platforms’ data to tackle disinformation” *European Law Blog* (October 19, 2021), <https://europeanlawblog.eu/2021/10/19/from-the-dsa-to-media-data-space-the-possible-solutions-for-the-access-to-platforms-data-to-tackle-disinformation/>, accessed March 13, 2023.

⁷⁰ See, for instance, the European Digital Media Observatory, “Report of the European Digital Media Observatory’s Working Group on Platform-to-Researcher Data Access” (May 31, 2022), <https://edmoprod.wpengine.com/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf>, accessed April 4, 2023.

⁷¹ Vermuelen, “The Keys to the Kingdom.” *Knight First Amendment Institute* (July 27, 2021), <https://knightcolumbia.org/content/the-keys-to-the-kingdom>, accessed March 20, 2023.

⁷² Digital Services Act, art. 40. See also Albert, “A guide to the EU’s new rules for researcher access to platform data” *Algorithm Watch* (December 7, 2022), <https://algorithmwatch.org/en/dsa-data-access-explained/>, accessed April 5, 2023.

⁷³ Regulation (EU) 2022/1925 of the European Parliament and of the Council of September 14, 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) [2022] OJ L265.

⁷⁴ European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data [COM/2020/66 final].

for use in the economy and society. The Data Governance Act⁷⁵ and the Data Act proposal⁷⁶ seek to strengthen mechanisms to increase data availability and harness the potential of industrial data, respectively. Lastly, the European Commission announced the creation of a dedicated media data space.⁷⁷ The media data space initiative, financed through the Horizon Europe and Digital Europe Programmes,⁷⁸ aims to support both PSM and commercial media operators to pool their content and customer data to develop innovative solutions.

14.3.2 Data Quality and Bias in Training Datasets

Another, closely related, consideration is data quality. There is a growing literature on the quality and representation issues with training, testing, and validation data, especially those in publicly available datasets and databases.⁷⁹ Moreover, generative AI raises controversies regarding the GDPR compliance of the training data⁸⁰ and brings a broader question of *extraction fairness*, defined as “legal and moral concerns regarding the large-scale exploitation of training data without the knowledge, authorization, acknowledgement or compensation of their creators.”⁸¹

The quality of training data and data annotation is crucial, for example, for hate speech and abusive language detection in comments. A 2022 report by the EU Agency for Fundamental Rights shows how tools that automatically detect or *predict* potential online hatred can produce biased results.⁸² The predictions frequently overreact to various identity terms (i.e., words indicating group identities like ethnic

⁷⁵ Regulation (EU) 2022/868 of the European Parliament and of the Council of May 30, 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) OJ L152.

⁷⁶ European Commission, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) [COM/2022/68 final].

⁷⁷ European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Europe’s Media in the Digital Decade: An Action Plan to Support Recovery and Transformation (December 3, 2020, COM/2020/784 final).

⁷⁸ In particular the Cloud Data and TEF Call (DIGITAL-2022-CLOUD-AI-03).

⁷⁹ See, for example, Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton, “Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (AIES ’20). Association for Computing Machinery, New York, NY, USA, 145–151; Osonde Osoba and William Welser IV, “An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence” (RAND Corporation, 2017).

⁸⁰ See, for instance, “Artificial intelligence: Stop to ChatGPT by the Italian SA Personal data is collected unlawfully, no age verification system is in place for children” (March 31, 2023), www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english, accessed April 5, 2023.

⁸¹ Helberger and Diakopoulos, “ChatGPT and the AI Act” (2023) *Internet Policy Review*, 12(1). For AI and fairness, see Chapter 5 of this book.

⁸² European Union Agency for Fundamental Rights, “Bias in Algorithms – Artificial Intelligence and Discrimination” (Publications Office of the European Union, 2022).

origin or religion), flagging text that is not actually offensive.⁸³ Research shows that social media content moderation algorithms have difficulty differentiating hate speech from discussion about race and often silence marginalized groups such as racial and ethnic minorities.⁸⁴ At the same time, underrepresentation of certain groups in a training dataset may result in them experiencing more abusive language than other groups.

There are blurred lines between what constitutes *hateful*, *harmful*, and *offensive* speech, and these notions are context dependent and culturally specific. Many instances of hate speech cannot be identified and distinguished from innocent messages by looking at single words or combinations of them.⁸⁵ Such contextual differentiation, between, for example, satirical and offensive uses of a word proves challenging for an AI system. This is an important technical limitation that may lead to over- and under-removal of content. Both can interfere with a range of fundamental rights such as the right to freedom of expression⁸⁶ (see Section 14.3.4), the right to data protection, as well as the right to nondiscrimination.

