LL.M. Thesis

Law and Technology in Europe

AI-Generated Content and the Pollution of the Information Sphere: A Freedom of Expression Analysis under Article 10 ECHR

Katharina Goth

27th June 2025

24.404 Words

Abstract

This thesis analyses whether and to what extent AI-generated content is protected by freedom of expression and information under Article 10 §1 ECHR, and under what conditions state interferences can be justified to protect the integrity of the information sphere. Synthetic content is increasingly present and has the potential to impact the digital information sphere. The analysis examines the extent to which it causes structural impairment. It therefore begins with a systematic categorisation of the phenomenon of *'pollution of the information sphere'* into qualitative, quantitative, and formal dimensions. A legal analysis then determines whether, and under what conditions, AI-generated content falls within the scope of protection of Article 10 §1 ECHR. The assessment relies on established methods of doctrinal legal interpretation. It shows that content which is either attributable to a human or that contributes functionally to public discourse is eligible for protection. State interferences can be justified if they are prescribed by law, serve a legitimate aim and are necessary for a democratic society. In this context, the analysis demonstrates that the protection of the information sphere can be subsumed under several legitimate aims under Article 10 §2 ECHR. While the qualitative pollution of the information sphere can clearly be restricted under certain conditions, addressing quantitative information pollution requires a more comprehensive, systemic regulatory solution. The results indicate that AI-generated content remains assessable within the existing ECHR framework. The taxonomy of information pollution developed in this thesis can serve as a methodological frame of reference for future legal and empirical studies.

Table of Contents

Introduction

AI-generated content is no longer a novelty; it is increasingly shaping what we read, believe and share. Synthetic texts in news headlines, product reviews, political articles or medical advice are permeating our everyday digital lives, often without us even realising it. Recent studies estimate that 30-40 % of online text is already fully or partially generated by AI and the trend is rising.[1] At the same time, the mass production of such content places new structural burdens on the digital information sphere, for example in the form of disinformation and information overload.[2] This *'pollution of the information sphere'* threatens the right to form informed opinions, freedom of expression and participation in democratic processes.[3] The risks associated with AI-generated content have already prompted first regulatory responses, such as the EU AI Act and the Digital Services Act. However, this could also affect the communication freedoms of real people. In this context, two issues arise: Is AI-generated content protected as *'expression'* under Article 10 §1 ECHR? Furthermore, can the protection of the information sphere constitute a legitimate aim for proportionate state interferences or do these novel phenomena present legal challenges to the assessment of the existing legal framework of Article 10 ECHR? These questions are the subject matter of this thesis.

The literature on the information sphere focuses on different aspects of the phenomenon. From a philosophical perspective, *Floridi* describes the term as an information system that includes all forms of content and the media through which they are conveyed[4], while military strategic approaches such as those of *Allen and Gilbert Jr.* define the information sphere as an independent operational domain.[5] In media studies and legal debates, the terms *infollution*[6],

[1] Dirk HR Spennemann, '"Delving into" the Quantification of Ai-Generated Content on the Internet (Syn- Thetic Data)' (Charles Sturt University 2025) 3 <https://arxiv.org/pdf/2504.08755> accessed 24 April 2025.
[2] MS Sridhar, 'Knowledge Dilution, Pollution, and Burden in Digital Media' [2024] Journal of Information and Knowledge 109, 110; Jiawei Zhou and others, 'Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions', *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (ACM 2023) 14 <https://dl.acm.org/doi/10.1145/3544548.3581318> accessed 26 May 2025.
[3] Kalina (ed) Bontcheva, 'Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities' (University of Sheffield 2024) 21 <https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf> accessed 26 May 2025.
[4] Luciano Floridi, *Philosophy and Computing: An Introduction* (Routledge 1999) 8.
[5] Patrick D Allen and Dennis P Gilbert Jr., 'Qualifying the Information Sphere as a Domain' (2010) 9 Journal of Information Warfare 39, 45.
[6] Qaisar Iqbal and others, 'Infollution (Information Pollution) Management, Filtering Strategy, Scalable Workforce, and Organizational Learning: A Conceptual Study' (2018) 10 Information Management and Business Review 1.

infodemic[7] and *infobesity*[8] have emerged to describe disruptions to the information sphere. *Nielsen and Cho* emphasise the growth of irrelevant, erroneous, or decontextualised content as a structural risk.[9] The role of AI-generated content is increasingly recognised as an amplifier of such disruptions but has so far hardly been systematically assessed under the ECHR, especially Article 10.

De Vries is one of the few who deal with the fundamental dogmatics of the ECHR in this context.[10] According to her, although Article 10 §1 ECHR refers to '*everyone'*, the ever-present human involvement in AI-generated content already allows it to be categorised under the scope of protection.[11] She also emphasises that freedom of expression protects not only the perspective of the speaker, but also that of the recipient.[12] This suggests that AI-generated content falls under Article 10 §1 ECHR, provided it entails human involvement or serves the interests of a human recipient. In a US context, *Massaro, Norton and Kaminski* argue similarly: The decisive factor is the value of the information for the audience, not the identity of the sender.[13] *Cheong* further advocates a contextual approach by arguing that only algorithms with recognisable *'expressive intent'* fall under the protection of the First Amendment, and that the scope of protection must depend on the respective context.[14] In Canadian law, Section 2(b) of the Charter of Rights and Freedoms only protects *'any activity that conveys*, *or attempts to convey meaning'*. *Goswami* emphasises that this protection should depend on the purpose behind the content, particularly whether the algorithm was used to convey meaning.[15]

The conditions under which AI-generated content may be legally restricted in the light of Article 10 §2 ECHR are also largely unclear. Much of the existing literature focuses on AI-supported moderation and how this can interfere with human forms of expression. For example,

---

7 David J Rothkopf, 'When the Buzz Bites Back' (*University of Delaware – Global Agenda 2004*, 2003) <https://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html> accessed 24 April 2025.

8 Prasanna Karhade and others, 'Digital Platforms and Infobesity: A Research Agenda' in Aravinda Garimella and others (eds), *The Role of e-Business during the Time of Grand Challenges*, vol 418 (Springer International Publishing 2021) 67 <https://link.springer.com/10.1007/978-3-030-79454-5_7> accessed 18 April 2025.

9 Jakob Nielsen, 'IM, Not IP (Information Pollution): A Steady Dose of Realtime Interruptions Is Toxic to Anyone's Health.' (2003) 1 Queue 76, 76–77; Paek J Cho, '"Infollution" and the Quality of Life' (2002) <https://stanford.edu/~ncho/Infollution_manuscript_PJCho_2002.pdf> accessed 28 May 2025.

10 Katja De Vries, 'Let the Robot Speak! AI-Generated Speech and Freedom of Expression' in Steffen Hindelang and Andreas Moberg (eds), *YSEC Yearbook of Socio-Economic Constitutions 2021*, vol 2021 (Springer International Publishing 2022).

11 ibid 99–100.

12 ibid 100.

13 Toni M Massaro, Helen Norton and Margot E Kaminski, 'SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment' (2017) 101 Minnesota Law Review 2481, 2502.

14 Inyoung Cheong, 'Freedom of Algorithmic Expression' (2023) 91 Social Science Research Network 680, 710–712.

15 Manasvin (Veenu) Goswami, 'Algorithms and Freedom of Expression' in Woodrow Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (1st edn, Cambridge University Press 2020) 558-566,568-573.

Hasimi and Poniszewska-Maranda show that automated filtering mechanisms can block legitimate content and thereby pose a risk to diversity of opinion.[16] *Moreno* analyses the regulation of deepfakes under the AI-Act and discusses the compatibility of such measures with the requirements of Article 10 ECHR.[17] While these studies primarily relate to the use of AI for content control, their proportionality considerations can also contribute to this analysis.

Despite numerous individual contributions to our understanding of phenomena such as disinformation and information overload, there has so far been a lack of systematic legal coverage of the *'pollution of the information sphere'* as a structured problem area. Existing works describe partial aspects without categorising them. Furthermore, it has yet to be clarified whether and to what extent AI-generated content is covered by Article 10 §1 ECHR. Although *de Vries* provides an initial impulse, she does not distinguish precisely between the personal and material scope of protection, nor does she analyse freedom of expression and freedom of information in detail. The present thesis fills this gap by systematically analysing and categorising both dimensions of protection. Finally, there is an absence of a structured examination framework for justifying state intervention in AI-generated content under Article 10 §2 ECHR. This work therefore develops a legal guideline that adapts the existing ECtHR's standards to generative systems and examines the legal challenges that may arise when the current legal framework is applied to new phenomena like AI and the increasing pollution of the information sphere. It thereby prepares the ground for solving future cases. This leads to the following research question:

> **RQ:** *To what extent does AI-generated content fall within the scope of freedom of expression and information under Article 10 §1 ECHR, and what legal challenges arise when justifying state interference to counter the pollution of the information sphere?*

Three sub-questions have been formulated to answer this main question, which also determine the structure of the work.

> **SQ1:** *How does the proliferation of AI-generated content contribute to the pollution of the information sphere?*

> **SQ2:** *In light of the pollution of the information sphere, does AI-*

[16] Lumbardha Hasimi and Aneta Poniszewska-Marańda, 'Detection of Disinformation and Content Filtering Using Machine Learning: Implications to Human Rights and Freedom of Speech' 6–9.
[17] Felipe Romero Moreno, 'Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content' (2024) 38 International Review of Law, Computers & Technology 297, 311–318.

generated content fall within the scope of protection under Article 10 §1 ECHR, particularly in relation to freedom of expression and freedom of information?

SQ3: *What legal challenges does Article 10 §2 ECHR pose when justifying state interference with AI-generated content in response to the pollution of the information sphere?*

The work takes a doctrinal legal approach. Article 10 § 1 ECHR is applied to AI-generated content based on recognised interpretation criteria, such as textual interpretation, contextual interpretation, historical context and the object and purpose. These methods follow the general principles of interpretation under Articles 31, 32 of the Vienna Convention on the Law of Treaties and are also regularly reflected in ECtHR practice.[18] They allow for a structured yet flexible understanding of Article 10 §2 ECHR in light of new communicative technology. The case law of the ECtHR provides the main frame of reference. This work therefore aims to contribute to the further development of principles for dealing with new forms of expression and to categorise AI-generated content within the Convention´s existing protection system.

The analysis finds that AI-generated content pollutes the information sphere in qualitative, quantitative, and formal terms. This impedes access to reliable information and the exercise of fundamental communicative freedoms. Accordingly, such content only falls within the scope of protection if it can be attributed to a human actor or if it functionally contributes to public discourse. Furthermore, state interference with AI-generated content is justified under Article 10 §2 ECHR if it pursues a legitimate aim and meets proportionality requirements. In particular, the protection of the integrity of the information sphere can be recognised as a legitimate aim under Article 10 §2 ECHR. Although new phenomena such as AI-generated content and the resulting pollution of the information sphere raise new questions of interpretation, they do not present any insurmountable legal challenges. The existing framework of Article 10 §2 ECHR, including ECtHR case law, can be coherently applied to these developments.

In order to answer the research question, the thesis is divided into three main chapters, each of which addresses one of the sub-questions. The first chapter develops a systematisation of the

[18] *Golder v the United Kingdom* [1975] ECtHR 57496/16 paras.29-36; *Demir and Baykara v Turkey* [2008] ECtHR [GC] 34503/97 paras.65-68.

‘*pollution of the information sphere’* into qualitative, quantitative, and formal dimensions, and analyses the extent to which AI-generated content contributes to the information pollution at these levels. The second chapter then builds on this by examining whether, and under what conditions, AI-generated content falls within the personal and material scope of protection of freedom of expression and freedom of information under Article 10 §1 ECHR. The third chapter considers whether protecting the information sphere can be recognised as a legitimate aim under Article 10 §2 ECHR, and which standards apply to state interference with AI-generated content. Two case studies then illustrate that the existing legal framework can be applied to AI-generated content and the pollution of the information sphere without major challenges. The aim is to provide a systematically viable and dogmatically substantiated answer to each of these questions.

Chapter 1: How AI Pollutes the Information Sphere

The first chapter addresses the question: *How does the proliferation of AI-generated content contribute to the pollution of the information sphere?* The aim is to analyse this phenomenon systematically, by examining its forms, mechanisms and effects. This chapter first sets out a working definition of the information sphere, followed by a brief classification of AI-generated content. In the main section, the concept of information pollution is systematised into the categories: qualitative, quantitative and formal pollution of the information sphere. These categories are used to analyse the extent to which AI-generated content contributes to information pollution. Finally, the direct consequences of a polluted information sphere are outlined briefly to illustrate the practical relevance of the phenomenon.

I. Defining the Information Sphere

For the legal analysis, it is essential to determine a precise definition of the information sphere. This section first presents philosophical and military strategic definitions and then develops a normative working definition as a basis for further analysis.

1. Philosophical Foundation and Strategic-Military Conceptualization

The term '*infosphere*' was first used by *Kenneth E. Boulding* in 1970 to describe a distinct and autonomous segment of the sociosphere in which '*each individual or group of individuals* [is] *a node in a global network of inputs and outputs of information, symbols and language*'.[19] Ten years later, *Alvin Toffler* used the term to describe the future changes brought about by modern communication structures in the transition to the information society, defining it as a sphere '*for producing and distributing information*'.[20] Philosopher *Luciano Floridi* coined the term '*infosphere*' in 1999 as a central element of the information philosophy.[21] He defines it as '*the whole system of services and documents, encoded in any semiotic and physical media, whose contents include any sort of data, information and knowledge* […]*, with no limitations either in size, typology or logical structure*.'[22]

Since the 2000s, the term '*information sphere*' has been used in military and strategic contexts. The US Department of Defense, for example, defined the information environment

[19] Kenneth E Boulding, *Economics as a Science* (Tata McGraw-Hill Publishing Company Ltd 1970) 15–16.
[20] Alvin Toffler, *The Third Wave* (William Morrow and Company 1980) 48–49.
[21] Floridi (n 4) 8.
[22] ibid.

as *'The aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information'*.[23] *Patrick D. Allen* and *Dennis P. Gilbert Jr.* suggested in 2009 that the information domain should be defined as a distinct operational domain alongside existing physical domains.[24] To them, it is not the technical infrastructure that matters, but rather '*The space of relationships among actors, information, and information systems that form a sphere of interest and influence in or through which information related activities, functions, and operations are undertaken to accomplish missions and exercise control over an opponent in order to achieve desired effects*'.[25]

Both perspectives offer valuable insights for a normative understanding of the structure, quality and need for protection of information environments.

2. A Normative Concept of the Information Sphere

For the purposes of legal analysis, this thesis develops a normative concept of the information sphere which integrates elements of both approaches. It is termed *'normative'* to illustrate that it serves as an analytical framework to evaluate conditions that are relevant to the exercise of human rights. The information sphere is a dynamic socio-technical structure in which human and algorithmic actors generate, exchange and process information. It structures public communication and democratic processes of negotiation. The information sphere can be characterised by four elements that shape its internal functioning and external impact: (1) *actors* - such as people, institutions or algorithmic systems; (2) *information* - meaningfully structured data, including facts, opinions or synthetic content; (3) *systems* - technical infrastructures for generating and disseminating information, such as social networks or AI models; and (4) *relational dynamics* - interactions between actors, platform mechanisms and algorithmic structures that influence the visibility, evaluation and reception of information. In the following, the term *information sphere* is used in this sense.

Joint Chiefs of Staff, United States Department of Defense, 'Department of Defense Dictionary of Military and Associated Terms' 1 110 <https://irp.fas.org/doddir/dod/jp1_02.pdf> accessed 24 April 2025.
[24] Allen and Gilbert Jr. (n 5) 45.
[25] ibid 43.

II.AI-Generated Content

In order to clearly define the object of analysis, this section provides a distinction of the varying forms of AI-generated content.

AI-generated content is not a rigid concept but operates on a spectrum of varying degrees of human and machine involvement. Even for contributions that appear to have been generated entirely by an AI-system, the question is whether true AI autonomy is even possible. Every output is at least to some degree based on human development, training data and control, so AI always remains embedded in a pre-determined human framework.[26] In this context, the distinction between AI-generated and AI-assisted content is central for legal categorisation.[27] AI-*generated* content includes works that are created by an AI-system without significant human influence, while AI-*assisted* content is characterised by a significant human contribution in which creative decisions are made and key design aspects are controlled.[28] Simply formulating a prompt or launching an AI system is not sufficient, if the human has no significant influence on the final result.[29]

This thesis defines the threshold for *'significant human influence'* as the point at which a human not only triggers or approves content, but determines the substance, framing or argumentative trajectory of the output. The following example illustrates this: Person A is planning a blog post and has already finalised the core content. A uses an LLM to produce a text with a different style. In this case, the main content comes from A. Person B, however, reads A's blog post but has neither their own arguments nor the time to analyse the content. Instead, they transfer the text to an LLM and ask it to create a counter-post. As B has not provided any content specifications, the new argumentation is mainly generated by the model. Therefore, the main content comes from the system. For the purposes of this paper, this constitutes AI-generated content. In practice, however, the distinction between human-generated and AI-generated content is often less clear. Later in the thesis, I will return to this issue, as the degree of human influence is central to determining whether content falls under Article 10 §1 ECHR and whether an interference can be justified.

[26] Vincenzo Calderonio, 'A Basis for Human Responsibility in Artificial Intelligence Computation' (arXiv, 21 January 2025) 10–11 <http://arxiv.org/abs/2501.12498> accessed 28 April 2025.

[27] Amanda Johnson, 'What Is the Difference between AI-Assisted and AI-Generated Content?' (3 June 2024) <https://www.clearscope.io/blog/ai-generated-vs-ai-assisted-content> accessed 9 May 2025.

[28] Johannes Fritz, 'Understanding Authorship in Artificial Intelligence-Assisted Works' [2025] Journal of Intellectual Property Law and Practice 1, 6.

[29] ibid 10.

AI-generated content appears in a variety of formats, depending on the system, application and purpose. Common forms include text-based contributions and AI-created images or videos. This content is generated by various systems, such as language models, image generators or text-to-speech programmes, and serves functions ranging from entertainment and information to influence. This thesis focuses on AI-generated content that aims to have a communicative impact, meaning that it is designed to engage with human recipients. Outputs with purely technical or functional purposes, such as code or data structures without communicative intent, are excluded.

III.Pollution of the Information Sphere

The growing prevalence of AI-generated content raises questions about its impact on the information sphere. In order to assess potential interferences with freedom of expression under Article 10 §2 ECHR, it is necessary to develop a systematic understanding of the phenomenon. This section first outlines the historical development and existing understandings of the pollution of the information sphere. It then introduces a taxonomy that distinguishes between three dimensions of information pollution: qualitative, quantitative and formal. After his, the specific contribution of AI-generated content to each category is analysed. Finally, the section highlights the direct consequences of information pollution for democratic discourse and individual decision-making as a groundwork for a normative legal assessment.

1. Origins and Dimensions of the Pollution of the Information Sphere

The phenomenon of *pollution of the information sphere* (or *information pollution*) is by no means a consequence of recent developments in the field of artificial intelligence. Rather, it has been described in the literature long before the emergence of generative systems and has been associated with different stages of information development. Back in 2003, *Nielsen* analysed the shift from email to instant messaging as a response to the overload of traditional communication channels - email, he said, had become an *'overloaded channel for readers'*.[30] *Cameron and Ju-Pak* also showed back in 2000 that information pollution is not limited to digital media: They examined print magazines in which advertorials were often not sufficiently labelled and were therefore barely recognisable as commercial messages for readers.[31] Both examples illustrate that information pollution can occur structurally in different forms of media, regardless of their digital or technological design. In the academic debate on the phenomenon

[30] Nielsen (n 9) 76.

