

AI and Media¹

Lidia Dutkiewicz, Noémie Krack, Aleksandra Kuczerawy & Peggy Valcke

Chapter 15

In: Nathalie A. Smuha (ed.), *Cambridge handbook on the law, ethics and policy of Artificial Intelligence*. Cambridge: Cambridge University Press. (2024), forthcoming

Abstract

This chapter discusses how AI technologies permeate the media sector. It sketches opportunities and benefits of the use of AI in media content gathering and production, in media content distribution, in fact-checking and content moderation. The chapter then zooms in on ethical and legal risks raised by AI-driven media applications: lack of data availability, poor data quality and bias in training datasets, lack of transparency, risks for the right to freedom of expression, threats to media freedom and pluralism online, and threats to media independence. Finally, the chapter introduces the relevant elements of the EU legal framework which aim to mitigate these risks, such as the Digital Services Act, the European Media Freedom Act proposal and the AI Act proposal.

Keywords: media, AI, content moderation, media freedom, media pluralism, recommender systems, Digital Services Act, AI Act, freedom of expression, social media

1. Introduction

Media companies can benefit from artificial intelligence (AI)² technologies to increase productivity and explore new possibilities for producing, distributing and re-using content. This chapter demonstrates the potential of the use of AI in media.³ It takes a selective approach to showcase a variety of applications in the following areas: Can ChatGPT write news articles?

¹ 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).

² For the definition of AI see Chapter 2 of this book.

³ This chapter takes a narrower understanding of media, focusing on traditional mass media outlets such as news media, public service media (PSM) 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, e.g., Georg 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 2020) <<https://data.europa.eu/doi/10.2861/294829>>.

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 re-use (Section 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 non-discrimination 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 Act⁴, the AI Act proposal⁵, and the European Media Freedom Act proposal⁶) tries to mitigate some of the risks to fundamental rights posed by the development and the use of AI in media (Section 3). Section 4 offers conclusions.

2. Opportunities of AI applications in media⁷

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 a day-to-day journalistic practice.⁸ 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.⁹

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, in 2014, Reuters News Tracer project used natural language processing techniques to decide which topics are “newsworthy”.¹⁰ It detected the bombing of hospitals in Aleppo and the terror

⁴ 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) 2022 (OJ L 277/1).

⁵ European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. 2021 [COM(2021) 206 final].

⁶ European Commission, Proposal for a Regulation of the European Parliament and of the Council establishing common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU 2022 [COM(2022) 457 final].

⁷ For a broad overview see, e.g., Filareti Tsalakanidou, ‘AI technologies and applications in media: State of Play, Foresight, and Research Directions’ (2022) *AI4Media* <https://www.ai4media.eu/wp-content/uploads/2022/03/AI4Media_D2.3_Roadmap_final.pdf>.

⁸ Charlie Beckett, ‘New Powers, New Responsibilities. A Global Survey of Journalism and Artificial Intelligence’ (*Blogs LSE*, 18 November 2019) <<https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>> accessed 18 March 2023.

⁹ *ibid.*

¹⁰ Jonathan Stray, ‘The Age of the Cyborg’, (*Columbia Journalism Review*, 30 November 2016) <https://www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php> accessed 18 March 2023.

attacks in Nice and Brussels before they were reported by other media.¹¹ 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.¹²

AI has also been proven useful in investigative journalism to assist journalists in tasks which could not be done by humans alone or would have taken 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.¹³

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 as the Associated Press, Forbes and The New York Times, have started to automate news content.¹⁴ Terms like “robot journalism”, “automated journalism” and “algorithmic journalism” have been used interchangeably to describe this phenomenon.¹⁵ In addition, generative AI tools such as ChatGPT,¹⁶ Midjourney¹⁷ or DALL-E¹⁸ is being used to illustrate news stories, simplify text for different audiences, summarize documents or writing potential headlines.¹⁹

2.2 AI in media content distribution²⁰

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*

¹¹ *ibid.*

¹² See, e.g., ReTV, <https://retv-project.eu/news-discourse-monitoring/>.

¹³ Marina Walker Guevara, ‘How Artificial Intelligence Can Help Us Crack More Panama Papers Stories’ (*International Consortium of Investigative Journalists*, 25 March 2019) <<https://www.icij.org/inside-icij/2019/03/how-artificial-intelligence-can-help-us-crack-more-panama-papers-stories/>> accessed 18 March 2023.

¹⁴ Andreas Graefe, ‘Guide to Automated Journalism’ (2016) *Columbia Journalism Review* <https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php>

¹⁵ Although they do not have the same meaning. See, e.g., Graefe, *Guide to Automated Journalism*, p.3; Matteo 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; Konstantin Nicholas Dörr, ‘Mapping the Field of Algorithmic Journalism’ (2015) *Digital Journalism*.