The consequence of using unreliable data could be the spread of misinformation⁸⁷ as illustrated by inaccurate responses to news queries from search engines using generative AI. Research into Bing's generative AI accuracy for news queries shows that there are detail errors and attribution errors, and the system also sometimes asserts the opposite of the truth.⁸⁸ This, together with the lack of media literacy, may cause an automation bias, that is, the uncritical trust in information provided by the automated system despite the information being actually incorrect.

⁸³ Ibid.

⁸⁴ Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner, "Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas" (2021) *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 466.

⁸⁵ Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, "The impact of algorithms for online content filtering or moderation 'Upload filters'" (Study Requested by the JURI Committee, September 2020).

⁸⁶ See, for example, Helberger, van Drunen, Eskens, Bastian, and Moeller, "A freedom of expression perspective on AI in the media – with a special focus on editorial decision making on social media platforms and in the news media" (2020) *European Journal of Law and Technology*, 11(3); Krack, Beudels, Valcke, and Kuczerawy, "AI in the Belgian Media Landscape. When Fundamental Risks Meet Regulatory Complexities," *Artificial Intelligence and the Law*, vol 13 (2nd rev ed., Jan De Bruyne and Cedric Vanleenhove (eds), *Intersentia*, 2023).

⁸⁷ Misinformation, as opposed to disinformation is not deliberate. The EU defines it as "false or misleading content shared without harmful intent though the effects can be still harmful," see "Tackling online disinformation" *European Commission* (June 29, 2022), <https://digital-strategy.ec.europa.eu/en/policies/online-disinformation>, accessed April 5, 2023.

⁸⁸ Diakopoulos, "Can We Trust Search Engines with Generative AI? A Closer Look at Bing's Accuracy for News Queries" *Medium* (February 17, 2023), <https://medium.com/@ndiakopoulos/can-we-trust-search-engines-with-generative-ai-a-closer-look-at-bings-accuracy-for-news-queries-179467806bcc>, accessed April 5, 2023.

14.3.3 Transparency

Transparency can mean many different things. Broadly speaking, it should enable people to understand how an AI system is developed, is trained, how it operates, and how it is deployed so that they can make more informed choices.⁸⁹ This section focuses on three aspects of transparency of AI in media.⁹⁰

The *first* aspect relates to *internal transparency*, which describes the need for journalists and other non-technical groups inside media organizations to have sufficient knowledge around the AI systems they use.⁹¹ The importance of closing the intelligibility gap around AI within a media organization is necessary for an understanding of how AI systems work to use them responsibly.⁹² The AI Act requires providers and deployers of AI systems including media organizations to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operations and use of AI systems on their behalf (art. 4).⁹³

The *second* aspect concerns *external transparency*, which refers to transparency practices directed toward the audience to make them aware of the use of AI. The AI Act requires providers of AI systems, such as OpenAI, to make it clear to users that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use (art. 50(1)).⁹⁴ As a rule, they must also mark generative AI outputs (synthetic audio, image, video or text content) as AI-generated or manipulated (art. 50(2)). For now it remains unclear what forms of transparency will be sufficient and whether they will be meaningful to the audience. Transparency requirements also apply to those who use AI systems that generate

⁸⁹ “Transparency and explainability (Principle 1.3)” OECD, <https://oecd.ai/en/dashboards/ai-principles/P7>, accessed April 5, 2023.

⁹⁰ Drawing on the distinction made by Cools and Koliska. See: Hannes Cools, Michael Koliska, “News Automation and Algorithmic Transparency in the Newsroom: The Case of the Washington Post” (2024) *Journalism Studies* 25(6): 662–80.

⁹¹ Schjøtt Hansen, Bocyte, Krack, Dutkiewicz, “Blog series: AI regulation is overlooking the need for third-party transparency in the media sector,” *Medium* (July 15, 2024), medium.com/ai-media-observatory/blog-series-ai-regulation-is-overlooking-the-need-for-third-party-transparency-in-the-media-sector-df84318c1cfa, accessed July 31, 2024.

⁹² Jones Bronwyn, Rhianne Jones, and Ewa Luger, “AI ‘Everywhere and Nowhere’: Addressing the AI Intelligibility Problem in Public Service Journalism” (2022) *Digital Journalism*, 10 (10): 1731–55. doi: 10.1080/21670811.2022.2145328.

⁹³ The AI Act does not provide details on whether and how (media) organizations will be supported in this work. It has been pointed out that supporting AI literacy in media organizations is highly resource-intensive and takes a lot of translational work. See, for example, DW Innovation, “AI in Media Tools: How to Increase User Trust and Support AI Governance” (June 21, 2024), innovation.dw.com/articles/ai-media-tools-user-trust, accessed July 31, 2024.