[31] Glen T Cameron, Kuen-Hee Ju-Pak and Bong-Hyun Kim, 'Information Pollution?: Labeling and Format of Advertorials' (2000) 21 Newspaper Research Journal 65, 71–72.

of information pollution, there are numerous terms such as *infollution*[32], *infobesity*[33], *infodemic*[34] or *information pollution*.[35] Some of these terms are used synonymously, while others cover only certain aspects of the problem.

In order to systematically capture the concept, on the basis of the relevant literature, I make a distinction between three categories: quantitative, qualitative and formal information pollution. Some phenomena can be assigned to several categories, particularly at the intersection of formal and quantitative pollution. However, the categorisation provides a useful foundation for the subsequent legal evaluation of state interferences.

2. Qualitative Pollution of the Information Sphere

a. Definition and Content

The qualitative information pollution describes the devaluation, distortion or targeted manipulation of information. It primarily concerns the second structural point of the information sphere: *information* - as meaningfully structured data including facts, opinions or synthetically generated content.

The conceptualisation and systematisation of qualitative information pollution is multi-layered in the literature and can be traced back historically. In 1984, *Orman* was one of the first to define information pollution as '*the contamination of information supply with incomplete, inconsistent, or irrelevant information*'.[36] In 1996 *Cai and Zhang* expanded the term to include a behavioural perspective: they understand information pollution as the accumulation of widespread, unwanted messages that can have potentially profound negative effects on social life and human behaviour; they thus emphasise the social reach and long-term impact of information pollution.[37] Based on a survey of trainee teachers, *Firat and Kurt* define information pollution as unnecessary, incorrect, outdated, unreliable, inconsistent and commercialised information.[38] The term *'infollution'* was introduced by *PaekJae Cho* in 2002

[32] Iqbal and others (n 6).
[33] Karhade and others (n 8) 67.
[34] Rothkopf (n 7).
[35] Priyanka Meel and Dinesh Kumar Vishwakarma, 'Fake News, Rumor, Information Pollution in Social Media and Web: A Contemporary Survey of State-of-the-Arts, Challenges and Opportunities' (2020) 153 Expert Systems with Applications 112986, 3.
[36] Levent Orman, 'Fighting Information Pollution with Decision Support Systems' (1984) 1 Journal of Management Information Systems 64, 64.
[37] Kai-Yuan Cai and Chao-Yang Zhang, 'Towards a Research on Information Pollution', *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)* (1996) as cited in; Şebnem Özdemir, 'Individual Contributions to Infollution (Information Pollution): Trust and Share' (2016) 7 24.
[38] Mehmet Firat and AdiLe Aşkim Kurt, 'The Views of Candidate Teachers of Information Technologies about

to describe the unintended negative effects of information technology.[39] The concept of *'information pollution'* was also taken up by *Nielsen* in 2003.[40] Both *Cho* and *Nielsen* emphasise the negative impact that information systems can have on society, for example by compromising the quality of information and flooding it with irrelevant or misleading content.[41] In their framework *Wardle and Derakhshan*, propose to use the term *'information pollution'* as an umbrella category for different forms of disrupted information flows, including misinformation, disinformation and malinformation.[42] *Whitney Phillips and Ryan Milner* prefer the term *'polluted information'* to emphasise that not only the information landscape but the information itself is polluted with toxic content; their ecologically inspired approach shifts the focus from intention to the consequences of low quality information, like the rise of the reactionary far right after the 2016 US elections.[43]

The variety of approaches to describing qualitative information pollution makes it clear that this is a multidimensional phenomenon which encompasses content-related inadequacies and deliberate distortions and takes into account harmful effects on society. Examples include manipulated election content that fuels political distrust[44], or misleading medical advice that results in dangerous self-treatment practices.[45] This category is closely linked to the quality of the information, which is defined by its accuracy, believability, completeness, relevance and reputation.[46] On the basis of the perspectives presented, qualitative information pollution can be summarised as follows: It refers to the devaluation, distortion or manipulation of the content

Information Pollution on the Internet' (WSEAS Press 2008) 104 <https://www.researchgate.net/publication/273202478_The_Views_of_Candidate_Teachers_of_Information_Technologies_about_Information_Pollution_on_the_Internet>.

39 Cho (n 9).

40 Nielsen (n 9) 76.

41 ibid 76–77; Cho (n 9).

42 Claire Wardle and Hossein Derakhshan, 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making' (2017) Council of Europe DGI(2017)09 5 <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 18 April 2025.

43 Whitney Phillips and Ryan M Milner, *You Are Here: A Field Guide for Navigating Polarized Speech, Conspiracy Theories, and Our Polluted Media Landscape* (The MIT Press 2021) as cited in; Yvonne M Eadon and Stacy E Wood, 'Combating Contamination and Contagion: Embodied and Environmental Metaphors of Misinformation' [2024] Convergence 13548565241255347, 13.

44 Lluis De Nadal and Peter Jančárik, 'Beyond the Deepfake Hype: AI, Democracy, and "the Slovak Case"' Harvard Kennedy School (HKS) Misinformation Review 1–2 <https://misinforeview.hks.harvard.edu/article/beyond-the-deepfake-hype-ai-democracy-and-the-slovak-case/> accessed 18 June 2025.

45 Karol Żmudka and others, 'ChatGPT – a Tool for Assisted Studying or a Source of Misleading Medical Information? AI Performance on Polish Medical Final Examination' (2024) 78 Annales Academiae Medicae Silesiensis 94, 102.

46 Richard Y Wang and Diane M Strong, 'Beyond Accuracy: What Data Quality Means to Data Consumers' (1996) 12 Journal of Management Information Systems 5, 22.

of information through errors, misleading information or loss of context. It includes both intentional and unintentional degradation of the quality of information and can occur regardless of its truthfulness, for example when true information is taken out of context, selectively presented or accompanied by misleading implications. Qualitative information pollution includes forms such as misinformation, disinformation and malinformation.

b. Contribution through AI-Generated Content

b.a.Recursive Training Loops

Recursive training loops are a major risk for qualitative pollution of the information sphere. Generative AI models rely on large amounts of high-quality but limited training data, which is why web data is increasingly being used.[47] It is predicted that suitable human generated data for LLMs could be exhausted by 2026, thus making training with artificial data inevitable.[48] The use of synthetic data is not only driven by necessity, but also by economics: Creating artificial content is much cheaper than spending time sourcing and curating real data.[49] This interaction leads to so-called recursive training loops (‘*autophagy*‘[50]), in which models reuse their own outputs.[51] Studies show that the accuracy of the content and the diversity of the information generated decrease with each iteration.[52] This leads to a downward spiral of information quality, where each successive cycle is based on inferior data, and the quality of the information generated continues to decline.[53] Technical countermeasures, such as the integration of fixed real data or fresh data sets, can only delay the decay without solving the structural problem.[54] The *Common Crawl* database for example, which is regularly updated with billions of new web pages, also uses data that may have already been generated by other generative models and forms the basis of models from well-known companies such as OpenAI (GPT-3) and LLaMA v1.[55]

---

[47]Xiaodan Xing and others, ‘On the Caveats of AI Autophagy’ (2025) 7 Nature Machine Intelligence 172, 172–173.

[48] Pablo Villalobos and others, ‘Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data’ (arXiv, 4 June 2024) 9 <http://arxiv.org/abs/2211.04325> accessed 22 April 2025.

[49] Jonathan Klein and others, ‘Synthetic Data at Scale: A Development Model to Efficiently Leverage Machine Learning in Agriculture’ (2024) 15 Frontiers in Plant Science 1, 12.

[50] Sina Alemohammad and others, ‘Self-Consuming Generative Models Go MAD’ (2024) 3 <https://scholars.houstonmethodist.org/en/publications/self-consuming-generative-models-go-mad> accessed 22 April 2025.

[51] Xing and others (n 47) 173.

[52]Gonzalo Martínez and others, ‘Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet’ in Fabio Cuzzolin and Maryam Sultana (eds), *Epistemic Uncertainty in Artificial Intelligence* (Springer Nature Switzerland 2024) 66–68.

[53] Xing and others (n 47) 174.

[54] ibid.

[55] Stefan Baack, ‘A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl’,

b.b.Hallucinations

Another phenomenon related to the quality of AI-generated content is AI *'hallucination'*. Hallucinations refer to content which is not representative, is senseless or fictitious information that nevertheless appears linguistically plausible.[56] Even OpenAI officially points out that while the answers generated by ChatGPT often sound convincing, their content may be incorrect or illogical.[57] The phenomenon affects large language models, and image, audio, video and multimodal AI applications.[58] In AI-generated text, this manifests itself in logic and inference errors, while image generators can produce anatomically incorrect representations or unrealistic details.[59] Citations and references are also frequently affected by this.[60]

The causes of hallucinations are many and varied, for example inconsistent human feedback as part of fine-tuning processes. [61] What all these systems have in common is that they generate content based on statistical probabilities, without any real understanding of meaning.[62] It is this architecture that leads to hallucinations: Because these systems are unable to detect errors and noise in the input data, and because they have no real semantic understanding, the output often contains errors, inconsistencies or made-up elements.[63] In addition, many systems show a strong tendency towards sycophancy, i.e. an effort to confirm user expectations, even at the expense of content accuracy.[64]

Hallucinations clearly contribute to qualitative information pollution, as they represent a devaluation and distortion of information content, characterised by inaccuracy, misleading

The 2024 ACM Conference on Fairness, Accountability, and Transparency (ACM 2024) 2201 <https://dl.acm.org/doi/10.1145/3630106.3659033> accessed 22 April 2025; Xing and others (n 47) 172–173; 'Open Data for the World: Common Crawl Frees Petabytes of Web Data' (22 May 2024) <https://datos.gob.es/en/blog/open-data-world-common-crawl-frees-petabytes-web-data> accessed 9 May 2025.

56 Gernot Beutel, Eline Geerits and Jan T Kielstein, 'Artificial Hallucination: GPT on LSD?' (2023) 27 Critical Care 148; Ziwei Ji and others, 'Survey of Hallucination in Natural Language Generation' (2023) 55 ACM Comput. Surv. 248:1, 248:3; Katja Filippova, 'Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data' [2020] Findings of the Association for Computational Linguistics: EMNLP 2020 864, 864.

57 'Introducing ChatGPT' (13 March 2024) <https://openai.com/index/chatgpt/> accessed 22 April 2025.

58 Yujie Sun and others, 'AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content' (2024) 11 Humanities and Social Sciences Communications 1, 2.

59 Dr Julien Siebert, 'Halluzinationen von generativer KI und großen Sprachmodellen (LLMs) - Blog des Fraunhofer IESE' (*Fraunhofer IESE*, 20 September 2024) <https://www.iese.fraunhofer.de/blog/halluzinationen-generative-ki-llm/> accessed 22 April 2025.

60 Terence Day, 'A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT' (2023) 75 The Professional Geographer 1024, 1025–1026.

61 Siebert (n 59); Sun and others (n 58) 9–12.

62 Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) 616 <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 22 April 2025.

63 Sun and others (n 58) 11.

64 Mrinank Sharma and others, 'Towards Understanding Sycophancy in Language Models' (2024) 2 <http://arxiv.org/abs/2310.13548> accessed 27 May 2025.

information or loss of context. They lead to a lasting impairment of the quality of information and therefore fall into the categories of unintentional misinformation and disinformation.

b.c.Decontextualisation

Another contribution of AI-generated content to qualitative information pollution could be the decontextualisation of information. Generative AI models such as LLMs create content purely based on statistical probabilities, without any understanding of the original context of origin or the meaning of the information used.[65] There could therefore be a theoretical risk that information will be taken out of its context and misrepresented. This could lead to distortions, particularly in the case of automated summaries or the processing of informal sources. However, this relationship has not yet been sufficiently researched. If this assumption is confirmed, the resulting loss of context would be a form of qualitative information pollution, as even correct content can be misleading.

b.d.Bias and Discrimination

Bias and discrimination contribute to the qualitative pollution of the information sphere. They distort information by overemphasising certain perspectives and marginalising others, thereby undermining objectivity and permanently compromising the integrity of the information sphere.[66] Generative AI systems exacerbate this problem because they rely on existing training data, which is already embedded with social biases and structural inequalities.[67] Especially the logics of distribution, which are driven by the platform, favour the visibility of polarising or stereotypical content, which further manifests existing biases.[68] AI-generated content therefore contributes to the passive reproduction of discrimination, and even actively acts as a catalyst for the dissemination of biased information.[69]

---

[65] Emily M Bender and Alexander Koller, 'Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data', *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics 2020) 5187 <https://www.aclweb.org/anthology/2020.acl-main.463> accessed 23 April 2025.

[66] See Kate Crawford, 'Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence', *Atlas of AI* (Yale University Press 2021) Chapter 4, 123-126 for a critical examination of discriminatory information as informational basis.

[67] Tolga Bolukbasi and others, 'Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings' NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems 4356, 3.

[68] William Lucas, 'The Influence of Social Media Algorithms on Racial and Ethnic Misinformation: Patterns and Impacts' (2024) 9.

[69] Yuqi Zhou and others, 'Source Echo Chamber: Exploring the Escalation of Source Bias in User, Data, and Recommender System Feedback Loop' (arXiv, 28 May 2024) 7 <http://arxiv.org/abs/2405.17998> accessed 23 April 2025.

b.e.Intentional Disinformation

Targeted deception through AI-generated content is a serious form of qualitative information pollution. Whereas disinformation used to require human intervention and limited resources, generative AI now enables the automated, scalable and deceptively authentic production of manipulative content.[70] The main forms of AI-based disinformation include deepfakes, synthetic text, and artificially generated images and audio.[71] Deepfakes, are a use of AI which take the quality of disinformation to a new level: Highly realistic video or audio, in which people appear to speak or act authentically[72], can be used in a targeted way to spread credible false narratives.[73] AI technologies are able to create deepfakes in a very short time, by imitating voices, facial expressions and gestures in a deceptively realistic way.[74] The technical perfection of this content increases its persuasiveness, and makes it harder to detect.[75] These forms of disinformation lead to the deliberate manipulation and distortion of information by creating deceptively authentic but false or misleading content.[76]

3. Quantitative Pollution of the Information Sphere

a. Definition and Content

Quantitative information pollution describes the structural overload of the information sphere caused by an excessive amount of content whose sheer volume and redundancy impairs the usability and visibility of information. In an information-rich environment, it is not the information itself that is scarce, but the attention of the recipients, whose limited cognitive capacity is overwhelmed by the mass of content.[77] As a result, relevant information is lost in the digital overload and efficient attention management becomes increasingly difficult.[78] Within the information sphere, the phenomenon of quantitative information pollution especially

[70] Beatriz Saab, 'Manufacturing Deceit: How Generative AI Supercharges Information Manipulation' (National Endowment for Democracy (NED) 2024) 4 <https://www.ned.org/manufacturing-deceit-how-generative-ai-supercharges-information-manipulation/> accessed 28 April 2025.

[71] Noémi Bontridder and Yves Poullet, 'The Role of Artificial Intelligence in Disinformation' (2021) 3 Data & Policy e32, e32-3.

[72] Agnieszka M Walorska, 'Deepfakes and Disinformation' (Friedrich-Naumann-Stiftung für die Freiheit 2020) 9 <https://shop.freiheit.org/#!/Publikation/897> accessed 23 April 2025.

[73] Bontridder and Poullet (n 71) e32-4.

[74] Mika Westerlund, 'The Emergence of Deepfake Technology: A Review' (2019) 9 Technology Innovation Management Review 39, 40.

[75] ibid 42.

[76] Mohamed R Shoaib and others, 'Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models' (IEEE 2023) 1 <http://arxiv.org/abs/2311.17394> accessed 28 April 2025.

[77] Leila Shahrzadi and others, 'Causes, Consequences, and Strategies to Deal with Information Overload: A Scoping Review' (2024) 4 International Journal of Information Management Data Insights 100261, 4.

[78] Josef Falkinger, 'Limited Attention as a Scarce Resource in Information-Rich Economies' (2008) 118 The Economic Journal 1596, 1596, 1602.

affects the structural characteristics of (2) *information* and (4) *relational dynamics*.

In the literature*, information overload* is not uniformly defined. Rather, it is discussed in close connection with related concepts such as i*nfobesity*, *information anxiety*, *cognitive overload* or *communication overload,* which describe the cognitive overload caused by information flows.[79] The overload acts like noise: a density of stimuli that undermines learning, attention and decision-making skills.[80] *Orman* recognised early on that information pollution can be understood not only qualitatively but also quantitatively as *'information overload'*, especially when decision-makers are confronted with irrelevant or inconsistent information.[81] As soon as the costs of processing information exceed the expected benefits, they are forced to discard information arbitrarily or without checking it.[82] *Bawden, Holtham and Courtney* echo this idea, by describing overload as a state in which information loses its functional value and creates confusion rather than orientation.[83] *Cai and Zhang* also speak of a '*widespread*' presence of unwanted messages that can disrupt the social balance of information.[84] *Cho* refers to information overload with the term '*infollution'*.[85] This study uses the definition of *Belabbes et al.,* as it captures the key characteristics of information overload:

> '*Information overload is a negative psychological state in which individuals feel that they are receiving too much information, which hinders their ability to carry out their tasks. IO manifests itself through emotional and cognitive challenges and is most likely to happen through intrinsic and extrinsic information characteristics, poorly defined information needs, their working environment, brain ability and cognition, or the information environment.* […]*'*[86]

[79] Mohamed Amine Belabbes and others, 'Information Overload: A Concept Analysis' 79 Journal of Documentation 144, 3.
[80] Orrin Edgar Klapp, *Overload and Boredom: Essays on the Quality of Life in the Information Society* (Greenwood Publishing Group Inc 1986) as cited in:; Angela Edmunds and Anne Morris, 'The Problem of Information Overload in Business Organisations: A Review of the Literature' (2000) 20 International Journal of Information Management 17, 18–19.
[81] Orman (n 36) 64–65.
[82] ibid.
[83] D Bawden, C Holtham and N Courtney, 'Perspectives on Information Overload' (1999) 51 ASLIB PROCEEDINGS 249.
[84] Cai and Zhang (n 37) as cited in; Özdemir (n 37) 24.
[85] Cho (n 9).
[86] Belabbes and others (n 79) 12–13.

b. Contribution through AI-Generated Content

b.a.Mass Generation of (Partially) Synthetic Content

AI-generated content promotes the mass production of (partially) synthetic content and thus contributes to the flooding of the information sphere.[87] The ability of generative AI systems to produce large amounts of text, images or video in a short period of time[88], leads to information pollution based on sheer quantity.[89] Automated systems power entire networks of websites that publish thousands of AI-generated articles every day, thereby flooding search results and digital information sources with irrelevant or repetitive content.[90] Information overload distorts perception and makes it difficult to distinguish between relevant and irrelevant information.[91] This effect is exacerbated by the algorithmic incentive structures of platforms that value reach and visibility based on quantity and interaction potential rather than quality.[92] TikTok´s 'For You' feed for example strongly amplifies videos that align with the user's past interactions.[93] As a result, repetitive AI-generated clickbait gains visibility, while diverse or expert content remains less visible.[94] In addition, AI makes it easier for human authors to produce faster and longer texts, which also contributes to the amount of information available in the information sphere.[95] The mass of AI-generated content thus exacerbates the state of information overload, in which the cognitive and emotional overload of recipients impairs their perception of relevant content.[96]

---

[87] Spennemann (n 1) 3.

[88] Divya, V and Mirza, Agha Urfi, 'Chapter 8, Transforming Content Creation: The Influence of Generative AI on a New Frontier', *Exploring the Frontiers of Artificial Intelligence and Machine Learning Technologies* (1st edn, San International Scientific Publications 2024) 2 <https://www.researchgate.net/publication/379654562> accessed 24 April 2025.