¹⁶ See OpenAI, ‘Introducing ChatGpt’ <<https://openai.com/blog/chatgpt>> accessed 5 April 2023.

¹⁷ Midjourney’ <<https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>> accessed 5 April 2023.

¹⁸ OpenAI, ‘DALL-E2’, <<https://openai.com/product/dall-e-2>> accessed 5 April 2023.

¹⁹ See also: Generative AI In the Newsroom, <<https://generative-ai-newsroom.com/>> accessed 5 April 2023.

²⁰ 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, e.g., Matt Carlson, ‘Order versus access: news search engines and the challenge to traditional journalistic roles’ (2007) *Media, Culture & Society*, 29(6), 1014–1030.

available online and use algorithms to guide users towards a narrower selection of material, according to a set of criteria chosen by their developers."²¹

In recent years, online news media (e.g. online newspapers' websites and apps) started engaging in 'news recommendation' practices.²² 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.²³

The 2022 Ada Lovelace report²⁴ provides an informative in-depth snapshot of the BBC's development and use of recommendation systems which gives insights about the role of recommendations in public service media (PSM).²⁵ As pointed out by the authors, developing recommendation systems for PSM requires an interrogation of the organizations' role in democratic societies in the digital age, i.e. how to translate the public service values²⁶ 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.²⁷ This is a considerable difference between the public and private sector. 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.²⁸

Research shows that social media platforms are using a mix of commercial criteria, but also vague public interest considerations in their content prioritization measures.²⁹ Importantly,

²¹ Ada Lovelace Institute 'Inform, educate, entertain... and recommend? Exploring the use and ethics of recommendation systems in public service media' (2022)

<<https://www.adalovelaceinstitute.org/report/informededucate-entertain-recommend/>>.

²² Judith Vermeulen 'The Algorithmic State of Mind: A Human Rights Frame for Governing News Recommendation' (2022) (Ghent University, Faculty of Law and Criminology).

²³ *ibid.*

²⁴ Ada Lovelace Institute, '*Inform, educate, entertain... and recommend?(...)*', p.4.

²⁵ 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 5 April 2023.

²⁶ Public service media organisations 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 organisations commit to strive to uphold these in practice. See EBU, 'Empowering Society, a declaration on the core values of public service media', <https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf>

²⁷ Ada Lovelace Institute, '*Inform, educate, entertain... and recommend?(...)*', p.4.

²⁸ See for e.g. Adam Mosseri, 'Shedding More Light on How Instagram Works' *AboutInstagram.com* (8 June 2021), <<https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>> accessed 22 March 2023.

²⁹ 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 on 5 April.

prioritization of some content demotes other.³⁰ 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.³¹ In fact, online platforms use a whole arsenal of techniques to (de)amplify the visibility or reach of some content.³² To illustrate, in the aftermath of Russian aggression on Ukraine, platforms announced they would ‘restrict access’ to RT and Sputnik media outlets.³³ 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.³⁴

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.³⁵ Ranking functions and optimization metrics in recommendation systems have become powerful determinants of access to media and news content.³⁶ This has consequences for both the fundamental right to freedom of expression and media freedom (see Section 3).

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).³⁷

³⁰ Keller uses the term ‘demote’ to cover any form of deamplification, including decreasing content’s algorithmic ranking or excluding it from features like recommendations. Daphne Keller, ‘Amplification and Its Discontents’ Knight First Amendment Institute at Columbia University (8 June 2021)

<<https://knightcolumbia.org/content/amplification-and-its-discontents>> accessed March 19, 2023.

³¹ Facebook, ‘What are recommendations on Facebook?’ *Facebook Help Center* <<https://www.facebook.com/help/1257205004624246>> accessed 5 April 2023.

³² See, e.g., Eric Goldman ‘Content Moderation Remedies’ (2021) 28 *Mich. Tech. L. Rev.* 1.

³³ Laura Kayali, ‘Facebook’s Parent Company Restricts EU Access to Russia’s RT, Sputnik’ *Politico* (28 February 2022) <<https://www.politico.eu/article/facebook-parent-company-restricts-eu-access-to-russia-rt-sputnik/>> accessed 5 April 2023.

³⁴ Elizabeth Culliford ‘Twitter Will Label, Reduce Visibility of Tweets Linking to Russian State Media’ *Reuters* (28 February 2022) <<https://www.reuters.com/technology/twitter-will-label-reduce-visibility-tweets-linking-russian-state-media-2022-02-28/>> accessed 17 January 2023.