⁹⁴ What “interaction” means in this context is unclear, but it could cover applications such as chatbots, newsbots, recommender systems, and automated writing systems. See: Helberger and Diakopoulos, “The European AI Act and How It Matters for Research into AI in Media and Journalism” (2022) *Digital Journalism*.

or manipulate images, audio or video content constituting a deep fake (art. 50(4) para 1). However, if the content is part of an evidently artistic, creative or satirical work, the disclosure should not hamper the display or enjoyment of the work. Moreover, deployers of an AI-generated or manipulated text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. There is an important exception for the media sector. If the AI-generated text has undergone a process of human review or editorial control within an organisation that holds editorial responsibility for the content (such as a publisher), disclosure is no longer necessary. This provision raises questions as to what will count as a *human review* or *editorial control* and who can be said to hold *editorial responsibility*.⁹⁵ Moreover, research shows that audiences want media organisations to be transparent and provide labels when using AI.⁹⁶

In addition to the AI Act, the DSA presents multiple layers of transparency obligations for the benefit of users that differ depending on the type of service concerned.⁹⁷ In particular, it requires transparency on whether AI is used in content moderation. All intermediary services must publish in their terms and conditions, in a “clear and unambiguous language,” a description of the tools used for content moderation, including AI systems that either automate or support content moderation practices (art. 14). In practice, this means that users must know why, when, and how online content is being moderated, including with the use of AI, and when human review is in place.

The DSA also regulates recommender system transparency. As mentioned earlier, recommender systems can have a significant impact on the ability of recipients to retrieve and interact with information online. Consequently, providers of online platforms are expected to set out in their terms and conditions in plain and intelligible language the main parameters used in their recommender systems and the options for users to modify or influence them (art. 27). The main parameters shall explain why certain information is suggested, and include, at least, the criteria that are most significant in determining the information suggested, and the reasons for the relative importance of those parameters. There are additional requirements

⁹⁵ Schjøtt Hansen, Bocyte, Krack, Dutkiewicz, ‘Blog series: AI regulation is overlooking the need for third-party transparency in the media sector’ *Medium* (July 15, 2024), medium.com/ai-media-observatory/blog-series-ai-regulation-is-overlooking-the-need-for-third-party-transparency-in-the-media-sector-df843118c1fa, accessed July 31, 2024.

⁹⁶ Fletcher, Kleis Nielsen, ‘What does the public in six countries think of generative AI in news?’ (May 28, 2024), reutersinstitute.politics.ox.ac.uk/what-does-public-six-countries-think-generative-ai-news#header-6, accessed July 31, 2024.

⁹⁷ Other transparency requirements include, for example, an obligation for VLOPs and VLOSEs to explain the design, the logic, the functioning and the testing of their algorithmic systems, including their recommender systems as well as transparency of online advertising. See Digital Services Act art. 40(3) and art. 26, respectively. See also Krack, Beudels, Valcke and Kuczerawy, *AI in the Belgian Media Landscape. When Fundamental Risks Meet Regulatory Complexities*.

imposed on the providers of VLOPs and VLOSEs to provide at least one option for their recommendation systems which is not based on profiling.

There are also further obligations for VLOPs and VLOSEs to perform an assessment of any systemic risks stemming from the design, functioning, or use of their services, including algorithmic systems (art. 34(1)). This risk assessment shall include the assessment of any actual or foreseeable negative effects on the exercise of fundamental rights, including the right to freedom of expression and the freedom and pluralism of the media (art. 34(1)(b)). When conducting risk assessments, VLOPs and VLOSEs shall consider, in particular, whether the design of their recommender systems and their content moderation systems influence any of the systemic risks. If so, they must put in place mitigation measures, such as testing and adapting their algorithms (art. 35).

Lastly, intermediary services (excluding micro and small enterprises) must publish, at least once a year, transparency reports on their content moderation activities, including a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error of the automated means (art. 15). Extra transparency reporting obligations apply to VLOPs (art. 42).

The *third* aspect concerns *third-party transparency*, which refers to the importance of having insights into how AI systems provided by third-party providers have been trained on and how they work.⁹⁸

In both the DSA and the AI Act, there are no explicit provisions that make such information widely available.⁹⁹

14.3.4 Risks for the Right to Freedom of Expression

Article 10 of the European Convention of Human Rights (ECHR), as well as Article 11 of the Charter of Fundamental Rights of the European Union (CFR),¹⁰⁰ guarantees the right to freedom of expression to everyone. The European Court of Human Rights (ECtHR) has interpreted the scope of Article 10 ECHR through an extensive body of case law. The right to freedom of expression includes the right to impart information, as well as the right to receive it. It protects the rights of individuals, companies, and organizations, with a special role reserved for media organizations and journalists. It is their task to inform the public about matters of public interest

⁹⁸ Schjøtt Hansen, Bocyte, Krack, Dutkiewicz, “Blog series: AI regulation is overlooking the need for third-party transparency in the media sector” *Medium* (July 15, 2024), medium.com/ai-media-observatory/blog-series-ai-regulation-is-overlooking-the-need-for-third-party-transparency-in-the-media-sector-df843118c1fa, accessed July 31, 2024.