[89] Bhaskar Mitra, Henriette Cramer and Olya Gurevich, 'Sociotechnical Implications of Generative Artificial Intelligence for Information Access' (arXiv, 16 July 2024) 3 <http://arxiv.org/abs/2405.11612> accessed 24 April 2025 This paper is now published in; Bhaskar Mitra, Henriette Cramer and Olya Gurevich, 'Sociotechnical Implications of Generative Artificial Intelligence for Information Access' in Ryen W White and Chirag Shah (eds), *Information Access in the Era of Generative AI* (Springer Nature Switzerland 2025).

[90] McKenzie Sadeghi and Lorenzo Arvanitis, 'Rise of the Newsbots: AI-Generated News Websites Proliferating Online' (*NewsGuard*, 1 May 2023) <https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating> accessed 24 April 2025.

[91] Belabbes and others (n 79) 4.

[92] Introductional Overview: Nicole Immorlica, Meena Jagadeesan and Brendan Lucier, 'Clickbait vs. Quality: How Engagement-Based Optimization Shapes the Content Landscape in Online Platforms', *Proceedings of the ACM Web Conference 2024* (ACM 2024) 36 <https://dl.acm.org/doi/10.1145/3589334.3645353> accessed 24 April 2025.

[93] Fabian Baumann and others, 'Dynamics of Algorithmic Content Amplification on TikTok' (arXiv, 26 March 2025) p.2-4 <http://arxiv.org/abs/2503.20231> accessed 18 June 2025.

[94] Sadiya Iqbal, 'How Social Media Algorithms Affect Content Visibility' (*Content Whale*, 28 April 2025) <https://content-whale.com/blog/impact-of-social-media-algorithms-on-content-visibility/> accessed 18 June 2025.

[95] Zhicheng Lin, 'Techniques for Supercharging Academic Writing with Generative AI' (2025) 9 Nature Biomedical Engineering 426, 426.

[96] Hu Zi-yang, 'AIGC Related Context: A New Communication Culture For Human' (2024) 14 Journal of

b.b.Automated Bots and Spam

Automated social bots are able to dominate digital discourse through mass contributions, and flood social media with spam, disinformation and irrelevant content.[97] The rise of generative AI is increasing the spread of spam, as LLMs are able to generate messages with a high frequency and linguistic quality that can increasingly evade traditional detection mechanisms.[98] An example of the use of automated bots is the mass distribution of manipulated reviews in online review systems, which deliberately deceive consumers and make access to reliable information more difficult.[99] In political contexts, bots are used specifically to manipulate public opinion, stifle debate and fill social media with automated posts.[100]

4. Formal Pollution of the Information Sphere

a. Definition and Content

Formal information pollution describes a disruption of the information sphere that is not aimed at the content quality (qualitative) or the volume (quantitative), but at the way in which information is presented, structured and made accessible. In digital environments, it is not only the content itself that is important, but also its presentation, especially given the growing volume of information.[101] Formal information pollution primarily concerns the fourth structural element of the information sphere: *relational dynamics* – the interactions and mechanisms that shape the presentation, accessibility and reception of information.

Formal pollution is closely related to quantitative information pollution[102]: Without structuring elements, relevant information runs the risk of being lost in information overload.[103] In the literature, it has so far usually only been described as an appearance of information overload[104], although its influence can be analysed as a separate category. Formal information pollution is characterised by unstructured presentations, such as those found on websites without clear

Literature and Art Studies 921, 922.

97 Emilio Ferrara and others, 'The Rise of Social Bots' (2016) 59 Communications of the ACM 96.

98 Stefano Cresci, 'A Decade of Social Bot Detection – Communications of the ACM' (2020) 63 Communications of the ACM 72, 77.

99 Yingying Ma and Fengjun Li, 'Detecting Review Spam: Challenges and Opportunities', *Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (IEEE 2012) 651 <http://eudl.eu/doi/10.4108/icst.collaboratecom.2012.250640> accessed 24 April 2025.

100 MC Forelle and others, 'Political Bots and the Manipulation of Public Opinion in Venezuela' [2015] SSRN Electronic Journal 1, 1–2.

101 Starr R Hiltz and Murray Turoff, 'Structuring Computer-Mediated Communication Systems to Avoid Information Overload' (1985) 28 Communications of the ACM 680, 680, 681.

102 Shahrzadi and others (n 77) 9.

103 Belabbes and others (n 79) 11.

104 Firat and Kurt (n 38) 102; Ramesh Pandita, 'Information Pollution, a Mounting Threat: Internet a Major Causality' (2014) 2 Journal of Information Science Theory and Practice 49, 52.

structure, navigation or meaningful categories.[105] It can occur when content is difficult to access, for example through paywalls or when a lack of visual hierarchy makes orientation difficult.[106] Formal information pollution thus weakens important functions of the information sphere: the targeted search, evaluation and categorisation of content.

b. Contribution through AI-generated Content

Although AI is contributing to the structuring, categorisation and filtering of content, the content generated by AI itself, whose lack of transparency in terms of origin, source base and creation process poses challenges for the integrity of the information sphere. AI generative models are often trained on opaque data sets whose origin and composition are not disclosed and often users are not even informed that the content was AI-generated in the first place.[107] For users of the infosphere, this opacity makes it difficult to assess the reliability of AI-generated content, which thereby undermines trust in such information.[108] Without verifiable references, a prerequisite for rational discourse, namely the ability to evaluate and categorise sources is undermined.

5. Direct Consequences

A polluted information sphere directly impacts how individuals perceive and process information. In particular, the efficient search for reliable information becomes much more difficult, as users in an overloaded and low-quality information environment find it difficult to distinguish relevant content from irrelevant or erroneous content.[109] *Cho et al.* show that information overload leads to disorientation and suboptimal decisions.[110] In addition, filter bubbles and echo chambers thrive in such conditions, because users tend to retreat into homogeneous information spheres to escape any cognitive overload.[111] Furthermore, the

[105] Pandita (n 104) 52.
[106] Max Levy, 'Information Pollution & the Public Discourse' (2022) 6 Journal of Law and Technology at Texas (JOLTT) 113, 139.
[107] See for example: Roxana Daneshjou and others, 'Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review' (2021) 157 JAMA Dermatology 1362, 1365.
[108] Jack Hardinges, Elena Simperl and Nigel Shadbolt, 'We Must Fix the Lack of Transparency Around the Data Used to Train Foundation Models' [2024] Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/xau9dza3/release/2> accessed 24 April 2025.
[109] Martin J Eppler and Jeanne and Mengis, 'The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines' (2004) 20 The Information Society 325, Table 4, 333.
[110] Byung-Kwan Lee and Wei-Na Lee, 'The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment' (2004) 21 Psychology & Marketing 159, 161.
[111] Emerson Palmieri, 'Online Bubbles and Echo Chambers as Social Systems' (5 March 2025) 13 <https://www.emerald.com/insight/content/doi/10.1108/K-09-2023-1742/full/html> accessed 25 April 2025 now published in; Emerson Palmieri, 'Online Bubbles and Echo Chambers as Social Systems' (2025) 54 Kybernetes 2457.

constant exposure to opaque or low-quality content accelerates the erosion of public trust in digital platforms, information sources and institutions.[112] To cope with information overload, users often rely on recommendation systems, which gradually suppress critical reflection through repeated confrontation with similar content.[113] Eventually this leads to the *'information fatigue syndrome'*, which weakens the ability to make decisions and judgements.[114] In response, people adapt coping strategies such as avoiding information or deliberate withdrawal, where they ignore potentially relevant but challenging information or they just limit their sources.[115] Ultimately, this behaviour reduces critical engagement with content and fosters a tendency to simplify complex issues or avoid inconvenient truths.[116] These consequences show that a polluted information sphere affects everyday-life by making it more difficult to access, process and evaluate information.

IV.Concluding Remarks on Chapter 1

The analysis demonstrates that AI-generated content contributes to the pollution of the information sphere. Through the mass generation of synthetic content, the occurrence of hallucinations, biases and targeted disinformation, the quality and the structure of the available information are permanently compromised. The classification into qualitative, quantitative and formal information pollution highlights that generative AI models can have a destabilising effect at all levels: They distort content, overload the information sphere with redundant data and deprive information of its context and comprehensibility. These developments have direct consequences for users and society. A polluted information sphere makes access to reliable information more difficult, promotes cognitive overload and leads to withdrawal tendencies and an erosion of trust in digital information sources.

These findings provide the foundation for further legal assessment, especially with regard to the question of whether the protection of the information sphere can be a legitimate aim for restricting freedom of expression.

[112] Mitra, Cramer and Gurevich (n 89) 10.
[113] Sofia Bonicalzi, Mario De Caro and Benedetta Giovanola, 'Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems' (2023) 42 Topoi 819, 829.
[114] Pandita (n 104) 55.
[115] David Bawden and Lyn Robinson, 'Information Overload: An Introduction', *Oxford Research Encyclopedia of Politics* (Oxford University Press 2020) 16.
[116] ibid 13.

Chapter 2: AI-Generated Content within the Scope of Article 10 §1 ECHR

This chapter examines the second sub-question: *In light of the pollution of the information sphere, does AI-generated content fall within the scope of protection under Article 10 §1 ECHR, particularly in relation to freedom of expression and freedom of information?* Article 10 §1 ECHR states that *'everyone has the right to freedom of expression'*, including the *'freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers'*. Given the increasing quantitative, qualitative and formal pollution of the information sphere by mass-produced, potentially misleading or repetitive content, the question arises whether AI-generated content can claim the protection of Article 10 §1 ECHR regarding freedom of expression and freedom of information. Article 10 §1 ECHR also protects other freedoms of communication, including freedom of the press, freedom of broadcasting, freedom of artistic and academic expression.[117] These sub-freedoms may also be relevant to AI-generated content. However, due to the limited scope of this thesis, the focus is on freedom of expression and information. Overlaps with other areas are included selectively.

The chapter shows that AI-generated content is not categorically excluded from protection under Article 10 §1 ECHR. However, this protection depends on whether the content can be functionally linked to public discourse or personally attributed to a human actor. The analysis is divided into two main sections: First, the material scope of protection is analysed. This includes an examination of whether AI-generated content can fall under the concept of *'expression'* and the extent to which Article 17 ECHR acts as a barrier to the scope of protection (I). Then, the personal scope of protection will be considered to clarify which natural or legal persons can invoke Article 10 §1 ECHR in the context of AI-generated content (II).

I. Material Scope of Protection

The first step in assessing the applicability of Article 10 §1 ECHR to AI-generated content is to examine whether such content falls within the material scope of protection. The analysis is conducted in two steps: First, the concept of *'expression'* is examined for its applicability to AI-generated content by using recognised interpretation methods, including textual, systematic, teleological and historical interpretation. Second, the extent to which the abuse clause

[117] Christoph Grabenwarter and Katharina Pabel, '§ 23. Politische Und Gemeinschaftsbezogene Grundrechte' in Grabenwarter and Pabel (eds), *Europäische Menschenrechtskonvention* (7th edn, 2021) para.2.

Article 17 ECHR can act as a barrier is analysed, as it may exclude AI-generated content from the protection of Article 10 §1 ECHR when such content deliberately undermines democratic structures. The result shows that AI-generated content is not generally excluded from the material scope of protection but requires contextual differentiation.

1. AI-Generated Content as '*expression*'

Whether AI-generated content falls within the scope of protection of Article 10 §1 ECHR depends on the interpretation of the term *'expression'*. The analysis follows a four-step structure: It begins with an examination of the wording, then turns to the systematic categorisation within the ECHR, and in the third step includes the historical contexts of origin and a teleological interpretation. Within each of these levels of interpretation, the extent to which the relevant criteria can be applied to AI-generated content is examined before a concluding assessment is made. The result shows that AI-generated content can only be protected if it is either an expression of individual autonomy or makes a functional contribution to public discourse.

a. Wording and Ordinary Meaning

The term *'expression'* in Article 10 §1 ECHR does not in principle exclude the inclusion of AI-generated content. The Article itself does not provide a more detailed definition of *'expression'*. The *Oxford English Dictionary* defines *'expression'* as, inter alia '*The action or process of representing, demonstrating, indicating, etc., something, as through an action, manner, appearance, state of affairs, etc*.'[118] This definition suggests a subjective act of communication and emphasises the personal dimension of expression.

The ECtHR has traditionally interpreted the concept of expression broadly, and recognized non-verbal forms such as works of art, clothing, gestures or symbols as worthy of protection, provided there is a communicative content.[119] Both the form and the content of expressed ideas and information are protected.[120] Against this background, the question arises as to whether AI-generated content can also be regarded as a form of *'expression'*. Certainly, it is not a new <u>form</u> of expression *per se*, but a new <u>source</u> of content. Nevertheless, it could be argued that machine-generated content is merely a different technical realisation of expression, comparable to the use of tools or software by humans.

[118] 'Expression' sense II.3.b <https://doi.org/10.1093/OED/8156500661> accessed 2 May 2025.

[119] *Müller and Others v Switzerland* [1988] ECtHR 10737/84 [27]; *Bédat v Switzerland* [2016] ECtHR [GC] 56925/08 [58]; *Gough v the United Kingdom* [2018] ECtHR 2153/15 [53]; *Ibragimova v Russia* [2022] ECtHR 68537/13 [22].

[120] *Mariya Alekhina and Others v Russia* [2018] ECtHR 38004/12 paras.197,203-206.

However, the link between the term "*expression*" and the personal carrier "*everyone*" in Article 10 §1 ECHR, which suggests a human form of expression, argues against a broad interpretation.[121] The wording is therefore not unambiguous. Nevertheless, considering the dynamic interpretation by the ECtHR, which regularly incorporates new forms of expression and technological developments, for example by recognising freedom of the press for online media[122] or an app protected by Article 10 §1 ECHR[123], as well as the frequently broad interpretation[124], the use of the word *'expression'* is no reason to exclude AI-generated content.

b. Systematic Interpretation

The systematic placement of Article 10 ECHR argues against the inclusion of AI-generated content within the scope of protection. Systematic interpretation examines the position of a provision within the structure of the Convention and its relation to other norms. The proximity of '*expression*' to '*everyone*' within the same sentence of Article 10 §1 ECHR suggests a close conceptual link, which points to a personal dimension of the right and indicates that the right to expression is granted specifically to identifiable legal subjects.[125] The ECHR is consistently oriented towards the protection of specific legal entities (natural or legal persons) whereas technical systems are neither the addressee nor the object of protection of the Convention. This basic orientation is also reflected in the title and preamble of the ECHR - European Convention on *Human* Rights.[126] Articles 34 and 35(3a) ECHR confirm this by reserving individual complaints exclusively to human beings, thus systematically emphasise that only legal positions supported by human beings are enforceable at all.[127]

c. Historical Context and Teleological Interpretation

Historical interpretation considers the legal, political and social context in which a provision was adopted and thereby helps to uncover the original understanding of a term. This approach is recognised under Article 32 of the Vienna Convention on the Law of Treaties and reflected in ECtHR practices.[128] The historical context of Article 10 ECHR shows that the protection of freedom of expression was conceived as a reaction to the systematic violations of human rights

---

121 De Vries (n 10) 99.
122 *Times Newspapers Ltd (nos 1 and 2) v United Kingdom* [2009] ECtHR App nos 3002/03 and 23676/03 [27].
123 *Magyar Kétfarkú Kutya Párt v Hungary* [2018] ECtHR [GC] 201/17 [37].
124 *Handyside v the United Kingdom* [1976] ECtHR 5493/72, ECLI:CE:ECHR:1976:1207JUD000549372 [49]; *Oberschlick v Austria* [1991] ECtHR 11662/85 [57].
125 De Vries (n 10) 99.
126 ibid.
127 ibid.
128 *Golder v. the United Kingdom* (n 18) paras.29-36; *Demir and Baykara v. Turkey* (n 18) paras.65-68; *Magyar Kétfarkú Kutya Párt v. Hungary* (n 123) para.136.

during the Second World War and as a counter-proposal to the Stalinism that was establishing itself in Eastern Europe.[129] With the Convention, the Council of Europe sought to create an institutional foundation that would safeguard liberal-democratic values and counter authoritarian tendencies in the long term.[130] The ECtHR made it clear early on that, '*freedom of expression constitutes 'one of the essential foundations' of […] a [democratic] society […]*'.[131] Historically, the focus lies therefore not primarily on the protection of the speaker, but on the protection of a liberal and democratic society.

Teleological interpretation focuses on the object and purpose of the ECHR as a living instrument.[132] This form of interpretation can also be found in Article 31 Vienna Convention on the Law of Treaties and is recognised by the ECtHR.[133] Closely linked to this origin, Article 10 §1 ECHR today has a dual protective function: on the one hand, it guarantees the individual freedom to form and express opinions as an expression of personal autonomy and human dignity.[134] On the other hand, it guarantees open communication as a prerequisite for democratic discourse and the forming of opinions.[135] The *telos* thus goes beyond the mere protection of the interests of individual speakers and aims at preserving a free, pluralistic public sphere.[136] This can be differentiated into two dimensions: Freedom of expression for personal autonomy and individual self-fulfilment and freedom of information for public discourse and individual self-fulfilment. These two dimensions are interdependent: The exercise of freedom of expression requires access to information, while a functioning information sphere relies on individuals being able to disseminate and debate ideas freely.

c.a.Freedom of Expression: Personal Autonomy and Individual Self-Fulfilment

The first dimension concerns the realisation of human dignity and individual self-fulfilment through the exercise of Article 10 ECHR.[137] It encompasses the content of the general *freedom of expression*, which in other legal systems is sometimes referred to more narrowly as '*freedom*

---

[129] Imara McMillan, 'Enforcement Through the Network: The Network Enforcement Act and Article 10 of the European Convention on Human Rights' (2019) 20 Chicago Journal of International Law 255, 268.
[130] Ed Bates, *The Evolution of the European Convention on Human Rights: From Its Inception to the Creation of a Permanent Court of Human Rights* (Oxford Academic 2011) 6–8.
[131] *Handyside v. UK* (n 124) para.49.
[132] *Tyrer v the United Kingdom* [1978] ECtHR 5856/72 para.31.
[133] *Selmouni v France* [1999] ECtHR [GC] 26856/06, 26449/08, 61815/09, 30395/10, 50999/10, 58287/10, 63120/10, 36942/11 para.74; *Leyla Şahin v Turkey* [2005] ECtHR [GC] 44774/98 para.137.
[134] Matthias Cornils, 'EMRK Art. 10 Freiheit Der Meinungsäußerung' in Gersdorf and Paal (eds), *BeckOK Informations- und Medienrecht* (47th edn, 2024) para.1.
[135] ibid; *Handyside v. UK* (n 124) para.49.
[136] *Centro Europa 7 S.r.l and Di Stefano v Italy* [2012] ECtHR [GC] 38433/09 [129].
[137] Cornils (n 134) para.1.

of speech'[138], a term which, however, linguistically abbreviates the fact that Article 10 §1 ECHR also protects other forms of expression far beyond speech.[139] The ECtHR has consistently interpreted the scope of Article 10 §1 ECHR broadly, taking into account the protection of individual self-fulfilment[140]: In *Handyside v. UK* the Court emphasised that freedom of expression applies not only to information or ideas that are perceived as pleasant or harmless, but also to those that '*offend, shock or disturb*'.[141] Article 10 §1 ECHR also covers untrue statements[142] (except when they are manifestly false factual claims[143]) and expressions that a person adopts as their own.[144] The ECtHR complements this openness of content with a broad protection of forms of expression.[145] What matters is not the specific form, the actual author or the specific content of the expression, but whether it is an expression of individual autonomy to form, hold and communicate opinions or thoughts.[146] Article 10 §1 ECHR therefore protects not only speech but also for example artistic expression[147], academic speech[148] or commercial expression[149].