³⁵ 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, 1-3 December 2021’, <<https://rm.coe.int/cdmsi-2021-009-guidance-note-on-the-prioritisation-of-pi-content-e-ado/1680a524c4>> accessed on 5 April 2023.

³⁶ *ibid.*

³⁷ Lucas 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 4 April 2023.

To illustrate, platforms such as DALIL help fact-checkers to spot questionable claims which then require subsequent verification.³⁸ 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 non-factual statements (claims) made on social media or analyze the relationships between accounts.³⁹ 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.⁴⁰ Some advanced processing and analytics methods can also be used to analyze different types of content and apply a trustworthiness scoring to online articles.⁴¹

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.⁵⁹ Some platforms have also been labelling content containing "synthetic and manipulated media".⁴²

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.⁴³ Collaborative tools that enable users to work together to find, organize and verify user-generated content are also on the rise.⁴⁴

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).*”⁴⁵ This section focuses on the instances where AI is used either by media

³⁸ EU Neighbours south, ‘AI-driven platform launched to accelerate Arabic language fact-checking’ (2 January 2023) <<https://south.euneighbours.eu/news/ai-driven-platform-launched-to-accelerate-arabic-language-fact-checking/>> accessed 4 April 2023.

³⁹ DW Innovation, ‘AI for Content Verification I: Status Quo and Current Limitations’ (DW Innovation 24 October 2022) <<https://innovation.dw.com/articles/ai-for-content-verification-i-status-quo-and-current-limitations>> accessed 4 April 2023.

⁴⁰ See WeVerify, <<https://weverify.eu/verification-plugin/>>, accessed 5 April 2023.

⁴¹ Francesco Saverio Nucci et al., ‘Artificial Intelligence Against Disinformation: The FANDANGO Practical Case (short paper)’ (International Forum on Digital and Democracy (IFDaD), Venice, Italy, 2020).

⁴² Twitter, ‘Synthetic and manipulated media policy’, *Twitter Help Centre* <<https://help.twitter.com/en/rules-and-policies/manipulated-media>> accessed 5 April 2023.

⁴³ See for e.g. vera.ai, <<https://www.veraai.eu/home>> accessed 5 April 2023.

⁴⁴ See Truly Media, <<https://www.truly.media>> accessed 5 April 2023. See also: AI4Media, ‘UC1: AI for Social Media and Against Disinformation’, *AI4Media* <<https://www.ai4media.eu/uc1-ai-for-social-media-and-against-disinformation/>> accessed 4 April 2023.

⁴⁵ Robert Gorwa, Reuben Binns and Christian Katzenbach, ‘Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance’ (2020) 7 *Big Data & Society*.

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.

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),⁴⁶ and encompass stereotypes, homophobic, racist, sexist, and xenophobic terms which may amount to hate speech.⁴⁷ The rise of incivility in online news comments negatively affects people’s perceptions of news article quality and increases hostility.⁴⁸ “Don’t read the comments” has become a mantra throughout the media.⁴⁹ 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.

Some introduced AI systems to support their moderation processes. To illustrate, both the New York Times⁵⁰ and the Washington Post⁵¹ 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 Der STANDARD) has developed an automated system to pre-filter problematic content, as well as a set of pre-emptive moderation techniques, including forum design changes to prevent problematic content from being posted in the first place.⁵²

⁴⁶ Kevin Coe, Kate Kenski, Stephen A. Rains ‘Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments’ (2014) 64 *Journal of Communication* 658.

⁴⁷ Gina Masullo Che, *Online Incivility and Public Debate: Nasty Talk* (Springer International Publishing AG 2017).

⁴⁸ 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).

⁴⁹ Becky 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.

⁵⁰ Alex Traub, ‘Why Humans, Not Machines, Make the Tough Calls on Comments’ *The New York Times* (26 October 2021) <<https://www.nytimes.com/2021/10/26/insider/why-humans-not-machines-make-the-tough-calls-on-comments.html>> accessed 5 April 2023.

⁵¹ WashPostPR, ‘The Washington Post leverages artificial intelligence in comment moderation’ *The Washington Post* (22 June 2017) <<https://www.washingtonpost.com/pr/wp/2017/06/22/the-washington-post-leverages-artificial-intelligence-in-comment-moderation/>> accessed 5 April 2023.

⁵² Ben Wagner, Johanne Kübler, Eliška Pirková, Rita Gsenger, Carolina Ferro ‘Reimagining content moderation and safeguarding fundamental rights. A study on community-led platforms’ *The Greens/EFA in the European Parliament* (3 May 2021) <https://www.greens-efa.eu/files/assets/docs/alternative_content_web.pdf> (accessed 4 April 2023).