⁹⁹ In the AI Act, requirements to provide some information about the training datasets and documentation around the capabilities and limitations of AI models only apply to general-purpose AI models or high-risk AI systems (see Recitals 66-67, Article 53, and Annex XII AI Act).

¹⁰⁰ According to CFR art. 52(3), the meaning and scope of rights in both instruments shall be the same.

and current events and to play the role of the public watchdog.¹⁰¹ The right applies offline and on the Internet.¹⁰²

One of the main risks for freedom of expression associated with algorithmic content moderation is over-blocking, meaning the unjustified removal or blocking of content or the suspension or termination of user accounts. In 2012, the Court of Justice of the EU held that a filtering system for copyright violations could undermine freedom of information since it might not distinguish adequately between lawful and unlawful content, which could lead to the blocking of lawful communications.¹⁰³ This concern is equally valid outside the copyright context. The technical limitations of AI systems, together with regulatory pressure from States who increasingly request intermediaries to take down certain categories of content, often based on vague definitions, incentivize platforms to follow a “if in doubt, take it down” approach.¹⁰⁴ There is, indeed, growing empirical evidence of platforms’ over-removal of content.¹⁰⁵ To illustrate, social media platforms have deleted hundreds of posts condemning the eviction of Palestinians from the Sheikh Jarrah neighborhood of Jerusalem¹⁰⁶ or restricted access to information about abortion.¹⁰⁷ Both examples are a consequence of the algorithmic content moderation systems either not being able to recognize context or not knowing underlying facts and legal nuances. Such automated removals, even if unintentional and subsequently revoked, potentially limit both the right to impart information (of users who post content online) and the right to receive information (of third parties who do not get to see the deleted content).

On the other hand, the under-blocking of certain online content may also have a negative impact on the right to freedom of expression. Not acting against illegal

¹⁰¹ Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland, App no 931/13 (ECtHR June 27, 2017); Von Hannover v. Germany (no 2.), Apps nos 40660/08 and 60641/08) (ECtHR February 7, 2012).

¹⁰² See, for example, Council of Europe, “Recommendation of the Committee of Ministers to member States on a Guide to human rights for Internet users” (adopted by the Committee of Ministers on April 16, 2014 at the 1197th meeting of the Ministers’ Deputies).

¹⁰³ Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV* [2012], para 50.

¹⁰⁴ Keller, “Empirical Evidence of Over-Removal by Internet Companies under Intermediary Liability Laws: An Updated List” *CIS Blog* (February 8, 2021), <https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>, accessed April 4, 2023.

¹⁰⁵ Keller and Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*.

¹⁰⁶ “Sheikh Jarrah: Facebook and Twitter silencing protests, deleting evidence” *Article 19* (May 10, 2021), www.article19.org/resources/sheikh-jarrah-facebook-and-twitter-silencing-protests-deleting-evidence/, accessed April 4, 2023; “Israel/Palestine: Facebook Censors Discussion of Rights Issues” *Human Rights Watch* (October 8, 2021), www.hrw.org/news/2021/10/08/israel/palestine-facebook-censors-discussion-rights-issues, accessed April 4, 2023.

¹⁰⁷ Kuczerawy and Dutkiewicz, “Accessing Information about Abortion: The Role of Online Platforms under the EU Digital Services Act” *VerfBlog* (July 28, 2022), <https://verfassungsblog.de/accessing-information-about-abortion/>, accessed March 28, 2023.

content and some forms of legal but harmful content (i.e., hate speech) may lead people (especially marginalized communities) to express themselves less freely or withdraw from participating in the online discourse.

In addition, in the context of fact-checking, AI cannot yet analyze entire, complex disinformation narratives and detect all uses of synthetic media manipulation.¹⁰⁸ Thus, an overreliance on AI systems to verify the trustworthiness of the news may prove detrimental to the right to freedom of expression.

To mitigate these risks, the DSA provides certain procedural safeguards. It does not force intermediary services to moderate content, but requires that any restrictions imposed on users' content based on terms and conditions are applied and enforced "in a diligent, objective and proportionate manner," with "due regard to the rights and legitimate interests of all parties involved" (art. 14(4)). Not only do they have to take *due regard* to fundamental rights in cases of content removal, but also when restricting the availability, visibility, and accessibility of information. What *due regard* means in this context will be defined in courts. Moreover, the DSA requires intermediary services to balance their freedom to conduct a business with other rights such as users' freedom of expression. Online platforms also have to provide a statement of reasons as to why the content has been removed or the account has been blocked and implement an internal complaint-handling system that enables users to lodge complaints (art. 21). Another procedural option is the out-of-court dispute settlement or a judicial remedy.¹⁰⁹

A novelty foreseen by the DSA is an obligation for VLOPs and VLOSEs to mitigate systemic risks such as actual or foreseeable negative effects for the exercise of fundamental rights, in particular freedom of expression and information, including the freedom and pluralism of the media, enshrined in Article 11 of the CFR, and foreseeable negative effects on civic discourse (art. 34).