If the purpose of protection is to be transferred to AI-generated content, there must be a case-by-case examination of whether the specific contribution can be recognised as an expression of individual autonomy and self-fulfilment. This means that a person must either have actively contributed to the creation of AI-generated content or, by adopting it as their own, make it attributable to them.[150] To illustrate this, we return to the example previously cited (p.8). Person B disseminates a text that was not written by them but was instead generated by a language model and briefly prompted by B. Despite the fact that the contribution did not emerge from their own cognitive processes, B appropriates the content, publishes it consciously, and thus positions themselves within the public discourse. Consequently, the blogpost becomes an articulation of their opinion, even if it materialised solely because of the publication.

---

138 See for example: United States Constitution 1791 Amendmend I.
139 *Müller and Others v. Switzerland* (n 119) para.58; *Gough v. the United Kingdom* (n 119) para.53; *Ibragimova v. Russia* (n 119) para.22.
140 *Handyside v. UK* (n 124) para.49; *Stoll v Switzerland* [2007] ECtHR [GC] 69698/01 para.101.
141 *Handyside v. UK* (n 124) para.49.
142 Grabenwarter and Pabel (n 117) para.7.
143 *E.s v Austria* [2018] ECtHR 38450/12 [55].
144 *Steel and Morris v the United Kingdom* [2005] ECtHR 68416/01 paras.91-95; Cornils (n 134) para.13.
145 *Müller and Others v. Switzerland* (n 119) para.27; *Bédat v. Switzerland* (n 119) para.58; *Gough v. the United Kingdom* (n 119) para.53; *Ibragimova v. Russia* (n 119) para.22.
146 *Magyar Kétfarkú Kutya Párt v. Hungary* (n 123) para.37.
147 *Müller and Others v. Switzerland* (n 119) para.27.
148 *Sorguç v Turkey* [2009] ECtHR 17089/03 para.35.
149 *Casado Coca v Spain* [1994] ECtHR 15450/89 para.35.
150 Cornils (n 134) para.13; Jan Oster, 'Haftung für Persönlichkeitsrechtsverletzungen durch Künstliche Intelligenz' (2018) 82 UFITA 14, 21.

Collaborative scenarios, such as the use of tools like *Grammarly* that suggest stylistic changes but leave the human in control over content and intent may likewise qualify as an expression of personal autonomy. A contrasting example makes the distinction even clearer: An automated bot posts content on Instagram multiple times daily, with said content generated by an LLM. The operator of the bot is solely responsible for its technical oversight; however, they do not engage in any selection processes, undertake content checks, or possess a clearly defined communicative objective. The utilisation of bots is employed to enhance visibility and interactions. Notwithstanding the relevance of the content, it does not constitute an expression of individual autonomy. Consequently, it does not represent an individual communicative act of self-fulfilment.

c.b.Freedom of Information: Public Discourse and Individual Self-Fulfilment

The second dimension concerns the other purpose of Article 10 ECHR, which is to protect public discourse and thus the foundations of a democratic society.[151] Freedom of information is attributed to this domain.[152] It is a closely related concept to freedom of the press and serves not only the individual interest in knowledge but also fulfils a central function in democratic discourse.[153] Alongside active freedom of information, which is defined as the right to inform others, Article 10 §1 ECHR also protects passive freedom of information, which is the right to unhindered access to information that is generally available.[154] The protection is not limited solely to the receipt of information, it also extends to the efforts made to gain access to sources of information.[155] Freedom of information thus makes a substantial contribution to the preparation and facilitation of informed public debates and thus to the openness and plurality of the democratic information sphere.[156] In this sphere, democratic decision-making unfolds through the dissemination of opinions and information, as well as through access to the views of others.[157] Moreover, freedom of information protects the individual self-fulfilment and therefore is not limited to political content but includes access to for example artistic or cultural content.[158]

---

[151] *Handyside v. UK* (n 124) para.49.
[152] Rainer Hofmann, 'Internationaler Menschenrechtsschutz, 1. Teil: Universeller Menschenrechtsschutz' (2012) 46 <https://www.jura.uni-frankfurt.de/43671833/Skript_Internationaler_Menschenrechtsschutz_SS_12.pdf?> accessed 2 May 2025.
[153] *Társaság a Szabadságjogokért v Hungary* [2009] ECtHR 37374/05 paras.27-28.
[154] Grabenwarter and Pabel (n 117) paras.8-9.
[155] *Shapovalov v Ukraine* [2012] ECtHR 45835/05 [68].
[156] *Társaság a Szabadságjogokért v. Hungary* (n 153) para.27.
[157] Dirk Voorhoof and Hannes Cannie, 'Freedom of Expression and Information in a Democratic Society: The Added but Fragile Value of the European Convention on Human Rights' (2010) 72 International Communication Gazette 407, 421.
[158] Cornils (n 134) para.8.

The protective purpose of freedom of information, to safeguard and promote public discourse, does not in itself preclude the inclusion of AI-generated content. In the context of active freedom of information, AI-generated content can make a functional contribution to public dialogue, for example by providing new perspectives or adding to existing debates. Even if there is no traditional human authorship, protection is not automatically called into question as long as the content promotes public communication and thus fulfils the fundamental purpose of Article 10 ECHR.[159] AI-generated content can also be covered under the passive freedom of information. What matters is not whether the information is generated by a machine or a human, but whether it is accessible and relevant to a human recipient.[160] The origin of the information is secondary to its contribution to the formation of individual and collective opinion.[161]

d. Discussion

On the basis of the interpretation results above, it must be assessed whether AI-generated content falls within the scope of protection of Article 10 §1 ECHR. The discussion is structured along the two central dimensions of protection: *freedom of expression* and *freedom of information*. It concludes that AI-generated content can generally be protected as part of public discourse under Article 10 §1 ECHR but enjoys only limited protection on the basis of the general *freedom of expression*, which protects the expression of individual autonomy.

d.a.Freedom of Expression: Personal Autonomy and Individual Self-Fulfilment

The general right to freedom of expression under Article 10 §1 ECHR only protects expression that is attributable to a human will. According to the wording and in line with the broad case law of the ECtHR, it is not decisive whether the form is verbal, written or symbolic. However, it is essential that there is a human act of expression or that such an act can be attributed to a human. This interpretation is also supported by the normative purpose of the article: The protection of individual self-development presupposes personal autonomy. In addition, systematic arguments support this perspective. The relation to the person-centred wording *'everyone'* and the ECHR's focus on individual legal protection make it clear that only human acts of expression are protected. Consequently, AI-generated content falls under Article 10 §1 ECHR if a person consciously makes it their own, for example, by deliberately publishing it, commenting on it or incorporating it into their own statements.

[159] De Vries (n 10) 100.
[160] ibid.
[161] Following the content-based approach, see Massaro, Norton and Kaminski (n 13) 2487–2491.; for a more intent-focused perspective, see Goswami (n 15) 569, 572.

d.b.Freedom of Information: Public Discourse

AI-generated content may fall within the scope of freedom of information under Article 10 §1 ECHR, provided that it makes a functional contribution to public discourse. The normative purpose of the human right allows for a functional interpretation of the content: The decisive factor is not who the information originates from, but whether it makes a functional contribution to public communication. Moreover, the wording *'everyone has the right to [...] receive and impart information and ideas [...]'* can be reconciled with such an interpretation: The human benefit is preserved, since any extension of the information sphere could in theory benefit the individual. Nor is there any systematic restriction, as long as a link to human reception is maintained. Nevertheless, it is important to note, that an overly broad interpretation of *'expression'* could lead to the protection of content that does not enrich public discourse, but instead undermines the integrity of the information sphere through redundancy, disinformation or algorithmically generated overproduction. However, the ECtHR understands the scope of protection very broadly and tends to address considerations like this rather at the level of justification of an interference.[162] Consequently, AI-generated content may fall within the scope of protection if it is intended to be published and received by society.

2. Limitation by Article 17 ECHR

The subsequent discussion aims to ascertain the extent to which Article 17 ECHR can limit the scope of protection afforded by Article 10 §1 ECHR. Freedom of expression does not only apply to content that is relevant to politics or society, but also includes seemingly irrelevant contributions, as these can also potentially contribute to the formation of public opinion[163], as well as content that '*offend, shock or disturb*'.[164] The ECtHR has deliberately adopted a broad approach to the scope of protection in this respect: Restrictions are not methodically applied at this level of Article 10 §1 ECHR, but are rather located at the level of the justification for an interference in Article 10 §2 ECHR.[165] This approach ensures the essential openness of public discourse in terms of content and prevents the state from prejudging the content of communications.[166] Nevertheless, the scope of protection of Article 10 §1 ECHR could be limited by the abuse-clause of Article 17 ECHR, if AI-generated content is *'aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention'*. The ECtHR shows a tendency that the limitation

[162] Cornils (n 134) paras.5,15-16.
[163] *C8 (canal 8) v France* [2023] ECtHR 58951/18, 1308/19 paras.44-47; Cornils (n 134) para.8.
[164] *Handyside v. UK* (n 124) para.49.
[165] *C8 (canal 8) v. France* (n 163) para.47; Grabenwarter and Pabel (n 117) para.5.
[166] Cornils (n 134) paras.5-6.

must be taken into account already at the level of the scope of protection.[167] The Court has applied Article 17 ECHR mainly to extremist speech, in particular racist speech[168] and Holocaust denial.[169] In addition, Article 17 ECHR has been applied in the review of party prohibition proceedings: political organisations cannot invoke Convention guarantees if their aim is to abolish fundamental democratic principles.[170]

AI-generated content can undoubtedly also be excluded from Convention protection under Article 17 ECHR if it conveys extremist narratives, as it is in this context not different from human-generated content. It should be noted, however, that the legal threshold for this is quite high and that such an exception would only be expected in obvious cases.[171]
When transferring the jurisprudence of party prohibitions to AI-generated content, machine-generated communications that are systematically used to destabilise democratic structures could be treated in the same way. Content whose sole purpose is to undermine fundamental democratic values could therefore be excluded from the scope of protection. A paradigmatic example of this could be the use of deepfakes that simulate public figures, like an audio deepfake of Slovak politician Michal Šimečka, where he apparently suggests electoral manipulation on the eve of an election.[172] Irrespective of this theoretical transferability, the procedural burden of justification and proof for such a very narrow exception to the scope of protection would be significant.[173] The mere possibility that AI-generated content could contribute to the pollution of the information sphere is not in itself sufficient to exclude it from the scope of protection. Harmful effects of the information pollution are therefore analysed at the level of justification under Article 10 §2 ECHR. Nevertheless, Article 17 ECHR should not be disregarded as a potential inherent obstacle in the review of the scope of protection in extreme cases.

[167] Overview: *Roj Tv a/S v Denmark (dec)* [2018] ECtHR 24683/14 paras.32-38.
[168] *Norwood v the United Kingdom (dec)* [2004] ECtHR 23131/03.
[169] *M'bala M'bala v France (dec)* [2015] ECtHR 25239/13 paras.39-42.
[170] *German Communist Party and Others v Germany* [1957] ECtHR 250/57 p.4-5; *Refah Partisi (the Welfare Party) and Others v Turkey* [2003] ECtHR [GC] 60936/12 [99].
[171] *Paksas v Lithuania* [2011] ECtHR [GC] 34932/04 [87]; *Perinçek v Switzerland* [2015] ECtHR [GC] 27510/08 [114].
[172] De Nadal and Jančárik (n 44) 1–2.
[173] Hannes Cannie and Dirk Voorhoof, 'The Abuse Clause in International Human Rights Law: An Expedient Remedy against Abuse of Power or an Instrument of Abuse Itself?', *Bullying and the Abuse of Power, James O'Higgins Norman, Sameer Hinduja (eds.)* (Inter-Disciplinary Press 2010) 121 <http://hdl.handle.net/1854/LU-1080884> accessed 9 May 2025.

3. Interim Conclusion

As a result, the material scope of protection of Article 10 § 1 ECHR for AI-generated content requires a distinction along its two dimensions: a human act of will or a functional contribution to public discourse. The analysis results in the following three key findings:

First, the general *freedom of expression* only protects content that is attributable to a human act of will. A contribution generated by generative AI therefore only falls within the scope of protection if a person appropriates this content and thereby acts in self-fulfilment and autonomy.

Second, AI-generated content may fall under *freedom of information* if it functionally contributes to public discourse, as the protection focuses on its communicative effect rather than its technical origin or authorship. Article 10 §1 ECHR thus retains its function as a guarantor of a democratic public sphere, even in the face of new forms of content production.

Third, Article 17 ECHR may also play a role as an intrinsic barrier. In extreme circumstances, AI-generated speech that serves solely to destabilise fundamental democratic principles may be excluded from the scope of protection. However, this exception must be applied narrowly and cannot be a blanket application for cases in which the information sphere is negatively affected.

In conclusion, a blanket exclusion from the material scope of Article 10 §1 ECHR in the case of AI-generated content is not dogmatically justified. Consequently, a contextualised evaluation is required that considers the distinct protective functions of these two elements: individual self-fulfilment and contributions to the democratic public sphere.

II.Personal Scope of Protection

Having clarified in the previous section the conditions under which AI-generated content falls within the material scope of protection under Article 10 §1 ECHR, the question now arises as to *who* enjoys this protection. The following section therefore analyses whether Article 10 §1 ECHR necessarily requires a rights holder and who can be considered as such in the context of AI-generated content.

1. No Protection Without a Holder

Even if it has been shown in the context of the material scope of protection that Article 10 §1 ECHR protects not only individual acts of expression but also public discourse as such[174], it does not protect it in isolation from individual legal positions. The wording of Article 10 §1 ECHR (*'everyone'*) and the case-law of the ECtHR show that natural and legal persons are the subject of freedom of expression.[175] Despite the substantive nature of this matter, the procedural structure of the Convention can also be used to justify the connection to individuals: According to Article 34 ECHR, only the person who is specifically affected as a *'victim'* can lodge a complaint.[176] The ECHR strictly rejects an *actio popularis* in the individual application and requires a sufficiently concrete and present impairment to qualify as a victim.[177] Mere suppositions, general criticism of the law or purely theoretical concerns will not suffice.[178]Although Article 33 ECHR also provides for an action by the State in the general interest, it is still a tool for the protection of individual human rights.[179] For Article 10 ECHR to be affected there must be a person that is at least sufficiently likely to be affected in the future.[180] A purely abstract protection of AI-generated content, with the argument that it could theoretically contribute to discourse, is therefore not enough. The personal scope of protection of Article 10 §1 ECHR requires a right-holder, either on the side of the sender or on the side of the recipient.

---

174 *Handyside v. UK* (n 124) para.49.
175 *Autronic Ag v Switzerland* [1990] ECtHR 17112/16, 57630/17, 37339/19 [47].
176 European Court of Human Rights, 'Practical Guide on Admissibility Criteria' (2024) para.24 <https://www.echr.coe.int/documents/d/echr/admissibility_guide_eng> accessed 7 May 2025.
177 ibid 24.
178 *Senator Lines Gmbh v Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom (déc)* [2004] ECtHR [GC] 23765/94; *Aksu v Turkey* [2012] ECtHR [GC] 4149/04, 41029/04 [50].
179 Grabenwarter and Pabel (n 117) para.2.
180 *Klass and Others v Germany* [1978] ECtHR 5029/71 para.41; *Soering v the United Kingdom* [1989] ECtHR 14038/88 paras.90-91.

2. Determining the Rights Holders under Article 10 §1 ECHR

Having established that the scope of protection requires an identifiable rights holder, it will now be clarified who can be considered as such in the context of AI-generated content. This section shows that AI systems themselves are not protected, whereas users and recipients may fall within the scope under certain conditions.

a. AI Systems as Potential Rights Holders under Article 10 §1 ECHR

AI is not able to claim human rights *de lege lata*, as it does not fulfil the criteria of being either a physical or legal person as defined by the ECtHR case-law.[181] However, given the general reference of Article 10 §1 ECHR to *'everyone'* and the ECtHR's dynamic adaptation of its interpretation to technological and social developments[182], an extension of the personal scope of protection could be considered.

There is currently a controversial debate in international jurisprudence as to whether AI systems should be granted (basic) legal capacity as an '*electronic person*'.[183] Those who oppose the legal capacity of AI attempt to provide a philosophical and neuroscientific justification for the definition of *'life'* and the distinction between humans and AI.[184] Although the boundaries between humans and technology are becoming increasingly blurred (e.g. exoskeletons or bioelectronics)[185], legal capacity remains a legal construct.[186] In this context, the ECtHR has established case-law in relation to non-physical (legal) persons: Legal persons may be covered by the scope of protection of Article 10 ECHR, since the right to freedom of expression is by its nature applicable to them.[187] The protection is functionally justified: Legal persons can be vehicles of public communication and fulfil the purpose of this right.[188] The decisive factor is

---

[181] European Court of Human Rights (n 176) 9.

[182] *Times Newspapers Ltd (nos. 1 and 2) v United Kingdom* (n 122) para.27; *Magyar Kétfarkú Kutya Párt v. Hungary* (n 123) para.37.

[183] Foundationally, Lawrence B Solum, 'Legal Personhood for Artificial Intelligences' (1992) 70 North Carolina Law Review 1231; See for example: Simon Chesterman, 'ARTIFICIAL INTELLIGENCE AND THE LIMITS OF LEGAL PERSONALITY' (2020) 69 International & Comparative Law Quarterly 819; Joshua C Gellers, *Rights for Robots: Artificial Intelligence, Animal and Environmental Law (Edition 1)* (Routledge 2020).

[184] Paulius Čerka, Jurgita Grigienė and Gintarė Sirbikytė, 'Is It Possible to Grant Legal Personality to Artificial Intelligence Software Systems?' (2017) 33 Computer Law & Security Review 685, 691–693; Bartosz Brożek and Marek Jakubiec, 'On the Legal Responsibility of Autonomous Machines' (2017) 25 Artificial Intelligence and Law 293, 303.

[185] Peter Joosten, 'Human Enhancement Technologies: 7 Examples' (*Peter Joosten*, 22 May 2020) <https://www.peterjoosten.org/human-enhancement-technologies/> accessed 8 May 2025; Bert-Japp Koops, Mireille Hildebrandt and David-Oliver Jaquet-Chiffelle, 'Bridging the Accountability Gap: Rights for New Entities in the Information Society?' 526.

[186] Oster (n 150) 35.

[187] *Autronic Ag v. Switzerland* (n 175) para.47; *Schweizerische Radio- Und Fernsehgesellschaft Srg v Switzerland* [2012] ECtHR 34124/06 para.41; Lorenz Dopplinger, 'Legal Persons as Bearers of Rights Under the ECHR' (2021) 5 University of Vienna Law Review 1, 4.

[188] Dopplinger (n 187) 4.

that the expression serves as a manifestation of human autonomy. The fact that AI becomes more similar to humans therefore does not establish legal capacity. The question is, whether, as in the case of legal persons, there is a functional need for it to be recognised as a legal entity.[189] There is currently no functional need to provide AI systems with protection of human rights in the context of freedom of expression. Freedom of expression fulfils two central functions: the protection of individual self-fulfilment and the guarantee of open public discourse (see in detail p.23f.). The first dimension presupposes an autonomous subject who achieves self-fulfilment and autonomy through their expression. However, AI systems are devoid of personal autonomy, intention and consciousness.[190] Consequently, they are unable to articulate their own perspective in the context of human rights but generate content through statistical pattern processing.[191] In the context of public discourse, there is no compelling rationale for affording legal protection to AI systems themselves. Although contributions from AI can influence or structure this discourse (see p.19f.), their protection is ensured indirectly via the people involved, such as users, developers or recipients. It is therefore not necessary for AI to be a direct bearer of Article 10 §1 ECHR.

b. AI Developer

Whether an AI developer is protected as a holder of freedom of expression under Article 10 §1 ECHR with regard to content generated by its AI depends largely on whether this content is attributable to them. For the purposes of this analysis, *'AI developer'* refers to the person or entity that designs, trains or provides the generative AI system that enables the automated content creation. At first glance, a comparison with tort law or product liability law might support such an attribution: There, the developer of a product is liable for damage caused by a source of risk created by him (e.g., faulty programming or violation of testing obligations).[192] When transferred to AI-generated content under human rights law, it could be argued that participation in the communication process via the generated content must also be attributed to the developer, as it was the developer that created the technical basis. However, the divergent nature of the protection offered by these two categories challenges this analogy.