Others, like Reuters or CNN, have removed their comment sections completely.⁵³ 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.⁵⁴ This, however, did not remove the problem of hateful speech. To the contrary, it amplified it.⁵⁵

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. To avoid liability for illegal content online 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').⁵⁶ 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 3.4).

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.⁵⁷ Many platforms have therefore turned to AI-based content moderation. Such automation can be used as a 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.⁵⁸ 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,

⁵³ Jiawei Liu and Douglas M. McLeod, 'Pathways to news commenting and the removal of the comment system on news websites' (2021) *Journalism*, 22(4), 867–881.

⁵⁴ *ibid.*

⁵⁵ United Nations, 'Hate Speech: Turning the tide' *UN News, Global perspective Human stories* (30 January 2023) <<https://news.un.org/en/story/2023/01/1132617>> accessed 5 April 2023; Luke Munn, 'Angry by design: toxic communication and technical architectures.' (2020) *Humanities and Social Sciences Communications*, 7, 53.

⁵⁶ Luca Belli and Jamila Venturini, 'Private Ordering and the Rise of Terms of Service as Cyber-Regulation' (2016) *Internet Policy Review* 5 (4).

⁵⁷ Emma 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), <<https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>> accessed 5 April 2023.

⁵⁸ Cambridge Consultants, 'Use of AI in online content moderation' (Report produced on behalf of Ofcom 2019) <<https://www.cambridgeconsultants.com/sites/default/files/uploaded-pdfs/Use%20of%20AI%20in%20online%20content%20moderation.pdf>> accessed 5 April 2023.

labelling content as fact-checked or adding a “graphic content” label to problematic content before or as users encounter it.⁵⁹

Algorithmic moderation systems help classify user-generated content based on either matching or prediction techniques.⁶⁰ These techniques present a number of technical limitations.⁶¹ 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 a growing empirical evidence of platforms’ over-removal of content coming from individuals and media organizations (see Section 3.4).⁶²

3. Legal and ethical challenges of AI applications in media

This section identifies legal and ethical challenges of AI in media across various stages of the media value chain described above. The section also shows how these challenges may be mitigated by the EU legal framework.⁶³

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.⁶⁴ This concentrated control over the data influences how editorial decision-making is automated (see Section 3.6).

Data availability is also of paramount importance for news verification and fact-checking activities. Access to social media data for scientific researchers is vital to analyze and mitigate the harms resulting from disinformation, political microtargeting, or the effect of social media on elections or children’s wellbeing.⁶⁵ Access to platforms’ data for researchers is currently

⁵⁹ Eric Goldman, ‘Content moderation remedies’ (2021) 28 *Michigan Technology Law Review* 1, 1-59.

⁶⁰ Robert Gorwa, Reuben Binns and Christian Katzenbach, ‘Algorithmic content moderation: Technical and political challenges in the automation of platform governance’, p.6.

⁶¹ Llansó et al, *Artificial Intelligence, Content Moderation, and Freedom of Expression*.

⁶² Daphne Keller and Paddy Leerssen, ‘Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation’ In Nathaniel Persily and Joshua A. Tucker (eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (SSRC Anxieties of Democracy, pp. 220-251). Cambridge: Cambridge University Press.

⁶³ 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.

⁶⁴ Max van Drunen and Denise Fechner, ‘Safeguarding Editorial Independence in an Automated Media System: The Relationship Between Law and Journalistic Perspectives’ (2022) *Digital Journalism*.

⁶⁵ Pasquetto Irene V. et al., ‘Tackling misinformation: What researchers could do with social media data’ (2020) *Harvard Kennedy School Misinformation Review*, v. 1, n. 8, p. 01-14; Jef Ausloos, Paddy Leerssen and Pim ten Thije, ‘Operationalizing Research Access in Platform Governance What to Learn from Other Industries?’

mainly governed by contractual agreements, platforms' own terms of service and public application programming interfaces (APIs). APIs 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 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*

Algorithm Watch (25 June 2020)

<https://www.ivir.nl/publicaties/download/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf> accessed 23 March 2023.

⁶⁶ See e.g. 'We Research Misinformation on Facebook. It Just Disabled Our Accounts' *The New York Times* (10 August 2021) <<https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html?referringSource=articleShare>> accessed 5 April 2023; Nicolas Kayser-Bril, 'AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook' *Algorithm Watch* (13 August 2021) <<https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>> accessed 5 April 2023.

⁶⁷ Irene 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' (13 April 2021).

⁶⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119. See also Chapter 9 AI and Privacy Law.