News personalization from the freedom of expression perspective looks paradoxical at first glance. As Eskens points out, "news personalisation may enhance the right to receive information, but it may also hinder or downplay the right to receive information and the autonomy with which news users exercise their right to receive information."¹¹⁰ Given that content prioritization practices have a potential for promoting trustworthy and reliable news, it can be argued that platforms should be required to ensure online access to content of general public interest. The Council of Europe, for instance, suggested that States should act to make public interest content more prominent, including by introducing new obligations for platforms and intermediaries, and also impose minimum standards such as transparency.¹¹¹ Legal

¹⁰⁸ DW Innovation, "AI for Content Verification I: Status Quo and Current Limitations," p. 7.

¹⁰⁹ See also Kuczerawy, "Remedying Overremoval: The Three-Tiered Approach of the DSA," *VerfBlog* (November 3, 2022), <https://verfassungsblog.de/remedying-overremoval/>, accessed April 5, 2023.

¹¹⁰ Eskens, "The fundamental rights of news users: The legal groundwork for a personalised online news environment" (PhD Thesis University of Amsterdam, 2021).

¹¹¹ Council of Europe, "Guidance Note on the Prioritisation of Public Interest Content Online," p. 7.

scholars have proposed *exposure diversity* as a design principle for recommender systems¹¹² or the development of “diversity-enhancing public service algorithms.”¹¹³ But who should decide what content is trustworthy or authoritative, and based on what criteria? Are algorithmic systems of private platforms equipped enough to quantify normative values such as *newsworthiness*? What safeguards would prevent States from forcing platforms to prioritize State-approved-only information or government propaganda? Besides, many of the problems with content diversity are at least to some extent user-driven – users themselves, under their right to freedom of expression, determine what kind of content they upload and share.¹¹⁴ Legally imposed public interest content recommendations could limit users’ autonomy in their news selection by paternalistically censoring the range of information that is available to them. While there are no such obligations in the DSA, some legislative proposals at the national level are currently reviewing such options.¹¹⁵

14.3.5 *Threats to Media Freedom and Pluralism Online*

Freedom and pluralism of the media are pillars of liberal democracies. They are also covered by Art. 10 ECHR and Art. 11 CFR. The ECtHR found that *new electronic media*, such as an online news outlet, are also entitled to the protection of the right to media freedom.¹¹⁶ Moreover, the so-called positive obligations doctrine imposes an obligation on States to protect editorial independence from private parties, such as social media.¹¹⁷

Social media platforms have on multiple occasions erased content coming from media organizations, including public broadcasters, and journalists. This is often illustrated by the controversy that arose around Facebook’s decision to delete a post by a Norwegian journalist, which featured the well-known Vietnam War photo of a nude young girl fleeing a napalm attack.¹¹⁸ Similarly, users sharing an article from The Guardian showing Aboriginal men in chains were banned from Facebook on

¹¹² Helberger, Karppinen, and D’Acunto, “Exposure diversity as a design principle for recommender systems” (2018) *Information, Communication & Society*, 21(2): 191–207.

¹¹³ Vermeulen, “Access Diversity through Online News Media and Public Service Algorithms. An Analysis of News Recommendation in Light of Article 10 ECHR” in James Meese and Sara Bannerman (eds), *The Algorithmic Distribution of News: Policy Responses* (Cham: Palgrave Macmillan, 2022), pp. 269–287.

¹¹⁴ Helberger, Pierson, and Poell, “Governing online platforms: From contested to cooperative responsibility” (2017) *The Information Society*.

¹¹⁵ See, for example, the UK Draft Online Safety Bill presented to Parliament by the Minister of State for Digital and Culture by Command of Her Majesty May 2021.

¹¹⁶ *OOO Regnum v. Russia*, App no 22649/08 (ECtHR, September 8, 2020).

¹¹⁷ van Drunen and Fechner, “Safeguarding Editorial Independence in an Automated Media System: The Relationship Between Law and Journalistic Perspectives.”

¹¹⁸ Scott and Isaac, “Facebook Restores Iconic Vietnam War Photo It Censored for Nudity” *The New York Times* (September 9, 2016), www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html, accessed April 4, 2023.

the grounds of posting nudity.¹¹⁹ Other examples include videos of activists and local news outlets that documented the war crimes of the regime of Bashar al-Assad in Syria¹²⁰ or a Swedish journalist's material reporting sexual violence against minors.¹²¹ This is due to technical limitations of the algorithmic content moderation tools and their inability to distinguish educational, awareness raising or journalistic material from other content.