[189] Jan-Erik Schirmer, 'Rechtsfähige Roboter?' (2018) 71 JuristenZeitung 660, 663.
[190] Deyi Li, Wen He and Yike Guo, 'Why AI Still Doesn't Have Consciousness?' (2021) 6 CAAI Transactions on Intelligence Technology 175, 178.
[191] Georg Northoff and Steven S Gouveia, 'Does Artificial Intelligence Exhibit Basic Fundamental Subjectivity? A Neurophilosophical Argument' (2024) 23 Phenomenology and the Cognitive Sciences 1097, 1098, 1109–1114.
[192] See for AI liability: DIRECTIVE (EU) 2024/2853 on liability for defective products and repealing Council Directive 85/374/EEC 2024 Recital 32; Maria Lillà Montagnani, Marie-Claire Najjar and Antonio Davola, 'The EU Regulatory Approach(Es) to AI Liability, and Its Application to the Financial Services Market' (2024) 53 Computer Law & Security Review 105984, 12; Oster (n 150) 16–19.

Whilst the primary objective of tort law and product liability law is to attribute responsibility for risks[193], freedom of expression protects personal self-fulfilment (1), and public discourse (2).

(1) Usually, the developer does not know what specific content their system will generate, as this depends largely on the user's prompt. Depending on the input, a language model such as ChatGPT can generate a neutral reference work, an insulting statement or a political statement, without the producer having substantive prior influence on the specific content. Furthermore, the content does not necessarily reflect the beliefs or intentions of the developer. Microsoft, for example, explicitly disassociated itself from racist and sexist statements generated by its chatbot *'Tay'*, although these statements were technically generated by Tay, they clearly contradicted the company's communication intentions.[194]

(2) Neither does a functional contribution to public discourse justify a protective position for the developer. Passive freedom of information is not engaged because the developer does not act as a recipient but as a technical facilitator of communication. It is true that the active freedom of information under Article 10 §1 ECHR also protects the right to pass on information. However, according to the ECtHR, this presupposes a certain degree of influence over the content of the information: In *Delfi AS v. Estonia* (2015) the Court only established obligations for platform operators in relation to Article 10 §2 ECHR because the news platform published its own content and explicitly invited users to comment on it.[195] At the same time, the ECtHR emphasised that purely technical providers, such as discussion forums without editorial intervention, are not covered.[196] Developers of generative AI typically do not take this active role. They neither control the content nor transmit specific information but limit themselves to upstream technical facilitation. This structural distance from the act of communication argues against attribution in the sense of Article 10 §1 ECHR. This is confirmed by *Magyar Jeti Zrt v. Hungary* (2018): The mere linking to a third-party website does not give rise to responsibility under Article 10 §1 ECHR, *inter alia*, where the person concerned has no control over the

[193] See for example: DIRECTIVE (EU) 2024/2853 on liability for defective products and repealing Council Directive 85/374/EEC Recital 2.
[194] Peter Lee, 'Learning from Tay's Introduction' (*The Official Microsoft Blog*, 25 March 2016) <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/> accessed 8 May 2025; MJ Wolf, KW Miller and FS Grodzinsky, 'Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications' (2017) 1 The ORBIT Journal 1.
[195] *Delfi as v Estonia* [2015] ECtHR [GC] 64569/09 paras.113-117.
[196] ibid 117.

communication.[197] An attribution can only be made if the producer systematically evokes certain statements through targeted control or a very narrow selection of training data content and thus functionally assumes the controlling role of the user. Usually, however, there is no such link[198], which means that the scope of protection of Article 10 §1 ECHR does not apply.

c. User

Article 10 §1 ECHR protects indirect forms of communication, provided that it is possible to attribute the content: The ECtHR recognises that freedom of expression is not limited to traditional forms of authorship but can also encompass acts of publication of other content.[199] In the context of freedom of expression, this self-expression can be regarded as an articulation of individual self-fulfilment. The scope of protection presupposes that the expression can at least reflexively be attributed to the personality of the individual (see above, p.24f.). On the other hand, freedom of information also protects actions by which individuals participate in the dissemination of information.[200] Both dimensions therefore require the user to assume responsibility for the content, but not necessarily original authorship of the entire content. Under these conditions, the user falls within the personal scope of protection of Article 10 §1 ECHR (see p.27f.).

d. Recipient

As previously outlined in the context of the material scope of protection, Article 10 §1 ECHR also encompasses passive freedom of information. The personal scope of protection is engaged when a person specifically accesses, or attempts to access, AI-generated content in order to obtain information (see p.28f.), for example by asking a generative tool like *LawConnect* or *Tendi* for legal or financial advice.

---

197 *Magyar Jeti Zrt v Hungary* [2018] ECtHR 11257/16 [75].
198 Jessica Ji, Josh A Goldstein, and Andrew J. Lohn, 'Controlling Large Language Model Outputs: A Primer' (Center for Security and Emerging Technology (CSET), Georgetown University 2023) 12 <https://cset.georgetown.edu/publication/controlling-large-language-models-a-primer/> accessed 9 May 2025.
199 *Jersild v Denmark* [1994] ECtHR [GC] 57891/17 [35].
200 Grabenwarter and Pabel (n 117) para.8.

III.Concluding Remarks on Chapter 2

AI-generated content may, under certain conditions, fall within the scope of protection of Article 10 §1 ECHR.

Within the material scope of protection, the general freedom of expression requires a recognisable act of human will, for example through deliberate publication. Freedom of information allows for a functional approach: Content that contributes to the formation of public opinion can be protected regardless of its technical origin. Extremist content that deliberately undermines democratic structures can also be excluded from the scope of protection under Article 17 ECHR. However, this exception must be applied narrowly and cannot be a blanket application for cases where the information sphere is negatively affected.

The personal scope of protection of Article 10 §1 ECHR requires that a specific holder of the human right can be identified. AI systems do not fulfil this requirement, as they do not have the capacity to exercise human rights. Furthermore, the developer of AI is not generally protected, as the content that is generated usually cannot be attributed to them due to a lack of influence on the content. A user, on the other hand, is entitled to protection if they appropriate the content or disseminate it in a targeted manner. The same applies to the recipient who seeks access to information relevant to the discourse. The personal scope of protection is thus engaged when a person, whether as sender or recipient, is involved in a specific act of communication.

In conclusio, neither a blanket exclusion nor an unlimited inclusion of AI-generated content is compatible with Article 10 §1 ECHR. The scope of protection is only engaged where there is human participation in the form of an individual act of expression or a functional contribution to or reception of public discourse.

Chapter 3: Limitations on AI-Generated Content Under Article 10 §2 ECHR

The third chapter examines the legal challenges that arise from Article 10 §2 ECHR when justifying state interference with AI-generated content in response to the pollution of the information sphere. In accordance with Article 10 §2 ECHR, content falling within the scope of Article 10 §1 ECHR may only be restricted if the interference is prescribed by law, is pursued for a legitimate aim, and is deemed necessary in a democratic society.[201] This chapter argues that, although AI-generated content entails new forms of expression and risks for the information sphere, these phenomena can be adequately assessed within the existing legal framework of Article 10 ECHR. Although the protection of the information sphere is not explicitly mentioned, it can be subsumed under several legitimate aims of Article 10 §2 ECHR and can also be deemed necessary in a democratic society under certain conditions. By an analysis of two practical use cases, this chapter demonstrates that the existing principles can be applied as part of a case-by-case assessment, which takes the content, origin and effect of the AI-generated content into account.

This chapter is divided into two main parts. The first part provides an overview of the legal framework under Article 10 §2 ECHR in abstract terms, including case law of the ECtHR (I). The second part applies these standards to two case studies to illustrate how the requirements of Article 10 §2 ECHR can be applied to novel phenomena (II).

I. Legal Framework of Article 10 §2 ECHR

As a first step, this section briefly outlines the legal context and criteria under which state measures qualify as interference with freedom of expression (1). Next, the requirements arising from the *'prescribed by law'* principle are analysed (2). Following this, it is examined whether the protection of the information sphere can be recognised as a legitimate aim within Article 10 §2 ECHR (3). Last, it is analyses whether, and under what conditions, restrictions on AI-generated content can be classified as *'necessary in a democratic society'* (4). Rather than assessing specific measures, this section aims to outline the legal framework that defines the scope within which state interferences can occur. A detailed analysis of specific measures would require a case-by-case assessment and is therefore developed in the subsequent section through two illustrative case studies.

[201] *Handyside v. UK* (n 124) paras.48-49.

1. State Interference

The state can interfere with freedom of expression in various ways. As well as traditional sovereign measures, the concept of interference covers *de facto* or indirect state actions that hinder or discourage the exercise of communicative freedoms.[202] Case law distinguishes between *'prior restraint'*[203] and subsequent punishments, such as sanctions[204] or convictions.[205] Not only content bans but also restrictions on the form of presentation or dissemination are intrusive.[206] State measures against AI-generated content can be considered an interference, provided they fall within the scope of protection, as outlined in Chapter 2. In such cases, they must be justified in accordance with Article 10 § 2 ECHR.

2. Prescribed by Law

Restrictions on the freedom of expression protected by Article 10 §1 ECHR require a legal basis, Article 10 §2 ECHR. The law must be accessible to the person concerned; internal administrative rules or unpublished guidelines do not satisfy this criterion.[207] Furthermore, the ECtHR requires sufficient certainty and predictability: The person concerned must have the opportunity to adapt their behaviour to the regulation, if necessary with legal assistance.[208] Indeterminate legal terms remain permissible as long as they are practicable and can be interpreted depending on the context.[209] Excessive demands for legal precision, however, would impede the necessary flexibility of the law and prevent meaningful application in dynamic areas of life.[210]

3. Legitimate Aim: Protection of the Information Sphere

Any state actions aimed at restricting AI-generated content must serve one of the legitimate aims set out in Article 10 §2 ECHR. The ECtHR also requires the aim to be clearly defined and to concern a legally protected interest of sufficient importance, that is recognised as a *'pressing social need'* in a democratic society.[211] The protection of the information sphere is not explicitly mentioned in the ECHR. Therefore, it must be clarified whether it can be subsumed under one

[202] Cornils (n 134) para.34.
[203] *Ahmet Yildirim v Turkey* [2012] ECtHR 3111/10 [47].
[204] *Danileț v Romania* [2024] ECtHR 16915/21 [48].
[205] *Savva Terentyev v Russia* [2018] ECtHR 10692/09 [53].
[206] Cornils (n 134) para.34.
[207] *The Sunday Times v the United Kingdom (no 1)* [1979] ECtHR 6538/74 para.49; Cornils (n 134) para.45.
[208] *Sanoma Uitgevers B.v v the Netherlands* [2010] ECtHR [GC] 38224/03 [81]; *Rtbf v Belgium* [2011] ECtHR 50084/06 [103].
[209] *Rtbf v. Belgium* (n 208) para.103.
[210] ibid 103; *Nit S.r.l v the Republic of Moldova* [2022] ECtHR [GC] 28470/12 [159].
[211] *Handyside v. UK* (n 124) para.48.

of the exhaustively listed aims in Article 10 §2 ECHR.[212] The following analysis shows that protecting the information sphere can be subsumed under several legitimate aims, in particular the protection of the reputation and rights of others including non-discrimination, national security and public safety and the prevention of disorder, provided that there are concrete effects on individual rights, democratic structures or the physical safety of those affected. The subsequent sections examine each of these aims in turn.

a. Protection of the Reputation or Rights of Others

Article 10 §2 ECHR permits restrictions on freedom of expression and information, *inter alia*, to protect the *'reputation or rights of others'*. The protection of the rights of recipients under Article 10 §1 ECHR and of persons concerned under Article 8 ECHR and the right to non-discrimination are the most directly affected rights in the context of AI-generated content and information pollution and will therefore be analysed in the subsequent section.

a.a.Article 10 §1 ECHR

The recipient's passive freedom to receive and impart information without interference (see p.26f.) may serve as a legitimate aim, that justifies state measures which interfere with the freedom of expression of third parties, such as providers of AI-generated content.

To date, the ECtHR has not explicitly ruled on whether protecting the information sphere itself falls within the scope of freedom of information under Article 10 §1 ECHR and thereby constitutes a legitimate aim under Article 10 §2 ECHR. Any further interpretation must be grounded in the purpose of Article 10 ECHR, which aims *inter alia* at participation in public discourse.[213] The fact that pollution of the information sphere is caused by a large number of private actors and dynamic technological processes argues against a broad interpretation. Interferences of individual rights are difficult to define concretely and could exceed the scope of protection previously recognised under Article 10 §1 ECHR, which primarily addresses concrete access to information.[214]

Nevertheless, several arguments support a broad interpretation. Firstly, Article 10 §1 ECHR also protects the efforts to access information, state authorities may not impair them

[212] Christoph Grabenwarter and Katharina Pabel, '§ 18. Struktur Der Grundrechtsprüfung' in Christoph Grabenwarter and Pabel Pabel (eds), *Europäische Menschenrechtskonvention* (7th edn, 2021) para.12.
[213] *Baka v Hungary* [2016] ECtHR [GC] 20261/12 [167].
[214] *Guerra and Others v Italy* [1998] ECtHR [GC] 58135/09 [52–53]; *Weber v Germany (dec)* [2015] ECtHR 70287/11 [25].

disproportionately.[215] This means that the scope of protection can include not only access to information itself, but also the conditions that precede it, provided they are a prerequisite for receiving information. The right to participate presupposes the existence of a functioning public discourse; without it, the protection provided by human rights would be meaningless.[216]

Secondly, the ECtHR has recognised that access to generally available information is a right that can also be protected against private third parties.[217] In principle, recipients can therefore be affected in their passive freedom of information by the action of third parties, such as the generation of AI-generated content. Threats to public discourse can come not only from the state, but also from private third parties, for example through the monopolisation of media power[218] or the structural destruction of discursive spaces due to flooding with irrelevant (quantitative infollution) or false content (qualitative infollution).

Thirdly, the state has a duty to organise its information system in a manner that enables informed participation in key issues; to fulfil this role, it must ensure that the information in its sphere meets standards of veracity, sincerity, and completeness, which implies that the information must be truthful, not misleading and sufficiently comprehensive to enable informed public debate.[219] While Article 10 §1 ECHR does not impose an obligation to provide information in a specific form or create new sources of information[220], it is recognised that the effective organization of access to information is crucial at least within the state's sphere of influence (formal infollution).[221] Therefore, a functioning, accessible and reliable information sphere is an integral part of the effective protection of human rights.[222] If private third parties' actions are capable of impairing freedom of information, it stands to reason that the structural nature of the public information sphere should be recognised as an asset worthy of protection. It is therefore coherent to extend the scope of protection to the private sphere of information, as the effectiveness of freedom of information depends on not only state conduct but also on the structural integrity of the information sphere, which includes privately operated infrastructures.

---

[215] Grabenwarter and Pabel (n 117) para.9.

[216] *Association Burestop 55 Et Autres c France* [2021] ECtHR 56176/18, 56189/18, 56232/18, 56236/18, 56241/18, 56247/18 [108]; Katie Pentney, 'The Right of Access to "Reliable" Information Under Article 10 Echr: From Meagre Beginnings to New Frontiers' (2024) 5 European Convention on Human Rights Law Review, The 230, 248–249.

[217] *Leander v Sweden* [1987] ECtHR 9248/81 [74]; *Khurshid Mustafa and Tarzibachi v Sweden* [2008] ECtHR 23883/06 [41]; *Gillberg v Sweden* [2010] ECtHR 41723/06 [82].

[218] *Informationsverein Lentia and Others v Austria* [1993] ECtHR 24753/13, 28610/13, 57854/13, 65553/13 38–39.

[219] *Association Burestop 55 Et Autres c. France* (n 216) para.108; Pentney (n 216) 248.

[220] *Guerra and Others v. Italy* (n 214) para.53; *Weber v. Germany (dec.)* (n 214) para.25.

[221] Pentney (n 216) 248–255.

[222] *Association Burestop 55 Et Autres c. France* (n 216) para.108; Pentney (n 216) 248–249.

Given the growing prevalence of AI-generated content, automated disinformation campaigns, and algorithmic distortions, the pollution of the information sphere has escalated to the point where it no longer merely involves selective misinformation but rather undermines the fundamental prerequisites of public discourse.[223] For the first time, there is a serious risk that passive freedom of information will be devalued systematically. Therefore, the protection of the information sphere is to be regarded as an integral part of passive freedom of information and can constitute a legitimate aim within the meaning of Article 10 §2 ECHR, in order to take targeted action against such interference.

a.b.Article 8 §1 ECHR

Article 8 §1 ECHR can constitute a legitimate aim to justify an interference with Article 10 §1 ECHR insofar as it concerns the protection of the right to private or family life.[224] The ECtHR has repeatedly emphasised that this includes the protection of honour and privacy.[225] The fact that AI-generated content can lead to an interference is illustrated by the following case: In the USA, it was revealed that an AI-supported tool had falsely accused conservative activist *Robby Starbuck* of taking part in the storming of the US Capitol on 6th January 2021. *Starbuck* initiated a defamation lawsuit against the platform operators due to the dissemination of this untrue, reputation-damaging claim.[226] This example shows how AI-generated content can massively damage the reputation and privacy of individuals. Such content constitutes qualitative pollution of the information sphere and can justify restrictions on freedom of expression under Article 10 §2 ECHR, in order to safeguard Article 8 ECHR.

a.c.Non-Discrimination

Protecting the rights of others can also entail the principle of non-discrimination. While Article 10 ECHR also protects offending, disturbing or shocking statements[227], *'attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating racist speech in the face of*

---

[223] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Defence of Democracy' (2023) COM(2023) 630 final s 3.1 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52023DC0630> accessed 13 May 2025; Amaka Onebunne, 'Algorithmic Bias and Media Manipulation: A Systematic Review of AI's Role in Shaping Public Perception and Political Discourse' (2022) 16 World Journal of Advanced Research and Reviews 1239, 1241, 1245, 1246.

[224] *Axel Springer Ag v Germany* [2012] ECtHR [GC] 39954/08 82–83.

[225] *Chauvy and Others v France* [2004] ECtHR 64915/01 para.70; *Pfeifer v Austria* [2007] ECtHR 12556/03 para.35; *Axel Springer Ag v. Germany* (n 224) para.77.

[226] 'Conservative Activist Robby Starbuck Sues Meta over AI Responses about Him' (*AP News*, 30 April 2025) <https://apnews.com/article/robby-starbuck-meta-ai-delaware-eb587d274fdc18681c51108ade54b095> accessed 13 May 2025.

[227] *Handyside v. UK* (n 124) para.49.

freedom of expression.'[228] Therefore, state measures against group-based disparagement or hate speech can serve the legitimate aim of *'protecting the rights of others'*.[229] When AI systems generate discriminatory statements and thereby pollute the information sphere qualitatively, this can justify proportional interferences.

b. National Security and Prevention of Disorder

Pursuant to Article 10 §2 ECHR, *'national security'* and the *'prevention of disorder'* can constitute legitimate aims to justify an interference. The primary purpose of protecting *'national security'* is to defend against espionage, the disclosure of secrets, and other threats to the functioning of state institutions and partly overlaps with the objective of maintaining state confidentiality.[230] The ECtHR interprets the term *'prevention of disorder'* narrowly.[231] It only covers significant disruptions to external order, such as civil unrest, riots, and specific threats to public peace.[232] There are two reasons why the protection of the information sphere should be included under these legitimate aims: Firstly, it is recognised that threats to national security are no longer confined to the physical realm.[233] Cyberattacks, digital influence campaigns and coordinated disinformation can jeopardise state stability, undermine trust in democratic institutions and disrupt collective decision-making processes.[234] The ECtHR takes a dynamic approach to interpretation[235] and has already considered digital phenomena in other contexts.[236] Against this background, it seems consistent to also include the protection of the digital information sphere as a part of the legitimate aims *'national security'* and *'prevention of disorder'*.