⁶⁹ Lidia Dutkiewicz 'From the DSA to Media Data Space: the possible solutions for the access to platforms' data to tackle disinformation' *European Law blog* (19 October 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 13 March 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' (31 May 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 4 April 2023..

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

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

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 around 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 Act⁷³ 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 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.

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

⁷³ Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 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].

⁷⁵ Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 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 (3 December 2020, COM/2020/784 final).

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

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 which 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 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 which may lead to over- and under-removal of content. Both can interfere

⁷⁹ See e.g., 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' (31 March 2023) <<https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english>> accessed 5 April 2023.

⁸¹ Natali Helberger and Nicholas 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).

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).

with a range of fundamental rights such as the right to freedom of expression⁸⁶ (see Section 3.4), the right to data protection, as well as the right to non-discrimination.

The consequence of using unreliable data could be the spread of misinformation⁸⁷ as illustrated by inaccurate responses for news queries from search engines using generative AI. Research into Bing's generative AI accuracy for news queries shows that there are detail errors, 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, i.e. the uncritical trust in information provided by the automated system despite the information being actually incorrect.

3.3 Transparency

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

The *first* aspect relates to transparency of automated journalism and the use of generative AI. It is becoming harder to distinguish between human and machine produced content. As pointed out by Monti, "*ethically speaking, identifying whether a human being or a machine is the writer of a piece of news seems to be necessary so the reader will be aware when reading an article and allow him or her the freedom to choose between traditional journalism and automated journalism.*"⁹⁰ The AI Act proposal contains transparency obligations for providers and users of certain AI systems. AI systems should be designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless

⁸⁶ See, e.g., Natali Helberger, Max van Drunen, Sarah Eskens, Mariella Bastian and Judith 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*, Vol 11 No. 3; Noémie Krack, Marie Beudels, Peggy Valcke and Aleksandra Kuczerawy 'AI in the Belgian Media Landscape. When Fundamental Risks Meet Regulatory Complexities', *Artificial Intelligence and the Law*, vol 13 (Second Revised Edition, 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* (29 June 2022) <<https://digital-strategy.ec.europa.eu/en/policies/online-disinformation>> accessed 5 April 2023.

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

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

⁹⁰ Monti, 'Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field', p. 3.

this is obvious from the circumstances and the context of use.⁹¹ What ‘interaction’ means in this context is unclear, but it could cover applications such as chatbots, newsbots, recommender systems, and automated writing systems.⁹² Moreover, users of AI systems should disclose if the content has been artificially generated or manipulated unless it is used for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences, and subject to appropriate safeguards for the rights and freedoms of third parties.⁹³ It has been argued, not without controversies, that artists using AI-generated (audiovisual) output, e.g. DALL-E-created, would not have to disclose that they are using AI.⁹⁴

The *second* aspect concerns transparency about whether and how social media platforms use AI systems to moderate and recommend content, including content coming from media organizations. The DSA presents multiple layers of transparency obligations that differ depending on the type of service concerned.⁹⁵ 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”, information on any policies, procedures, measures and tools used for the purpose of content moderation, including about “algorithmic decision-making” and human review (art. 14). In other words, online intermediaries are free to decide what kind of content they do not wish to host, even if this content is not actually illegal. They have to, however, make it clear to their users why, when and how they moderate content, including with the use of AI.

The DSA also regulates recommender system transparency. As mentioned above, 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

⁹¹ AI Act proposal, art 52(1). Note that the General approach of the Council of the European Union proposed to amend this article by adding "the unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use."

⁹² Natali Helberger and Nicholas Diakopoulos, ‘The European AI Act and How It Matters for Research into AI in Media and Journalism’ (2022) *Digital Journalism*.

⁹³ Note that the General approach of the Council of the European Union proposed to amend this article so the disclose obligation does not apply "where the content is part of an evidently creative, satirical, artistic or fictional work or programme subject to appropriate safeguards for the rights and freedoms of third parties."

⁹⁴ Katerina Yordanova and Thomas Gils, ‘Council of the EU - Compromise Text AI Act’ *Knowledge Center Data & Society* <<https://data-en-maatschappij.ai/en/policy-monitor/council-of-the-eu-compromise-text-ai-act>> accessed 28 March 2023.

⁹⁵ 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*.

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 which are most significant in determining the information suggested, and the reasons for the relative importance of those parameters. There are additional requirements 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 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 for 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 on the use of automatic means (art. 15). Extra transparency reporting obligations apply to VLOPs (art. 33).

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),⁹⁶ 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 right 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, current events and to play the role of the public watchdog.⁹⁷ The right applies offline and on the Internet.⁹⁸

⁹⁶ According to CFR art. 52(3), the meaning and scope of rights in both instruments shall be the same.