In order to prevent removals of content coming from media organizations, a so-called *media exemption*¹²² was proposed during the discussions of the DSA proposal, aiming to ensure that the media would be informed and have the possibility to challenge any content moderation measure before its implementation. The amendments were not included in the final text of the DSA. There is no special protection or any obligation of prior notice to media organizations in the DSA. Media organizations and journalists can invoke the same procedural rights that apply to all users of online platforms. One can also imagine that mass-scale algorithmic takedowns of media content, suspension or termination of journalists' accounts by VLOPs could amount to a *systemic risk* in a form of a negative effect on the exercise of the freedom and pluralism of the media.¹²³ However, what qualifies as *systemic*, and when a threshold of systemic risk to freedom and pluralism of the media is reached, remains undefined.

Recognizing media service providers' role in the distribution of information and in the exercise of the right to receive and impart information online, the European Media Freedom Act (EMFA) grants media service providers special procedural rights vis-à-vis VLOPs. Where a VLOP considers that content provided by recognized media service providers¹²⁴ is incompatible with its terms and conditions, it

¹¹⁹ Tylor, "Facebook blocks and bans users for sharing Guardian article showing Aboriginal men in chains" *The Guardian* (June 15, 2020), www.theguardian.com/technology/2020/jun/15/facebook-blocks-bans-users-sharing-guardian-article-showing-aboriginal-men-in-chains, accessed April 4, 2023. Note that a spokeswoman for Facebook apologized for the mistake and that the post was restored.

¹²⁰ Alimardani and Elswah, "Digital Orientalism: #SaveSheikhJarrah and Arabic Content Moderation"; see also Hadi Al Khatib and Dia Kayyali "YouTube Is Erasing History" *The New York Times* (October 23, 2019), www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.html, accessed April 5, 2023.

¹²¹ Oversight Board decision 2021-016-FB-FBR.

¹²² Amendments 511 and 513 to Recital 38 and Article 12 of the Digital Services Act proposal (January 15, 2022). Note that the term "media exemption" is contested; other terms like "non-interference principle" are used interchangeably. See, for example, EBU. "The Digital Services Act must safeguard freedom of expression online" (January 18, 2022), www.ebu.ch/files/live/sites/ebu/files/News/Position_Papers/open/2022/220118-DSA-media-statement-final.pdf, accessed April 4, 2023.

¹²³ Buijs "The Digital Services Act and the implications for news media and journalistic content (Part 1)" *DSA Observatory* (September 29, 2022), <https://dsa-observatory.eu/2022/09/29/digital-services-act-implications-for-news-media-journalistic-content-part-1/>, accessed April 5, 2023.

¹²⁴ EMFA grants this and other procedural rights to media who declare that they are media service providers and meet the conditions of art. 18(1) EMFA. Interestingly, one of the conditions is a declaration that a media service provider does not provide AI-generated content without subjecting it to

should “duly consider media freedom and media pluralism” in accordance with the DSA and provide, as early as possible, the necessary explanations to media service providers in a statement of reasons as referred to in the DSA and the P2B Regulation (recital 50, art. 18). In what has been coined as a *non-interference principle*¹²⁵, VLOPs should provide the media service provider concerned, prior to the suspension or restriction of visibility taking effect, with an opportunity to reply to the statement of reasons within 24 hours of receiving it.¹²⁶ Where, following or in the absence of a reply, a VLOP takes the decision to suspend or restrict visibility, it shall inform the media service provider concerned without undue delay. Moreover, media service providers’ complaints under the DSA and the P2B Regulation shall be processed and decided upon with priority and without undue delay. Importantly, EMFA’s Article 18 does not apply where VLOPs suspend or restrict the visibility of content in compliance with their obligations to protect minors, to take measures against illegal content or in order to assess and mitigate systemic risks.¹²⁷

Next to media freedom, media pluralism and diversity of media content are equally essential for the functioning of a democratic society and are the corollaries of the fundamental right to freedom of expression and information.¹²⁸ Media pluralism is recognized as one of the core values of the European Union.¹²⁹

In recent years, concerns over the decline of media diversity and pluralism have increased.¹³⁰ Online platforms “have acquired increasing control over the flow, availability, findability and accessibility of information and other content online.”¹³¹ Considering platforms’ advertising-driven business model based on a profit maximization, they have more incentives to increase the visibility of content that would keep users more engaged. It can be argued that not only does this fail to promote

human review or editorial control (art. 18(1)(e)). At the same time, VLOPs have the right to reject such self-declarations where they consider that those conditions are not met.

¹²⁵ Papaevangelou, “The Non-Interference Principle”: Debating Online Platforms’ Treatment of Editorial Content in the EU’s Digital Services Act’. *European Journal of Communication*, 2023.