Secondly, even though the protection of democratic values is not explicitly stated as a legitimate aim in Article 10 §2 ECHR, the ECtHR has recognised that protecting democracy is part of *'national security'* and *'prevention of disorder'*.[237] Freedom of expression and free public discourse are considered essential elements of a democratic society.[238] Damaging or

[228] *Vejdeland and Others v Sweden* [2012] ECtHR 1813/07 para.55.
[229] *Feret c Belgique* [2009] ECtHR 15615/07 paras.59,64.
[230] Cornils (n 134) para.52.
[231] *Perinçek v. Switzerland* (n 171) paras.146-154.
[232] ibid paras.146,152; Cornils (n 134) para.53.
[233] Kathan Patel and Dhaval Chudasama, 'National Security Threats in Cyberspace' (2021) 4 National Journal of Cyber Security Law 12, 19.
[234] Masabah Bint E Islam and others, 'AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework' (2024) 2 Blockchains 458, 466–467; Patel and Chudasama (n 233) 15; Spencer McKay and Chris Tenove, 'Disinformation as a Threat to Deliberative Democracy' (2021) 74 Political Research Quarterly 703, 708.
[235] *Tyrer v. the United Kingdom* (n 132) para.31.
[236] *Ahmet Yildirim v. Turkey* (n 203) para.54; *Delfi as v. Estonia* (n 195) para.65.
[237] *Refah Partisi (the Welfare Party) and Others v. Turkey* (n 170) paras.65-67.
[238] *Handyside v. UK* (n 124) para.49.

endangering public discourse through information pollution can therefore be seen as a threat to the democratic process as a whole. For example, shortly before election day in 2024 Slovakian elections, AI-generated deepfake videos that showed the candidates making controversial statements were distributed online to influence voter behaviour and undermine trust in the electoral process.[239] Such disinformation (qualitative infollution) can destabilise democratic institutions and pose a threat to national security. As a result, protecting the integrity of the information sphere falls within the legitimate aims of *'national security'* and *'prevention of disorder'* under Article 10 §2 ECHR.

c. Public Safety

Another legitimate aim for justifying interferences with freedom of expression is the protection of '*public safety'*, Article 10 §2 ECHR. The ECtHR interprets this ground narrowly.[240] It recognises situations in connection with the fight against terrorism or acts that may lead to increased violence[241], as well as measures to contain tense situations at risk of escalating violence.[242] Public safety may also be affected in the context of a polluted information sphere, especially in the case of qualitative information pollution, such as AI-generated targeted disinformation, hate messages or inflammatory content that creates an immediate threat to certain individuals or groups.

d. Interim Conclusion

Despite not being explicitly mentioned, the analysis shows that the protection of the information sphere can be subsumed under several legitimate aims of Article 10 §2 ECHR. This especially applies to protecting the reputation and rights of others, public safety, national security, and the prevention of disorder. In each case, the prerequisite is that the polluted information sphere has a concrete impact on the reputation or rights of individuals, democratic structures, or the physical safety of people. The extent to which information pollution poses a threat to these interests must be evaluated empirically and on a case-by-case basis.

---

[239] 'Election Disinformation in Slovakia – Disinfo Bulletin 9 April' (*EDMO European Digital Media Observatory*) <https://ec.europa.eu/newsroom/edmo/newsletter-archives/52231> accessed 24 June 2025.
[240] *Stoll v. Switzerland* (n 140) para.54.
[241] *Karatas v Turkey* [1999] ECtHR [GC] 56079/21, 57743/21, 58274/21 [51]; *Gözel Et Özer c Turquie* [2010] ECtHR 43453/04, 31098/05 [55]; *Leroy c France* [2008] ECtHR 36109/03 [36].
[242] *Pentikäinen v Finland* [2015] ECtHR [GC] 11882/10 paras.75-76.

4. Necessary in a Democratic Society

Whether restrictions on AI-generated content are compatible with Article 10 §2 ECHR depends on whether they are *'necessary in a democratic society'*. This is the central criterion for the proportionality test, in which a context-specific balancing of conflicting legal positions takes place. While the Court recognises a certain margin of appreciation on the part of the contracting states, it reserves the right to independently review the necessity.[243]

The analysis in this section follows four steps: Firstly, the legal requirements for necessity are presented (a). Secondly, the protective weight of the legitimate aims involved is assessed systematically (b). Thirdly, the worthiness of protection of freedom of expression is examined (c). Fourth, the need for a case-by-case assessment for the balancing is outlined (d). The analysis demonstrates that an interference with freedom of expression can be compatible with Article 10 §2 ECHR if they address a specific risk to the legitimate aims, are based on a case-by-case assessment and are proportionate in response to the nature, origin and democratic relevance of the content. The following considerations are intended to outline relevant criteria and areas of conflict as a structured preliminary examination that provides points of reference for further research.

a. The Necessity Standard under Article 10 §2 ECHR

The phrase *'necessary in a democratic society'* in Article 10 §2 ECHR refers to the general principle of proportionality, which encompasses suitability, necessity and proportionality.[244] Although the ECtHR has developed these criteria, in practice it regularly examines them as part of an integrated overall assessment, without making a strict distinction.[245] The ECtHR does not view its role as a substitute for national decision-making; rather, it sees itself as having a *'supervisory function'* to monitor compliance with the limits of Convention law.[246] The intensity of this control depends on the margin of appreciation of the states concerned, which in turn depends on the severity of the interference and the social importance of the expression in question.[247] The decisive factor is whether the state measure corresponds to a *'pressing social need'*[248], which applies to all types of expression, but is stricter where the expression is

---

[243] Cornils (n 134) para.58.
[244] ibid 54; Eirik Bjorge, 'Proportionality' in Eirik Bjorge (ed), *Domestic Application of the ECHR: Courts as Faithful Trustees* (Oxford University Press 2015) 155, 161.
[245] See e.g.: *Schweizerische Radio- Und Fernsehgesellschaft Srg v. Switzerland* (n 187) paras.56ff.
[246] *Nagla v Latvia* [2013] ECtHR 73469/10 [93]; *Halet v Luxembourg* [2023] ECtHR [GC] 21884/18 [177].
[247] *Animal Defenders International v the United Kingdom* [2013] ECtHR [GC] 48876/08 para.102; *Ashirov and International Memorial v Russia* [2023] ECtHR 25246/07 para.13; Cornils (n 134) para.58.
[248] *Handyside v. UK* (n 124) paras.48-49; Cornils (n 134) para.54.

significant for forming public opinion (for example, in the case of political speech[249]).[250] Conversely, a wider scope is recognised for commercially motivated or purely private communication.[251]

b. Protective Weight of the Legitimate Aims

In order to determine whether, and under what conditions, a state can interference with AI-generated content under Article 10 §2 ECHR, the weight of the legitimate aims must be assessed in light of the potential threats posed by synthetic content.

b.a.Protection of the Reputation and Rights of Others

Passive freedom of information safeguards every individual's right to access information and ideas without state interference (see p.26f.). This right is vital to democracy and individual self-fulfilment, as it ensures that citizens receive the information they need to exercise their democratic rights competently and to make informed political decisions.[252] It also supports informed decisions about a wide range of areas of life and thereby strengthens the ability to organise life freely and responsibly.[253]

As explained above, the exercise of this right depends on a functioning, accessible and reliable information sphere. If this structure is impaired, the effective exercise of passive freedom of information is also affected (see pp.39f.). Against this background, the extent to which the increasing dissemination of AI-generated content impairs the exercise of this right and the weight of such an impairment must be analysed. A comprehensive evaluation of all possible scenarios is beyond the scope of this thesis. The ECtHR would have to examine on a case-by-case basis to what extent the conditions for justifying an interference are met.

The right to respect for private life and the protection of reputation safeguards social identity in the digital and social sphere.[254] Article 8 §1 ECHR is exceptionally important, because it protects personal dignity by safeguarding individuals from unwanted public exposure and by

---

[249] *Wingrove v the United Kingdom* [1996] ECtHR 58080/11 [58]; *Ashby Donald Et Autres c France* [2013] ECtHR 36769/08 [39].
[250] Cornils (n 134) para.57.
[251] ibid 57; Dirk Voorhoof, 'Restrictions on Television Advertising and Article 10 of the European Convention on Human Rights.' [1993] INTERNATIONAL JOURNAL OF ADVERTISING 189, 206.
[252] Candace Cummins Gauthier, 'Right to Know, Press Freedom, Public Discourse' (1999) 14 Journal of Mass Media Ethics 197, 198; Rubén Marciel, 'On Citizens' Right to Information: Justification and Analysis of the Democratic Right to Be Well Informed' (2023) 31 Journal of Political Philosophy 358, 361.
[253] Marciel (n 252) 369.
[254] *S and Marper v the United Kingdom* [2008] ECtHR [GC] 30562/04, 30566/04 [66]; Bettina Schöndorf-Haubold, *Das Recht auf Achtung des Privatlebens: Problemstellungen im Digitalbereich: eine rechtsvergleichende Perspektive - Deutschland* (Generaldirektion Wissenschaftlicher Dienst, Europäisches Parlament 2018) 3.

ensuring they have control over their social image and how their personal information is disseminated.[255]

AI-generated content affects the exercise and protection of Article 8 §1 ECHR *inter alia* through the inaccuracy of synthetic content: Models such as LLMs can easily generate false, outdated or distorted facts about real people (see p.13f.). LLMs have also been shown to reproduce personal data from their training sets and thereby violate the right to privacy under Article 8 §1 ECHR.[256] This can lead to invasions of privacy and their reputation, for example when individuals are misquoted, linked to fictitious statements, or portrayed in compromising contexts (see the example of Robby Starbuck, p.41). In this context, the potential for mass reputational damage is problematic, as the public perception of a person can be distorted in a short period of time, regardless of any intention to manipulate.[257] Once generated and disseminated, false information or defamatory content can be included in subsequent training data.[258] This perpetuates the original violation, because once disseminated, misrepresentations can be repeated in future model generations, and cause long-term reputational damage.[259]

b.b.National Security and Public Order

The ECHR is considered to be an instrument for protecting democratic societies, which require open democratic discourse.[260] If the information sphere is impaired, the central prerequisites of democratic systems are jeopardised, especially public control, pluralistic debate and democratic decision-making processes.[261] AI-generated content can impair these democracy-relevant functions, for example by disseminating systematic disinformation that undermines trust in

---

[255] James Q Whitman, 'The Two Western Cultures of Privacy: Dignity versus Liberty' (2004) 113 The Yale Law Journal 1151, 1161.

[256] Xabier Lareo, 'Large Language Models (LLM) | European Data Protection Supervisor' (13 June 2025) <https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm> accessed 24 June 2025.

[257] Andi Terziu, 'AI – the Threats It Poses to Reputation, Privacy and Cyber Security, and Some Practical Solutions to Combating Those Threats' (*Taylor Wessing*, 18 April 2024) <https://www.taylorwessing.com/en/global-data-hub/2024/cyber-security---weathering-the-cyber-storms/ai---the-threats-it-poses-to-reputation> accessed 16 May 2025.

[258] Bill Tomlinson, Donald J Patterson and Andrew W Torrance, 'Turning Fake Data into Fake News: The AI Training Set as a Trojan Horse of Misinformation' (2023) 60 San Diego Law Review 650.

[259] Ilia Shumailov and others, 'The Curse of Recursion: Training on Generated Data Makes Models Forget' [2023] arXiv 1, 3.

[260] *Handyside v. UK* (n 124) para.49; Steven Greer, *The European Convention on Human Rights: Achievements, Problems and Prospects* (Cambridge University Press 2006) 195–199.

[261] *Informationsverein Lentia and Others v. Austria* (n 218) para.38; Niklas Gudowsky and Ulrike Bechtold, 'The Role of Information in Public Participation' (2013) 9 Journal of Deliberative Democracy 22–23; Francesco Pira, 'Disinformation a Problem for Democracy: Profiling and Risks of Consensus Manipulation' (2023) 8 Frontiers in Sociology 1, 7; See in context of social media: Joshua A Tucker and others, 'Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature' (Social Science Research Network (SSRN) 2018) 4–5, 52–53.

institutions and elections.[262] In addition to misleading or manipulative statements, the structural effects of mass information production are increasingly destabilising public discourse (see p.19f.). The AI-caused information overload makes it difficult for state actors to accurately gauge public opinion and for citizens to critically monitor the actions of representatives, and thereby undermines the prerequisites democratic representation and accountability.[263] Moreover, the wide-ranging misuse of AI could intensify social tensions, escalate political conflicts, and provoke public unrest, which are developments that are likely to jeopardise the stability of a democratic community.[264] However, it should be noted that the ECtHR interprets *'national security'* rather restrictive.[265]

In addition to national security, public order can be threatened deeply by AI-generated content. A robust information sphere promotes public order by facilitating rational and critical discussions about shared concerns and thereby contributes to inclusive social coexistence.[266] However, there is growing concern that the spread of emotionally charged misinformation may hinder people's ability to form informed opinions and contribute to social tensions, especially in information environments, where access to reliable information is limited or distorted by the reinforcement of polarisation patterns and portraying certain groups in a distorted way.[267]

b.c.Public Safety

'Public safety' is a fundamental aspect of democratic societies. It protects individual legal interests, such as life, physical integrity and property, and guarantees the conditions for a peaceful society.[268] The large-scale dissemination of disinformation, provocative content or incitement to violence, assisted by automated means, may increase social tensions and give rise to risks to public safety.[269] Nevertheless, the ECtHR interprets *'public safety'* narrowly.[270] Justifying interferences on expression under this aim requires a case-specific assessment supported by concrete and reliable evidence.

---

[262] Wilson Wong, 'Trends in Political Science Research: Artificial Intelligence and Voter Disinformation' (2025) 75 International Political Science Abstracts 205, 210.
[263] Sarah Kreps and Doug Kriner, 'How AI Threatens Democracy' (2023) 34 Journal of Democracy 122, 123.
[264] Danhuai Guo and others, 'AIGC Challenges and Opportunities Related to Public Safety: A Case Study of ChatGPT' (2023) 4 Journal of Safety Science and Resilience 329, 333.
[265] *Stoll v. Switzerland* (n 140) para.54; European Court of Human Rights, 'Guide on Article 10 of the European Convention on Human Rights – Freedom of Expression' (Council of Europe 2022) paras.543-544.
[266] C Calhoun, 'The Public Sphere in the Field of Power' (2010) 34 Social Science History 301, 302.
[267] Andrew M Guess and Benjamin A Lyons, 'Misinformation, Disinformation, and Online Propaganda' in Joshua A Tucker and Nathaniel Persily (eds), *Social Media and Democracy* (Cambridge University Press 2020) 27.
[268] *S.a.s v France* [2014] ECtHR [GC] 43835/11 para.115.
[269] Guo and others (n 264) 333.
[270] *Stoll v. Switzerland* (n 140) para.54; European Court of Human Rights (n 265) paras.543-544.

c. Criteria for Justifying Restrictions on Freedom of Expression

Freedom of expression and information is a cornerstones of a democratic society.[271] Its individual and democratic-functional importance means, that the burden of justification is significant.[272] The ECtHR requires any restriction to be based on a *'pressing social need'*.[273] Nevertheless, freedom of expression does not take precedence over other legitimate legal interests *a priori*.[274] Rather, each case requires a balance of interest, in which freedom of expression may be overridden if the opposing right is deemed more important.[275] The type and severity of the interference are decisive factors here.[276] The ECtHR has developed extensive case law in this regard.[277] The following section presents the relevant factors when assessing the proportionality of restrictions aimed at protecting the integrity of the information sphere.

c.a.Factual Statements vs. Value Judgements

The ECtHR distinguishes between factual statements and value judgements.[278] Value judgements enjoy broad protection from state interference as they are an expression of individual freedom of opinion and personal self-fulfilment.[279] However, this categorisation is problematic in the case of AI-generated content for two reasons. Firstly, it is often unclear whether the statement can be attributed to a natural person.[280] Secondly, the AI itself lacks subjectivity (see p. 32f.). Consequently, the content is either a factual claim, which is subject to stricter examination of their factual basis,[281] or value judgements that lack a personal perspective. As previously mentioned, purely AI-generated content does not open the dimension of self-fulfilment on which the intensive protective effect of Article 10 ECHR is based (see p.27f.). These therefore do not fall under the increased protection. This may be assessed differently if human input (e.g. through targeted control, editing, or creative selection) has been incorporated into the content.

[271] *Handyside v. UK* (n 124) para.49.
[272] Cornils (n 134) para.54.
[273] *Barthold v Germany* [1985] ECtHR 27378/16, 30348/16, 59267/16, 52237/17, 57432/17, 62904/17 [55].
[274] Cornils (n 134) para.54.
[275] European Court of Human Rights (n 265) para.108.
[276] ibid 265.
[277] ibid paras.94-105.
[278] *Lingens v Austria* [1986] ECtHR 9815/82 [46].
[279] Cornils (n 134) para.65.
[280] Brian Porter and Edouard Machery, 'AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably' (2024) 14 Scientific Reports 26133, 8.
[281] *Mcvicar v the United Kingdom* [2002] ECtHR 60450/13 para.84; Cornils (n 134) para.63.

c.b.True vs. Untrue Statements

False statements are also protected by Article 10 §1 ECHR, but restrictions apply in cases of objectively inaccurate factual claims that are not in the public interest.[282] AI-generated content often pollutes the information sphere qualitatively, for example by generating disinformation systematically. Even if the creators are not classically culpable due to a lack of human intent, this does not diminish the justification for intervention, as the effect on public discourse is important. Content that contributes to qualitative information pollution therefore only enjoys a lower level of protection.

c.c.Relevance of the Content for the Public Debate

According to ECtHR case law, statements that contribute to a debate of public interest enjoy increased protection.[283] Therefore, AI-generated content benefits from this if it factually contributes to public debate. However, when such content is used mainly to maximise reach as described in Chapter 1 in the context of quantitative information pollution (for example when TikTok promotes videos based on user activity) or when it serves targeted disinformation like deepfakes as discussed under qualitative information pollution, the protection is lower.

c.d. ‚Chilling Effect‘

The ECtHR regularly considers the possibility of a *‘chilling effect’* when examining interference with freedom of expression.[284] Even measures that involve mild sanctions may be deemed disproportionate if they are likely to have such an effect on expression.[285] In the context of generative AI, however, there is no such effect on the machine itself, since it is not a holder of human rights. Nevertheless, it must be examined whether restrictions or sanctions have a *de facto* deterrent effect on people who use AI as a means of expression.

c.e.Human Involvement

From the perspective of freedom of expression of the speakers, the extent to which self-fulfilment is affected must be considered. The greater the degree of human involvement, the more severe the interference.

[282] Grabenwarter and Pabel (n 117) para.52.
[283] *Stoll v. Switzerland* (n 140) para.106.
[284] European Court of Human Rights (n 265) paras.57-58.
[285] *Brasilier c France* [2006] ECtHR 71343/01 para.43.

d. Balance

It is not possible to assess the proportionality of state measures against AI-generated content based on an abstract balance of the legal positions affected. Instead, the above-described factors must be weighed up against each other in a case-by-case assessment. This requires specific empirical findings on the concrete impact of AI-generated content pollution on the relevant protected goods. Generalised assessments and general solutions are inadequate. Therefore, the following section provides two use cases to illustrate how the dogmatic standards are applied to hypothetical state interferences.

e. Interim Conclusion

In conclusio, the analysis shows that a graduated approach is required that differentiates between the type, origin and impact of the content. Content containing more human involvement and personal expression enjoys greater protection, whereas disinformation enjoys less protection and at the same time poses risks to rights and democratic processes. Simultaneously, recipients' interest in information must be protected. Socially relevant content deserves special protection.