⁹⁷ *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, App no 931/13 (ECtHR 27 June 2017); *Von Hannover v. Germany* (no 2.), Apps nos. 40660/08 and 60641/08 (ECtHR 7 February 2012).

⁹⁸ See, e.g., 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 16 April 2014 at the 1197th meeting of the Ministers' Deputies).

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.⁹⁹ This concern is equally valid outside the copyright context. The technical limitations of AI systems, together with a regulatory pressure from States who increasingly request intermediaries to take down certain categories of content, often based on vague definitions, incentivizes platforms to follow a “if in doubt, take it down” approach.¹⁰⁰ There is, indeed, growing empirical evidence of platforms’ over-removal of content.¹⁰¹ To illustrate, social media platforms have deleted hundreds of posts condemning the eviction of Palestinians from the Sheikh Jarrah neighborhood of Jerusalem¹⁰² or restricted access to information about abortion.¹⁰³ 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 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.¹⁰⁴ Thus, an

⁹⁹ Case C-360/10 *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV*, [2012] para 50.

¹⁰⁰ Daphne Keller, ‘Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List’ *CIS Blog* (8 February 2021) <<https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>> accessed 4 April 2023.

¹⁰¹ Keller and Leerssen, *Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation*.

¹⁰² ‘Sheikh Jarrah: Facebook and Twitter silencing protests, deleting evidence’ *Article19* (10 May 2021) <<https://www.article19.org/resources/sheikh-jarrah-facebook-and-twitter-silencing-protests-deleting-evidence/>> accessed 4 April 2023; ‘Israel/Palestine: Facebook Censors Discussion of Rights Issues’ *Human Rights Watch* (8 October 2021) <<https://www.hrw.org/news/2021/10/08/israel/palestine-facebook-censors-discussion-rights-issues>> accessed 4 April 2023.

¹⁰³ Aleksandra Kuczerawy and Lidia Dutkiewicz, ‘Accessing Information about Abortion: The Role of Online Platforms Under the EU Digital Services Act’ *VerfBlog* (28 July 2022) <<https://verfassungsblog.de/accessing-information-about-abortion/>> accessed 28 March 2023.

¹⁰⁴ DW Innovation, ‘AI for Content Verification I: Status Quo and Current Limitations’, p. 7.

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 to 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.¹⁰⁵

A novelty foreseen by the DSA is an obligation for VLOPs and VLOSEs to mitigate systematic 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*".¹⁰⁶ 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.¹⁰⁷ Legal scholars

¹⁰⁵ See also: Aleksandra Kuczerawy, 'Remedying Overremoval: The Three-Tiered Approach of the DSA', *VerfBlog* (3 November 2022) <<https://verfassungsblog.de/remedying-overremoval/>> accessed 5 April 2023.

¹⁰⁶ Sarah Eskens, 'The fundamental rights of news users: The legal groundwork for a personalised online news environment' (PhD-Thesis University of Amsterdam 2021).

¹⁰⁷ Council of Europe, 'Guidance Note on the Prioritisation of Public Interest Content Online', p.7.

have proposed “exposure diversity” as a design principle for recommender systems¹⁰⁸ or the development of “diversity-enhancing public service algorithms”.¹⁰⁹ 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.¹¹⁰ 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 national level are currently reviewing such options.¹¹¹

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 that “new electronic media”, such as an online news outlet, are also entitled to protection of the right to media freedom.¹¹² Moreover, the so-called positive obligations doctrine imposes an obligation on States to protect editorial independence from private parties, such as social media.¹¹³

Social media platforms have on multiple occasions erased content coming from media organizations, including public broadcasters, and journalists. This is well-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.¹¹⁴ Similarly, users sharing an article from The Guardian showing Aboriginal men in

¹⁰⁸ Natali Helberger, Kari Karppinen and Lucia D’Acunto, ‘Exposure diversity as a design principle for recommender systems’ (2018) *Information, Communication & Society*, 21:2, 191-207.

¹⁰⁹ Judith 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*, 269–287. Cham: Palgrave Macmillan.

¹¹⁰ Natali Helberger, Jo Pierson and Thomas Poell, ‘Governing online platforms: From contested to cooperative responsibility’ (2017) *The Information Society*.

¹¹¹ See e.g. 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.

¹¹² *OOO Regnum v. Russia*, App no 22649/08 (ECtHR, 8 September 2020).

¹¹³ van Drunen and Fechner, ‘*Safeguarding Editorial Independence in an Automated Media System: The Relationship Between Law and Journalistic Perspectives*’.