¹²⁶ A shorter timeframe could apply in the event of a crisis as defined in Article 36(2) of the DSA in order to take into account, in particular, an urgent need to moderate the relevant content in such exceptional circumstances.

¹²⁷ See Articles 28, 34 and 35 DSA.

¹²⁸ Committee of Ministers, “Recommendation CM/Rec(2007)2 on media pluralism and diversity of media content” January 31, 2007. See also the similar Committee of Ministers, “Recommendation No. R (99) 1 on measures to promote media pluralism” adopted on January 19, 1999.

¹²⁹ CFR, art. 11; Treaty on European Union Articles 2 and 6.

¹³⁰ Mathias A. Färdigh, “Monitoring media pluralism in the digital era: Application of the Media Pluralism Monitor in the European Union, Albania, Montenegro, the Republic of North Macedonia, Serbia and Turkey in the year 2021. Country report: Sweden” Centre for Media Pluralism and Media Freedom (CMPF); Media Pluralism Monitor (MPM); 2022. Parcu, “New digital threats to media pluralism in the information age” (2020) *Competition and Regulation in Network Industries*, 21(2): 91–109; CMPF-CiTiP-IViR-SMIT, “Study on Media Plurality and Diversity Online.”

¹³¹ Committee of Ministers, “Recommendation CM/Rec(2018)(1)[1] of the Committee of Ministers to member States on media pluralism and transparency of media ownership” March 7, 2018.

diversity, but it strongly reduces it.¹³² The reduction of plurality and diversity of news content resulting from platforms' content curation policies may limit users' access to information. It also negatively affects society as a whole, since the availability and accessibility of diverse information is a prerequisite for citizens to form and express their opinions and participate in the democratic discourse in an informed way.¹³³

14.3.6 Threats to Media Independence

The growing dependence on automation in news production and distribution has a profound impact on editorial independence as well as on the organizational and business choices of media organizations. One way in which automation could potentially challenge editorial independence is media reliance on non-media actors such as engineers, data providers, and technology companies that develop or fund the development of the datasets or algorithms used to automate editorial decision-making.¹³⁴

(News) media organizations depend more and more on platforms to distribute their content. The phenomena of *platformed publishing* refers to the situation where news organizations have no or little control over the distribution mechanisms decided by the platforms.¹³⁵ Moreover, media organizations optimize news content to make it *algorithm ready*, for example, by producing popular content which is attractive for the platforms' recommender systems.¹³⁶ The entire news cycle, from production, distribution, to consumption of news "is (re)organized around platforms, their rules and logic and thus influenced and mediated by them."¹³⁷ Individuals and newsrooms, therefore, depend structurally on platforms, which affects the functioning and power allocation within the media ecosystem.¹³⁸

Moreover, platforms provide essential technical infrastructure (e.g., cloud computing and storage), access to AI models, or stand-alone software.¹³⁹ This increases the potential for so-called *infrastructure capture*¹⁴⁰ and risks shifting even more

¹³² Stasi, "Ensuring Pluralism in Social Media Markets: Some Suggestions" (2020) Working Paper, EUI RSCAS, 2020/05, Centre for Media Pluralism and Media Freedom.

¹³³ Council of Europe, Commissioner for Human Rights, "Media Pluralism and Human Rights, Issue Discussion paper," (2011), <https://rm.coe.int/16806da515>, accessed April 5, 2023; *Lingens v. Austria* App no 9815/82 (EctHR July 8, 1986); *Castells v. Spain*, App No 11798/85 (EctHR April 23, 1992).

¹³⁴ Van Drunen and Fechner, "Safeguarding Editorial Independence in an Automated Media System: The Relationship between Law and Journalistic Perspectives."

¹³⁵ Nielsen and Ganter, "The power of platforms" *Reuters Institute* (April 29, 2022), <https://reutersinstitute.politics.ox.ac.uk/news/power-platforms>, accessed April 5, 2023.

¹³⁶ Seipp, Helberger, de Vreese, and Ausloos, "Dealing with Opinion Power in the Platform World: Why We Really Have to Rethink Media Concentration Law" (2023) *Digital Journalism*.

¹³⁷ *Ibid.*

¹³⁸ *Ibid.*

¹³⁹ Simon, "Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy" (2022) *Digital Journalism*, 10(10): 1832–1854.

¹⁴⁰ Nechushtai, "Could Digital Platforms Capture the Media through Infrastructure?" (2018) *Journalism* 19(8): 1043–1058.

control to platform companies, at the expense of the media organizations autonomy and independence.