5. Conclusion

This section has demonstrated that restrictions on AI-generated content can be in line with Article 10 §2 ECHR. An interference must have a clear legal basis, pursue a legitimate aim and be necessary to fulfil a *'pressing social need'* in a democratic society.

Although Article 10 §2 ECHR does not explicitly mention the integrity of the information sphere, it can be subsumed under several listed aims. These include the protection of the rights of others, public safety, national security, and the prevention of public disorder. The prerequisite is always that the information pollution has a concrete impact on the individual legal positions.

Whether or not an intervention is necessary in a democratic society, depends on the type of content, and the nature, origin and effect of the pollution of the information sphere. Content that undermines democratic processes, such as low-quality or disinformation, enjoys only a low level of protection. In contrast, content that is relevant to the public discourse is worthy of a higher level of protection. A generalised assessment is not a viable option; rather, a context-related, case-by-case assessment is required.

II.Application of Article 10 §2 ECHR in Use Cases

The second part of the chapter considers how the standards set out in the previous section can be applied to concrete state measures against AI-generated content. The aim is to demonstrate the conditions under which such interventions can be justified in order to counteract the pollution of the information sphere. The analysis is based on two use cases which address the different forms of information pollution. The first is the state prohibition of deceptive political deepfakes; the second is the obligation of intermediaries to label and filter AI-generated content. The examples illustrate, that the existing legal framework from the previous section can be applied coherently to new phenomena like AI-generated content and the pollution of the information sphere.

1. Use Case for Qualitative Infollution: State Responses to Political Deepfakes

The first use case focuses on qualitative information pollution caused by politically motivated deepfakes. Taking the 2023 Slovakian elections as an example, this case study analyses the conditions under which European regulatory measures can be justified under Article 10 §2 ECHR. The analysis follows the structure of legal review as developed in the previous section. It concludes that, in the light of the minimal human input and the severe distortion of public discourse, the legitimate aims to protect national security, prevent disorder and protect the rights of others outweighs the expressive interest of the disseminator.

a. Factual Background

A notable case of deepfake disinformation occurred in the lead-up to the 2023 Slovakian elections: Two days before the election, during the legally prescribed 48-hour campaign break, a manipulated audio recording was circulated on social media. In the recording, the lead candidate of the Progressive Slovakia party, Michal Šimečka, and journalist Monika Tódová could allegedly be heard planning to buy votes from members of the Roma minority.[286] The fact that only one audio track was used, and that the publication took place shortly before the election, made it difficult to rectify the situation.[287] Given the close result that the SMER party won by a margin of 5%, it is impossible to determine the exact influence of the deepfake.[288] However, the content of the recording appeared to be intended to demobilise Šimečka's

[286] Tommaso Canetta, 'Prebunking AI-Generated Disinformation Ahead of EU Elections – EDMO' <https://edmo.eu/publications/prebunking-ai-generated-disinformation-ahead-of-eu-elections/> accessed 24 June 2025; Mateusz Łabuz and Christopher Nehring, 'On the Way to Deep Fake Democracy? Deep Fakes in Election Campaigns in 2023' (2024) 23 European Political Science 454, 464–465.
[287] Łabuz and Nehring (n 286) 465.
[288] De Nadal and Jančárik (n 44) 3,5.

electorate.[289] In a polarised, time-critical electoral environment, this form of qualitative information pollution could potentially have a decisive effect on the election.

b. State Interference

There is no specific regulation in Slovakia that explicitly prohibits political deepfakes. Influence could only be exerted indirectly, for example through general criminal law, such as Section 373 of the Criminal Code (Trestný zákon 300/2005), which penalises the dissemination of defamatory false allegations. This analysis therefore focuses on European instruments, namely the Digital Services Act (DSA) and the AI Act. Even though the EU is not a party to the ECHR, the rights of the ECHR are part of the general principles of EU law and guide the interpretation, Article 6 §3 TEU.

The DSA requires very large online platforms and search engines to identify and mitigate systemic risks, such as deepfake-based election interference, this kind of content must be clearly labelled (Articles 34(1)(c), 35(1)(k)). Transparency obligations exist with regard to users and authorities. While the AI Act does not impose a general ban on deepfakes, it does establish transparency obligations in Article 50(2)), and Article 50(4). These obligations aim to enable users to make an informed assessment (Article 50(5)), but do not apply to natural persons for purely private use (Article 2(10)). Lower requirements exist for artistic and law enforcement purposes (Article 50(4)). Although deepfakes are not generally considered to be high-risk systems, systems designed to influence elections may fall under Annex III (8)(b). In this case, the strict requirements of Chapter III (risk management, transparency, quality assurance and monitoring) apply.

c. Prescribed by Law

The provisions of the DSA and the AI Act fulfil the legal basis requirements set out in Article 10 §2 ECHR. Both regulations are publicly accessible, sufficiently specific and foreseeable. They specify clear addressees, obligations, and sanction mechanisms, and they effectively limit state discretion. Those affected can therefore adapt their behaviour accordingly.

d. Legitimate Aim

In the context of political deepfakes, several legitimate aims within the meaning of Article 10 §2 ECHR can be used to justify interference with freedom of expression. Firstly, the protection of the rights of others is relevant. This includes the passive freedom of information

[289] Łabuz and Nehring (n 286) 465.

of recipients, Article 10 §1 ECHR. Deepfakes that are created with the intention of deceiving people can undermine trust in public discourse and thus impair the ability of recipients to rely on truthful information. Therefore, protecting the information sphere from qualitative information pollution is a prerequisite for the effective exercise of freedom of information. In the Šimečka case, a violation of Article 8 ECHR is also likely. The deceptively authentic-sounding deepfake audio recording accused him of vote-buying, thus damaged his public reputation in the process. This constitutes an intrusion into the private life even for a public figure, in which case the ECtHR acknowledges that public figures must have a higher level of resilience.[290] As well as protecting individual rights, maintaining the democratic order can also be a legitimate aim. The dissemination of manipulated content immediately before elections, as occurred in Slovakia, can undermine confidence in the fairness and transparency of the electoral process.[291] Protection against such influence falls under the categories of *'national security'* and *'prevention of disorder'*.

e. Necessary in a Democratic Society

When examining the necessity of interference in a democratic society, the importance of the legitimate aim must be balanced with the impact on freedom of expression.

The protection of passive freedom of information, as set out in Art. 10 §1 ECHR protects not only access to specific information but also requires a functioning and reliable information sphere, which is the foundation of democratic decision-making and a prerequisite for the democratic order itself.[292] If this structure is undermined by qualitatively flawed AI-generated content, democratic discourse risks losing its ability to guide informed public debate. This is particularly serious when deepfakes are deliberately employed to erode trust in democratic processes, manipulate voter behaviour, and make the distinction between truth and falsehood impossible.[293] In such cases, the state's interest in protection carries substantial weight, as it is not only individual rights that are jeopardised, but the democratic system itself.

On the other hand, there is the scope of the protection provided by Article 10 ECHR for individuals creating or disseminating deepfakes. While in principle, restraint is required when taking measures against political statements, since they are vital for democratic debate and are

---

[290] *Von Hannover v Germany (no 2)* [2012] ECtHR [GC] 40660/08, 60641/08 para.110.
[291] Łabuz and Nehring (n 286) 465.
[292] *Association Burestop 55 Et Autres c. France* (n 216) para.108; Pentney (n 216) 248–249.
[293] Łabuz and Nehring (n 286) 468.

thus afforded special protection., this protection is reduced in the case of demonstrably false or deceptive statements involving a high degree of manipulation. Article 10 ECHR does protect incorrect statements of fact in principle. However, objectively false statements which do not contribute to public debate are afforded less protection.[294] This is evident in the case of manipulative disinformation as seen in the present case: The deepfake falsely suggested that a leading candidate had planned election fraud, which is an allegation with a clear factual but incorrect basis. Moreover, the level of human input is low in the case of deepfakes, as the AI generated the main elements of content and tone of voice. Unlike with satirical depictions or artistic debate, there is no recognisable human expression apart from the initial prompt and the act of dissemination. Nevertheless, the risk of a chilling effect on legal forms of expression must also be considered. This risk, however, can be reduced by clear, foreseeable and proportionate regulations[295], for example by legally defined exceptions for artistic, satirical and journalistic contexts, as set out in Article 50(4) AI Act.

Overall, the state's interest in protection outweighs the freedom of expression of the disseminator in the case of manipulative political deepfakes. The targeted dissemination of an objectively false, synthetically generated sound recording in a highly sensitive election context interferes deeply with the democratic decision-making process. The expression of individual self-realisation is hardly affected here, as the AI generates the expressive content. At the same time, the risk of considerable damage to trust in democratic processes is evident. The DSA and the AI Act ensure a differentiated assessment that limits the interference with freedom of expression: Art. 14 (4) DSA obliges providers to act in accordance with fundamental rights, particularly with regard to freedom of expression. Art. 50 (4) AI Act allows exceptions to strict labelling requirements for artistic, creative, or satirical works. While the transparency obligations do still allow the dissemination of opinions, they require transparency about the artificial origin of manipulated content. This makes it easier for recipients to categorise content, protects the integrity of public debates, and enables affected individuals, such as politicians, to defend their reputation more effectively. Therefore, the state intervention remains proportionate and contributes to a balanced protection of conflicting rights.

Another issue is that the transparency obligations and sanction mechanisms in the DSA and AI

[294] *E.s. v. Austria* (n 143) para.55; Grabenwarter and Pabel (n 117) para.52.
[295] Laurent Pech, 'The Concept of Chilling Effect; Its Untapped Potential to Better Protect Democracy, the Rule of Law, and Fundamental Rights in the EU' [2021] Open Society Justice Initiative, Open Society Foundations 5.

Act target platform operators, providers and users of AI systems in professional contexts, instead of the individuals who create and disseminate political deepfakes (Article 2(1)(2) DSA, Article 2(10) AI Act). Nevertheless, it is not disproportionate to focus responsibility on systemically relevant players. Large platforms and providers have effective means at their disposal to counter the risk of the rapid and widespread dissemination of manipulative content.[296] Given the severe threat that deceptively genuine deepfakes pose to the information sphere and thereby the democratic processes, this prioritisation is objectively justified.

f. Interim Conclusion

The analysis of the Slovakian deepfake case shows how harmful qualitative information pollution through targeted, AI-generated disinformation can be when it comes to undermining electoral integrity and violating personal rights. European regulations such as the DSA and the AI Act provide a clear and proportionate framework to contain such threats. These fulfil the requirements of a legal basis within the meaning of Art. 10 §2 ECHR, while pursuing legitimate aims such as protecting the rights of others and preserving national security and preventing disorder. The targeted dissemination of an objectively false, manipulated audio recording shortly before an election constitutes a severe interference with the information sphere in form of qualitative pollution. Given the minimal human involvement and the high deceptive impact, the freedom of expression of the disseminator is less worthy of protection. Consequently, the state's interest in protection takes precedence, given that the regulations allow for exceptions to the strict requirements in the case of artistic and political expression.

2. Use Case for Quantitative and Formal Infollution: Labelling and Filter Mechanism

In addition to the targeted manipulation of content, the sheer quantity of synthetic content can pollute the digital information sphere. This section therefore analyses a second case in which the focus is on the quantity and formal structure of information. The suitability of measures such as labelling obligations and filter functions as mild measures against overload is examined. Again, the analysis follows the criteria of Article 10 §2 ECHR. The result shows that such measures enable a fair balance between the recipient's interests and freedom of expression and are therefore compatible with Article 10 §2 ECHR.

[296] Philipp J Schneider and Marian-Andrei Rizoiu, 'The Effectiveness of Moderating Harmful Online Content' (2023) 120 Proceedings of the National Academy of Sciences of the United States of America (PNAS) 2–3.

a. Factual Background

AI-generated content is already having a significant impact on the quantity and form of information in the information sphere (see p.9f.). The large amount of synthetic content, and the increased fragmentation of the digital information sphere due to algorithms are undermining the functionality of public communication: Studies show that a substantial proportion of online content is already AI-generated, which indicates that the information sphere is already structurally polluted.[297] In theory, the right to receive information remains accessible under these conditions: it is possible to navigate confusing content by using search engines, searching specifically for trustworthy sources, or simply avoiding questionable content. In practice, however, this ideal is being challenged by the increasing user overload: Studies show that only 0.63% of users open the second page of search results[298], that cognitive load when processing information is increasing[299] and that many people abandon their search before finding the information they are looking for.[300] Another structural problem is that, although many users prefer non-AI-generated content[301], they are often unable to recognise whether content has been AI-generated.[302] While the right to receive information is not legally negated, it is impaired in practice.

b. State Interference

The AI Act combats formal and quantitative information pollution by imposing certain transparency obligations. According to Article 50(2), providers of generative AI systems must label synthetic content in a machine-readable way. Deployers must disclose artificially generated or manipulated content under Article 50(4), unless it is used exclusively for artistic, satirical or law enforcement purposes. Article 50(5) emphasises the goal of users being able to distinguish between genuine and artificially generated content. The AI Act thereby takes a structural approach; it does not interfere with individual expressions of opinion, but rather

---

[297] Spennemann (n 1) 2–7; Yiluo Wei and Gareth Tyson, 'Understanding the Impact of AI-Generated Content on Social Media: The Pixiv Case', *Proceedings of the 32nd ACM International Conference on Multimedia* (ACM 2024) 6820 <https://dl.acm.org/doi/10.1145/3664647.3680631> accessed 16 May 2025.

[298] Brian Dean, 'We Analyzed 4 Million Google Search Results. Here's What We Learned About Organic CTR' (*Backlinko*, 27 August 2019) <https://backlinko.com/google-ctr-stats> accessed 16 May 2025.

[299] Brianda Sígolo and Helen Casarin, 'Contributions of Cognitive Load Theory to Understanding Information Overload: A Literature Review' (2024) 22 RDBCI Revista Digital de Biblioteconomia e Ciência da Informação e024027, 4.

[300] Nina Paulsen and Tra My Nguyen, 'Jeder Dritte scheitert bei der Internetrecherche | Presseinformation | Bitkom e. V.' (*bitkom*, 27 January 2020) <https://www.bitkom.org/Presse/Presseinformation/Jeder-Dritte-scheitert-bei-der-Internetrecherche> accessed 16 May 2025.

[301] 'Truth: People Value What They Know Most - Other People' (*Baringa*, 25 February 2025) <https://www.baringa.com/en/insights/balancing-human-tech-ai/truth/> accessed 16 May 2025.

[302] Johanna Fleckenstein and others, 'Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays' (2024) 6 Computers and Education: Artificial Intelligence 100209, 4.

addresses the systemic risks posed by non-transparent, mass-distributed AI content. It does not reduce the quantity of content but facilitates its categorisation. It thereby reduces formal information pollution and at the same time minimises the cognitive overload experienced by users caused by quantitative information pollution.
Building on the mandatory labelling of synthetic content set out in Article 50 AI Act, this thesis goes a step further. One possible next step would be to introduce a mandatory filter for search engines that sorts content according to its origin in a traffic light system: This would involve red for content generated solely by AI, yellow for mixed content, and green for content generated by humans. This traffic light system would not restrict access to information, but rather structure it, thereby giving users the ability to decide what kind of content they wish to access and thus facilitating passive freedom of information.

c. Prescribed by Law

The AI Act fulfils the requirements of a law under Article 10 §2 ECHR: it is accessible and sufficiently specific, and it enables those affected to adapt their behaviour. Any hypothetical supplementary regulation on a mandatory filter mechanism for search engines would also have to fulfil these requirements, being clearly formulated, accessible to and foreseeable for those affected by it.

d. Legitimate Aim

The obligation to label and enable the filtering of AI-generated content aims to safeguard the structural requirements for a functioning digital information sphere. Even if the protection of the information sphere is not explicitly mentioned in Article 10 §2 ECHR, it can be subsumed under the recognised legitimate aims of protecting of others' rights, ensuring national security, and preventing disorder.
Firstly, passive freedom of information under Article 10 §1 ECHR presupposes a functioning information sphere.[303] In an overloaded, formally unstructured information environment, important content for individual users risks being lost in the digital '*noise*'.[304] If information cannot be found or recognised as trustworthy, the right to information is effectively devalued (p.39f.). Labelling and filtering obligations provide structural orientation and thus ensure access to relevant information.[305] These obligations therefore not only protect public communication in the abstract, but also the individual right to information in concrete terms.

---

303 Pentney (n 216) 248–249.
304 Klapp (n 80) as cited in; Edmunds and Morris (n 80) 18–19.
305 Shahrzadi and others (n 77) 9.

Secondly, the interference can also be categorised as falling under the aim of protection of the democratic order, and therefore as a legitimate aim of national security and preventing disorder. A functioning, transparent information sphere is essential for pluralistic opinion formation and democratic participation.[306] When users can no longer tell whether a contribution originates from a human or a machine, trust in public debates and digital information sources weakens.[307] Therefore, measures that counteract this loss of orientation and stabilise democratic discourse serve to protect core democratic values.

e. Necessary in a Democratic Society

The proposed labelling and filtering mechanisms must be necessary in a democratic society within the meaning of Article 10 §2 ECHR. This requires a careful balancing of the freedom of expression of those who create or disseminate AI-generated content and the legitimate aims presented above.

On the one hand, quantitative and formal information pollution poses a structural threat to the digital information sphere. It makes access to reliable content more difficult and can lead to cognitive overload, information fatigue, and ultimately, alienation from public communication. The proposed measures provide guidance and reduce the cognitive burden. They thus strengthen passive freedom of information, promote trust in digital information sources, and contribute to stabilising democratic discourse. The ECtHR has recognised that states can take structural measures to ensure pluralism and accessibility in the media environment, provided that this is necessary for meaningful public discourse.[308]

On the other hand, there is a risk that AI-generated content could be systematically suppressed due to filtering possibilities. This could result in such content receiving less visibility, effectively restricting its reach, particularly if algorithmic sorting mechanisms are based on user preferences. This could interfere with the speaker's freedom of expression.

Nevertheless, this is a milder measure than alternatives such as a complete ban on AI-generated content. The proposed regulation does not constitute a form of state censorship; it merely allows users to recognise its origin and make their own decisions based on this knowledge. This autonomy strengthens the right to receive information in accordance with Article 10 §1 ECHR rather than restrict the exercise of freedom of expression on side of the speaker. Furthermore, speakers are free to create their contributions without AI support to achieve greater reach. This

[306] Kalli Giannelos, 'Recommendations for a Healthy Digital Public Sphere' (2023) 38 Journal of Media Ethics 80, 82–83.

[307] Kreps and Kriner (n 263) 127.

[308] *Vgt Verein Gegen Tierfabriken v Switzerland* [2001] ECtHR 24699/94 paras.73-74,78.

measure therefore structurally protects the balance between freedom of communication and democratic participation. Moreover, the technical burden on AI developers and search engine providers appears to be limited. Implementing a basic filter function is comparable to existing tools like a time or language filter. Therefore, the measures are feasible in practice.
Overall, labelling and enabling filters are proportionate measures: They allow AI-generated content to remain in the public sphere, while strengthening recipients' freedom of choice and trust in the democratic information system.

f. Interim Conclusion

Labelling AI-generated content and providing the option to filter it is a mild, structural means of counteracting quantitative and formal information pollution. This approach safeguards passive freedom of information, strengthens trust in digital communication spaces, and promotes democratic participation. Although the visibility of AI-generated content may decrease, it can still be disseminated. Consequently, the intervention appears to be proportionate and therefore necessary in a democratic society.