¹¹⁴ Mark Scott and Mike Isaac, ‘Facebook Restores Iconic Vietnam War Photo It Censored for Nudity’ *The New York Times* (9 September 2016) <<https://www.nytimes.com/2016/09/10/technology/facebook-vietnam-war-photo-nudity.html>> accessed 4 April 2023.

chains were banned from Facebook on the grounds of posting nudity.¹¹⁵ Other examples include videos of activists and local news outlets that documented the war crimes of the regime of Bashar al-Assad in Syria¹¹⁶ or a Swedish journalist's material reporting sexual violence against minors.¹¹⁷ 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"¹¹⁸ 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. They are, however, not entirely out of the picture. The European Media Freedom Act (EMFA) proposal¹¹⁹ grants certain media service providers procedural privileges which cannot be enjoyed by other, non-media, users. These privileges include a prior notice with statement of reasons before the suspension takes effect, priority in complaint-handling, and, in case of frequent restrictions or suspensions, an option to engage in a "meaningful and effective dialogue" to terminate and avoid them in the future (art. 17).

There is no special protection or any obligation of a prior notice to media organizations in the DSA. Media organizations and journalists can however invoke the same procedural rights which apply to all users of online platforms (see above). 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 for the exercise of the

¹¹⁵ Josh Tylor, 'Facebook blocks and bans users for sharing Guardian article showing Aboriginal men in chains' *The Guardian* (15 June 2020) <<https://www.theguardian.com/technology/2020/jun/15/facebook-blocks-bans-users-sharing-guardian-article-showing-aboriginal-men-in-chains>> accessed 4 April 2023. Note that a spokeswoman for Facebook apologized for the mistake and that the post was restored.

¹¹⁶ Mahsa Alimardani and Mona Elswah, 'Digital Orientalism: #SaveSheikhJarrah and Arabic Content Moderation'; see also: Hadi Al Khatib and Dia Kayyali 'YouTube Is Erasing History' *The New York Times* (23 October 2019) <<https://www.nytimes.com/2019/10/23/opinion/syria-youtube-content-moderation.html>> accessed 5 April 2023.

¹¹⁷ Oversight Board decision 2021-016-FB-FBR.

¹¹⁸ Amendments 511 and 513 to Recital 38 and Article 12 of Digital Services Act proposal (15 January 2022). Note that the term 'media exemption' is contested; other terms like 'non-interference principle' are used interchangeably. See e.g. EBU, 'The Digital Services Act must safeguard freedom of expression online' (18 Jan 2022) <https://www.ebu.ch/files/live/sites/ebu/files/News/Position_Papers/open/2022/220118-DSA-media-statement-final.pdf> accessed 4 April 2023.

¹¹⁹ Note that the European Media Freedom Act proposal is at the beginning of the legislative process.

freedom and pluralism of the media.¹²⁰ This will have to be mitigated according to the DSA risks mitigation measures.

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.¹²¹ Media pluralism is recognized as one of the core values of the European Union.¹²²

In recent years, concerns over the decline of media diversity and pluralism have increased.¹²³ Online platforms “*have acquired increasing control over the flow, availability, findability and accessibility of information and other content online*”.¹²⁴ 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 diversity, but it strongly reduces it.¹²⁵ 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.¹²⁶

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 organizational and business choices of media

¹²⁰ Doris Buijs ‘The Digital Services Act and the implications for news media and journalistic content (Part 1)’ *DSA Observatory* (29 September 2022) < <https://dsa-observatory.eu/2022/09/29/digital-services-act-implications-for-news-media-journalistic-content-part-1/>> accessed 5 April 2023.

¹²¹ Committee of Ministers, ‘Recommendation CM/Rec(2007)2 on media pluralism and diversity of media content’ 31 January 2007. See also the similar Committee of Ministers, ‘Recommendation No. R (99) 1 on measures to promote media pluralism’ adopted on 19 January 1999.

¹²² CFR, art 11; Treaty on European Union Articles 2 and 6.

¹²³ 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. Pier Luigi 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*’.

¹²⁴ 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’ 7 March 2018.

¹²⁵ Maria Luisa Stasi, ‘Ensuring Pluralism in Social Media Markets: Some Suggestions’ (2020) Working Paper, EUI RSCAS, 2020/05, Centre for Media Pluralism and Media Freedom.

¹²⁶ Council of Europe, Commissioner for Human Rights, ‘Media Pluralism and Human rights, Issue Discussion paper’, (2011) <<https://rm.coe.int/16806da515>> accessed on 5 April 2023; *Lingens v. Austria* App no 9815/82 (ECtHR 8 July 1986); *Castells v. Spain*, App No 11798/85 (ECtHR 23 April 1992).