The relationship between AI, media, and platforms, raises broader questions about the underlying political, economic, and technological power structures and platforms' opinion power.¹⁴¹ To answer these challenges, legal scholars have called for rethinking media concentration rules¹⁴² and media law in general.¹⁴³ However, the considerations about opinion power of platforms, values, and media independence are somehow missing from the current EU regulatory initiatives. The EMFA rightly points out that providers of video-sharing platforms and VLOPs "play a key role in the organisation of content, including by automated means or by means of algorithms," and some "have started to exercise editorial control over a section or sections of their services" (recital 11). While it does mention "the formation of public opinion" as relevant parameter in the assessment of media market concentrations (art. 21), it does not provide a solution to address the concerns about the dependency between platforms' AI capacities and media organizations.¹⁴⁴

14.4 CONCLUSIONS

AI will continue to transform media in ways we can only imagine. Will news articles be written by fully automated systems? Will the proliferation of synthetic media content dramatically change the way we perceive information? Or will virtual reality experiences and new forms of interactive storytelling replace traditional (public interest) media content? As AI technology continues to advance, it is essential that the EU legal framework keeps pace with these developments to ensure that the use of AI in media is responsible, ethical, and beneficial to society as a whole. After all, information is a public good and media companies cannot be treated as any other businesses.¹⁴⁵

The DSA takes an important step in providing procedural safeguards to mitigate risks for the right to freedom of expression and freedom of the media posed by online platforms' content moderation practices. It recognizes that the way VLOPs

¹⁴¹ Helberger, "The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power" (2020) *Digital Journalism*, 8(6): 842–854; see also Orla Lynskey, "Regulating 'Platform Power'" (2017) LSE Legal Studies Working Paper No. 1/2017.

¹⁴² See, for example, Helberger, "The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power"; Seipp, Helberger, de Vreese and Ausloos, "Dealing with Opinion Power in the Platform World: Why We Really Have to Rethink Media Concentration Law."

¹⁴³ Tambini, "A theory of media freedom" (2021) *Journal of Media Law*, 13(2): 135–152.

¹⁴⁴ Other than art. 18 EMFA, mentioned earlier, which is limited in scope.

¹⁴⁵ See the speech of Ursula Von der Leyen (President of the European Commission) for the release of the MFA: European Commission, "European Media Freedom Act" (2022), https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/european-media-freedom-act_en, accessed April 5, 2023. See also Explanatory Memorandum of the European Commission, Proposal for a Regulation of the European Parliament and of the Council establishing a common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU 2022 [COM(2022) 457 final].

and VLOSEs moderate content may cause systemic risks to the freedom and pluralism of the media and negatively affect civic discourse. The EMFA also aims to strengthen the position of media organizations vis-à-vis online platforms. However, it remains to be seen how effective a 24-hour *non-interference* rule will be given the high threshold of who counts as a *media service provider* and which content falls within the scope of Art. 18 EMFA.

Many of the AI applications in (social) media, such as recommender systems, news bots, or the use of AI to generate or manipulate content are likely to be covered by the AI Act. A strong focus on external transparency both in the AI Act and in the DSA can be seen as a positive step to ensure that users become more aware of the extensive use of AI in (social) media.

However, many aspects of the use of AI in and by media such as the intelligibility gap, societal risks raised by AI (including worker displacement and environmental costs), as well as reliance of tech companies for access to high-quality media content to develop AI systems,¹⁴⁶ remain only limitedly addressed. Media organizations' dependency on social media platforms recommender systems and algorithmic content moderation as well as power imbalances in access to AI infrastructure should also be tackled by the European legal framework.

It is equally important to facilitate and stimulate responsible research and development of AI in the media sector, particularly in local and small media organizations, to avoid the AI divide. In this regard, it is worth mentioning that the Council of Europe's Committee of Experts on Increasing Resilience of the Media adopted Guidelines on the responsible implementation of AI systems in journalism.¹⁴⁷ The Guidelines offer practical guidance to news media organizations, States, tech providers and platforms that disseminate news, on how AI systems should be used to support the production of journalism. The Guidelines also include a checklist for media organizations to guide the procurement process of AI systems by offering questions that could help in scrutinizing the fairness of a procurement contract with an external provider (Annex 1).

Now the time has come to see how these regulations are enforced and whether they will enable a digital level playing field. To this end, policymakers, industry stakeholders, and legal professionals must work together to address the legal and ethical implications of AI in media and promote a fair and transparent use of AI.

¹⁴⁶ For an overview of how the media sector is responding to content crawling for model training, see Bocyte, Dutkiewicz, "How is the Media Sector Responding to Content Crawling for Model Training" *Medium* (June 18, 2024), medium.com/ai-media-observatory/how-is-the-media-sector-responding-to-content-crawling-for-model-training-9812ac2916d8, accessed August 1, 2024.

¹⁴⁷ Committee of Experts on Increasing Resilience of the Media (MSI-RES), *Council of Europe* (November 30, 2023), rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6, accessed August 1, 2024.