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 which develop or fund the development of the datasets or algorithms used to automate editorial decision-making.¹²⁷

(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.¹²⁸ Moreover, media organizations optimize news content to make it “algorithm ready”, e.g. by producing popular content which is attractive for the platforms’ recommender systems.¹²⁹ 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*”.¹³⁰ Individuals and newsrooms, therefore, depend structurally on platforms, which affects the functioning and power allocation within the media ecosystem.¹³¹

Moreover, platforms provide essential technical infrastructure (e.g., cloud computing and storage), access to AI models or stand-alone software.¹³² This increases the potential for so-called ‘infrastructure capture’¹³³ and risks shifting even more 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.¹³⁴ To answer these challenges, legal scholars have called for rethinking media concentration rules,¹³⁵ and media law in general.¹³⁶ However, the considerations about opinion

¹²⁷ van Drunen and Fechner, ‘*Safeguarding Editorial Independence in an Automated Media System: The Relationship Between Law and Journalistic Perspectives*’.

¹²⁸ Rasmus Kleis Nielsen and Sarah Anne Ganter, ‘The power of platforms’ *Reuters Institute* (29 April 2022) <<https://reutersinstitute.politics.ox.ac.uk/news/power-platforms>> accessed 5 April 2023.

¹²⁹ Theresa Josephine Seipp, Natali Helberger, Claes de Vreese and Jef Ausloos, ‘Dealing with Opinion Power in the Platform World: Why We Really Have to Rethink Media Concentration Law’ (2023) *Digital Journalism*.

¹³⁰ *ibid.*

¹³¹ *ibid.*

¹³² Felix M. Simon, ‘Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy’ (2022) *Digital Journalism*, 10:10, 1832-1854.

¹³³ Efrat Nechushtai, ‘Could Digital Platforms Capture the Media through Infrastructure?’ (2018) *Journalism* 19 (8): 1043–1058.

¹³⁴ Natali 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.

¹³⁵ See e.g. 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.*’

¹³⁶ Damian Tambini, ‘A theory of media freedom’ (2021) *Journal of Media Law*, 13:2, 135-152.

power of platforms, values, and media independence are somehow missing from the current EU regulatory initiatives. The EMFA proposal rightly points out that providers of video-sharing platforms and VLOPs “*play a key role in the content organization, including by automated means or algorithms*”, and some “*have started to exercise editorial control over a section or sections of their services*”.¹³⁷ 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.¹³⁸

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.¹³⁹

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 and VLOSEs moderate content may cause systemic risks to the freedom and pluralism of the media and negatively affects civic discourse. The EMFA proposal is likely to further strengthen the position of media organizations vis-à-vis online platforms.

Many of the AI applications in (social) media, such as recommender systems, news bots or automated news articles systems are likely to be covered by the AI Act. It remains to be seen whether and how the use of generative AI in media will be regulated. At the same time, it still

¹³⁷ European Media Freedom Act proposal, rec 8.

¹³⁸ Other than art. 17 EMFA proposal, mentioned above, which is limited in scope.

¹³⁹ 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 5 April 2023. See also Explanatory Memorandum of the European Commission, Proposal for a Regulation of the European Parliament and of the Council establishing common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU 2022 [COM(2022) 457 final].

needs to be determined whether the use of AI in media constitutes a ‘high risk’ for a democratic society and should be subject to stringent legal obligations.

The growing media organizations’ dependency on social media platforms distribution systems and AI infrastructure as well as power imbalances should also be tackled by the European legal framework. In this regard, an important initiative is currently being developed by the Council of Europe’s Committee of Experts on Increasing Resilience of the Media, namely the Draft Guidelines on the use of digital tools including artificial intelligence (AI) for journalism/by journalists.¹⁴⁰ The Guidelines aim to offer concrete and implementable recommendations and best practices around the different stages in the implementation of AI. They will also include a specific section addressing AI developers, platforms as the main distributors of news content and Member States.¹⁴¹

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.

¹⁴⁰ 'Committee of Experts on Increasing Resilience of the Media (MSI-RES) held its first meeting' *Council of Europe* (11 April 2022) <<https://www.coe.int/en/web/freedom-expression/-/committee-of-experts-on-increasing-resilience-of-the-media-msi-res-held-its-first-meeting>> accessed 5 April 2023.

¹⁴¹ Committee of Experts on Increasing Resilience of the Media (MSI-RES), 'Meeting report' (2nd meeting, 29-30 September 2022 Strasbourg, Palais de l’Europe, Room 8, and online) <<https://rm.coe.int/msi-res-2022-09-2nd-meeting-report-en-18-10-2022/1680a9719c>> accessed 5 April 2023.