III.Concluding Remarks on Chapter 3

The chapter demonstrated that restrictions on AI-generated content can be consistent with Article 10 §2 ECHR, provided that they are prescribed by law, pursue a legitimate aim, and are considered necessary in a democratic society. The assessment of necessity always depends on the specific context. The content, origin and effect of the statement are decisive factors. False information intended to manipulate only enjoys a low level of protection, whereas posts relating to public discourse deserve greater protection.
The two case studies analysed (qualitative information pollution through deepfakes in election campaigns and structural overload due to formal and quantitative information pollution) demonstrate that interference with freedom of expression is justified when the pollution of the information sphere has a tangible impact on individuals' rights, national security or the prevention of disorder.
This chapter showed that new phenomena, such as AI-generated content, can be encompassed by the existing assessment framework of Article 10 §2 ECHR. The developed criteria can be applied without fundamental legal challenges. Although the protection of the information sphere is not specifically mentioned in Article 10 §2 ECHR, it can be subsumed under the exhaustive list of legitimate aims without any fundamental legal challenges, given the new scale, speed and scope of these phenomena.

Conclusion and Further Research

This thesis analysed the extent to which AI-generated content falls within the scope of the protection of freedom of expression and information under Article 10 §1 ECHR, and the conditions under which state restrictions can be justified in order to protect the integrity of the information sphere.

Generative systems pollute the information sphere qualitatively through disinformation and distortion, quantitatively through information overload, and formally through the loss of context and structure. This makes access to reliable information more difficult and jeopardises trust, democratic processes, and individual self-fulfilment.

AI-generated content does not fall fully under the scope of protection of Article 10 §1 ECHR. The decisive factor is either a human act of attribution or a functional connection to public discourse. Users who deliberately disseminate content or make it their own, and recipients of the content, are eligible for protection.

Interference by the state in AI-generated content is generally compatible with Article 10 §2 ECHR, provided that it pursues a legitimate aim and is necessary in a democratic society and fulfils a *'pressing social need'*. The protection of the integrity of the information sphere can be subsumed under several legitimate aims. Qualitative information pollution, such as targeted disinformation, may be restricted if it poses a specific threat. A context-related, case-by-case assessment is always required, which takes into account the type, origin and impact of the content. The quantitative pollution of the information sphere is more difficult to address directly, as broad restrictions on volume or access raise concerns about censorship and sharply interfere with the freedom of expression for the speakers affected. However, protection against quantitative pollution can be achieved alongside the fight against formal pollution through structural measures like labelling and transparency rules combined with mandatory filtering tools in search engines. These measures reduce the user's cognitive overload and enable autonomous information choices without content being banned outright.

The analysis shows that, although AI-generated content presents new challenges, these can be addressed within the existing framework of Article 10 ECHR. The established case law on the criteria for assessing state interference can be applied to AI-generated content without encountering any significant legal challenges.

Further research should systematically and empirically investigate the extent to which AI-generated content contributes to information pollution. The categorisation of information pollution as qualitative, quantitative or formal, developed in this paper, should serve as an analytical frame of reference to record the effects on democratic communication conditions in a differentiated and reliable manner. Based on these findings, it should also be examined whether and in which constellations positive obligations for states arise from Article 10 ECHR.

Bibliography

Secondary Sources

1. Alemohammad S and others, 'Self-Consuming Generative Models Go MAD' (2024) <https://scholars.houstonmethodist.org/en/publications/self-consuming-generative-models-go-mad> accessed 22 April 2025

2. Allen PD and Gilbert Jr. DP, 'Qualifying the Information Sphere as a Domain' (2010) 9 Journal of Information Warfare 39

3. Baack S, 'A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl', *The 2024 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2024) <https://dl.acm.org/doi/10.1145/3630106.3659033> accessed 22 April 2025

4. Bates E, *The Evolution of the European Convention on Human Rights: From Its Inception to the Creation of a Permanent Court of Human Rights* (Oxford Academic 2011)

5. Baumann F and others, 'Dynamics of Algorithmic Content Amplification on TikTok' (arXiv, 26 March 2025) <http://arxiv.org/abs/2503.20231> accessed 18 June 2025

6. Bawden D, Holtham C and Courtney N, 'Perspectives on Information Overload' (1999) 51 ASLIB PROCEEDINGS 249

7. Bawden D and Robinson L, 'Information Overload: An Introduction', *Oxford Research Encyclopedia of Politics* (Oxford University Press 2020)

8. Belabbes MA and others, 'Information Overload: A Concept Analysis' 79 Journal of Documentation 144

9. Bender EM and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 22 April 2025

10. Bender EM and Koller A, 'Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data', *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics 2020) <https://www.aclweb.org/anthology/2020.acl-main.463> accessed 23 April 2025

11. Beutel G, Geerits E and Kielstein JT, 'Artificial Hallucination: GPT on LSD?' (2023) 27 Critical Care 148

12. Bjorge E, 'Proportionality' in Eirik Bjorge (ed), *Domestic Application of the ECHR: Courts as Faithful Trustees* (Oxford University Press 2015)

14. Bolukbasi T and others, ‘Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings’ NIPS’16: Proceedings of the 30th International Conference on Neural Information Processing Systems 4356

15. Bonicalzi S, De Caro M and Giovanola B, ‘Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems’ (2023) 42 Topoi 819

16. Bontcheva K (ed.), ‘Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities’ (University of Sheffield 2024) <https://edmo.eu/wp-content/uploads/2023/12/Generative-AI-and-Disinformation_-White-Paper-v8.pdf> accessed 26 May 2025

17. Bontridder N and Poullet Y, ‘The Role of Artificial Intelligence in Disinformation’ (2021) 3 Data & Policy e32

18. Boulding KE, *Economics as a Science* (Tata McGraw-Hill Publishing Company Ltd 1970)

19. Brożek B and Jakubiec M, ‘On the Legal Responsibility of Autonomous Machines’ (2017) 25 Artificial Intelligence and Law 293

20. Cai K-Y and Zhang C-Y, ‘Towards a Research on Information Pollution’, *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)* (1996)

21. Calderonio V, ‘A Basis for Human Responsibility in Artificial Intelligence Computation’ (arXiv, 21 January 2025) <http://arxiv.org/abs/2501.12498> accessed 28 April 2025

22. Calhoun C, ‘The Public Sphere in the Field of Power’ (2010) 34 Social Science History 301

23. Cameron GT, Ju-Pak K-H and Kim B-H, ‘Information Pollution?: Labeling and Format of Advertorials’ (2000) 21 Newspaper Research Journal 65

24. Canetta T, ‘Prebunking AI-Generated Disinformation Ahead of EU Elections – EDMO’ <https://edmo.eu/publications/prebunking-ai-generated-disinformation-ahead-of-eu-elections/> accessed 24 June 2025

25. Cannie H and Voorhoof D, ‘The Abuse Clause in International Human Rights Law: An Expedient Remedy against Abuse of Power or an Instrument of Abuse Itself?’, *Bullying and the Abuse of Power, James O’Higgins Norman, Sameer Hinduja (eds.)* (Inter-Disciplinary Press 2010) <http://hdl.handle.net/1854/LU-1080884> accessed 9 May 2025

26. Čerka P, Grigienė J and Sirbikytė G, ‘Is It Possible to Grant Legal Personality to Artificial Intelligence Software Systems?’ (2017) 33 Computer Law & Security Review 685

27. Cheong I, ‘Freedom of Algorithmic Expression’ (2023) 91 Social Science Research Network 680

29. Chesterman S, 'ARTIFICIAL INTELLIGENCE AND THE LIMITS OF LEGAL PERSONALITY' (2020) 69 International & Comparative Law Quarterly 819

30. Cho PJ, '"Infollution" and the Quality of Life' (2002) <https://stanford.edu/~ncho/Infollution_manuscript_PJCho_2002.pdf> accessed 28 May 2025

31. 'Conservative Activist Robby Starbuck Sues Meta over AI Responses about Him' (*AP News*, 30 April 2025) <https://apnews.com/article/robby-starbuck-meta-ai-delaware-eb587d274fdc18681c51108ade54b095> accessed 13 May 2025

32. Cornils M, 'EMRK Art. 10 Freiheit Der Meinungsäußerung' in Gersdorf and Paal (eds), *BeckOK Informations- und Medienrecht* (47th edn, 2024)

33. Crawford K, 'Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence', *Atlas of AI* (Yale University Press 2021)

34. Cresci S, 'A Decade of Social Bot Detection – Communications of the ACM' (2020) 63 Communications of the ACM 72

35. Cummins Gauthier C, 'Right to Know, Press Freedom, Public Discourse' (1999) 14 Journal of Mass Media Ethics 197

36. Daneshjou R and others, 'Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review' (2021) 157 JAMA Dermatology 1362

37. Day T, 'A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT' (2023) 75 The Professional Geographer 1024

38. De Nadal L and Jančárik P, 'Beyond the Deepfake Hype: AI, Democracy, and "the Slovak Case"' Harvard Kennedy School (HKS) Misinformation Review <https://misinforeview.hks.harvard.edu/article/beyond-the-deepfake-hype-ai-democracy-and-the-slovak-case/> accessed 18 June 2025

39. De Vries K, 'Let the Robot Speak! AI-Generated Speech and Freedom of Expression' in Steffen Hindelang and Andreas Moberg (eds), *YSEC Yearbook of Socio-Economic Constitutions 2021*, vol 2021 (Springer International Publishing 2022)

40. Dean B, 'We Analyzed 4 Million Google Search Results. Here's What We Learned About Organic CTR' (*Backlinko*, 27 August 2019) <https://backlinko.com/google-ctr-stats> accessed 16 May 2025

41. Divya, V and Mirza, Agha Urfi, 'Chapter 8, Transforming Content Creation: The Influence of Generative AI on a New Frontier', *Exploring the Frontiers of Artificial Intelligence and Machine Learning Technologies* (1st edn, San International Scientific Publications 2024) <https://www.researchgate.net/publication/379654562> accessed 24 April 2025

42. Dopplinger L, 'Legal Persons as Bearers of Rights Under the ECHR' (2021) 5 University of Vienna Law Review 1

44. Eadon YM and Wood SE, 'Combating Contamination and Contagion: Embodied and Environmental Metaphors of Misinformation' [2024] Convergence 13548565241255347

45. Edmunds A and Morris A, 'The Problem of Information Overload in Business Organisations: A Review of the Literature' (2000) 20 International Journal of Information Management 17

46. 'Election Disinformation in Slovakia – Disinfo Bulletin 9 April' (*EDMO European Digital Media Observatory*) <https://ec.europa.eu/newsroom/edmo/newsletter-archives/52231> accessed 24 June 2025

47. Eppler MJ and and Mengis J, 'The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines' (2004) 20 The Information Society 325

48. European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Defence of Democracy' (2023) COM(2023) 630 final <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52023DC0630> accessed 13 May 2025

49. European Court of Human Rights, 'Guide on Article 10 of the European Convention on Human Rights – Freedom of Expression' (Council of Europe 2022)

50. ——, 'Practical Guide on Admissibility Criteria' (2024) <https://www.echr.coe.int/documents/d/echr/admissibility_guide_eng> accessed 7 May 2025

51. 'Expression' <https://doi.org/10.1093/OED/8156500661> accessed 2 May 2025

52. Falkinger J, 'Limited Attention as a Scarce Resource in Information-Rich Economies' (2008) 118 The Economic Journal 1596

53. Ferrara E and others, 'The Rise of Social Bots' (2016) 59 Communications of the ACM 96

54. Filippova K, 'Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data' [2020] Findings of the Association for Computational Linguistics: EMNLP 2020 864

55. Firat M and Kurt AA, 'The Views of Candidate Teachers of Information Technologies about Information Pollution on the Internet' (WSEAS Press 2008) <https://www.researchgate.net/publication/273202478_The_Views_of_Candidate_Teachers_of_Information_Technologies_about_Information_Pollution_on_the_Internet>

56. Fleckenstein J and others, 'Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays' (2024) 6 Computers and Education: Artificial Intelligence 100209

57. Floridi L, *Philosophy and Computing: An Introduction* (Routledge 1999)

59. Forelle MC and others, ‘Political Bots and the Manipulation of Public Opinion in Venezuela’ [2015] SSRN Electronic Journal 1

60. Fritz J, ‘Understanding Authorship in Artificial Intelligence-Assisted Works’ [2025] Journal of Intellectual Property Law and Practice 1

61. Gellers JC, *Rights for Robots: Artificial Intelligence, Animal and Environmental Law (Edition 1)* (Routledge 2020)

62. Giannelos K, ‘Recommendations for a Healthy Digital Public Sphere’ (2023) 38 Journal of Media Ethics 80

63. Goswami M (Veenu), ‘Algorithms and Freedom of Expression’ in Woodrow Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (1st edn, Cambridge University Press 2020)

64. Grabenwarter C and Pabel K, ‘§ 18. Struktur Der Grundrechtsprüfung’ in Christoph Grabenwarter and Pabel Pabel (eds), *Europäische Menschenrechtskonvention* (7th edn, 2021)

65. ——, ‘§ 23. Politische Und Gemeinschaftsbezogene Grundrechte’ in Grabenwarter and Pabel (eds), *Europäische Menschenrechtskonvention* (7th edn, 2021)

66. Greer S, *The European Convention on Human Rights: Achievements, Problems and Prospects* (Cambridge University Press 2006)

67. Gudowsky N and Bechtold U, ‘The Role of Information in Public Participation’ (2013) 9 Journal of Deliberative Democracy

68. Guess AM and Lyons BA, ‘Misinformation, Disinformation, and Online Propaganda’ in Joshua A Tucker and Nathaniel Persily (eds), *Social Media and Democracy* (Cambridge University Press 2020)

69. Guo D and others, ‘AIGC Challenges and Opportunities Related to Public Safety: A Case Study of ChatGPT’ (2023) 4 Journal of Safety Science and Resilience 329

70. Hardinges J, Simperl E and Shadbolt N, ‘We Must Fix the Lack of Transparency Around the Data Used to Train Foundation Models’ [2024] Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/xau9dza3/release/2> accessed 24 April 2025

71. Hasimi L and Poniszewska-Marańda A, ‘Detection of Disinformation and Content Filtering Using Machine Learning: Implications to Human Rights and Freedom of Speech’

72. Hiltz SR and Turoff M, ‘Structuring Computer-Mediated Communication Systems to Avoid Information Overload’ (1985) 28 Communications of the ACM 680

73. Hofmann R, ‘Internationaler Menschenrechtsschutz, 1. Teil: Universeller Menschenrechtsschutz’ (2012) <https://www.jura.uni-frankfurt.de/43671833/Skript_Internationaler_Menschenrechtsschutz_SS_12.pdf?> accessed 2 May 2025

75. Immorlica N, Jagadeesan M and Lucier B, ‘Clickbait vs. Quality: How Engagement-Based Optimization Shapes the Content Landscape in Online Platforms’, *Proceedings of the ACM Web Conference 2024* (ACM 2024) <https://dl.acm.org/doi/10.1145/3589334.3645353> accessed 24 April 2025

76. ‘Introducing ChatGPT’ (13 March 2024) <https://openai.com/index/chatgpt/> accessed 22 April 2025

77. Iqbal Q and others, ‘Infollution (Information Pollution) Management, Filtering Strategy, Scalable Workforce, and Organizational Learning: A Conceptual Study’ (2018) 10 Information Management and Business Review 1

78. Iqbal S, ‘How Social Media Algorithms Affect Content Visibility’ (*Content Whale*, 28 April 2025) <https://content-whale.com/blog/impact-of-social-media-algorithms-on-content-visibility/> accessed 18 June 2025

79. Islam MBE and others, ‘AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework’ (2024) 2 Blockchains 458

80. Ji J, Goldstein JA, and Andrew J. Lohn, ‘Controlling Large Language Model Outputs: A Primer’ (Center for Security and Emerging Technology (CSET), Georgetown University 2023) <https://cset.georgetown.edu/publication/controlling-large-language-models-a-primer/> accessed 9 May 2025

81. Ji Z and others, ‘Survey of Hallucination in Natural Language Generation’ (2023) 55 ACM Comput. Surv. 248:1

82. Johnson A, ‘What Is the Difference between AI-Assisted and AI-Generated Content?’ (3 June 2024) <https://www.clearscope.io/blog/ai-generated-vs-ai-assisted-content> accessed 9 May 2025

83. Joint Chiefs of Staff, United States Department of Defense, ‘Department of Defense Dictionary of Military and Associated Terms’ 1 <https://irp.fas.org/doddir/dod/jp1_02.pdf> accessed 24 April 2025

84. Joosten P, ‘Human Enhancement Technologies: 7 Examples’ (*Peter Joosten*, 22 May 2020) <https://www.peterjoosten.org/human-enhancement-technologies/> accessed 8 May 2025

85. Karhade P and others, ‘Digital Platforms and Infobesity: A Research Agenda’ in Aravinda Garimella and others (eds), *The Role of e-Business during the Time of Grand Challenges*, vol 418 (Springer International Publishing 2021) <https://link.springer.com/10.1007/978-3-030-79454-5_7> accessed 18 April 2025

86. Klapp OE, *Overload and Boredom: Essays on the Quality of Life in the Information Society* (Greenwood Publishing Group Inc 1986)

87. Klein J and others, ‘Synthetic Data at Scale: A Development Model to Efficiently Leverage Machine Learning in Agriculture’ (2024) 15 Frontiers in Plant Science 1

89. Koops B-J, Hildebrandt M and Jaquet-Chiffelle D-O, ‘Bridging the Accountability Gap: Rights for New Entities in the Information Society?’

90. Kreps S and Kriner D, ‘How AI Threatens Democracy’ (2023) 34 Journal of Democracy 122

91. Łabuz M and Nehring C, ‘On the Way to Deep Fake Democracy? Deep Fakes in Election Campaigns in 2023’ (2024) 23 European Political Science 454

92. Lareo X, ‘Large Language Models (LLM) | European Data Protection Supervisor’ (13 June 2025) <https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm> accessed 24 June 2025

93. Lee B-K and Lee W-N, ‘The Effect of Information Overload on Consumer Choice Quality in an On-Line Environment’ (2004) 21 Psychology & Marketing 159

94. Lee P, ‘Learning from Tay’s Introduction’ (*The Official Microsoft Blog*, 25 March 2016) <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/> accessed 8 May 2025

95. Levy M, ‘Information Pollution & the Public Discourse’ (2022) 6 Journal of Law and Technology at Texas (JOLTT) 113

96. Li D, He W and Guo Y, ‘Why AI Still Doesn’t Have Consciousness?’ (2021) 6 CAAI Transactions on Intelligence Technology 175

97. Lin Z, ‘Techniques for Supercharging Academic Writing with Generative AI’ (2025) 9 Nature Biomedical Engineering 426

98. Lucas W, ‘The Influence of Social Media Algorithms on Racial and Ethnic Misinformation: Patterns and Impacts’ (2024)

99. Ma Y and Li F, ‘Detecting Review Spam: Challenges and Opportunities’, *Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (IEEE 2012) <http://eudl.eu/doi/10.4108/icst.collaboratecom.2012.250640> accessed 24 April 2025

100. Marciel R, ‘On Citizens’ Right to Information: Justification and Analysis of the Democratic Right to Be Well Informed’ (2023) 31 Journal of Political Philosophy 358

101. Martínez G and others, ‘Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet’ in Fabio Cuzzolin and Maryam Sultana (eds), *Epistemic Uncertainty in Artificial Intelligence* (Springer Nature Switzerland 2024)

102. Massaro TM, Norton H and Kaminski ME, ‘SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment’ (2017) 101 Minnesota Law Review 2481

103. McKay S and Tenove C, ‘Disinformation as a Threat to Deliberative Democracy’ (2021) 74 Political Research Quarterly 703

104. McMillan I, ‘Enforcement Through the Network: The Network Enforcement Act and Article 10 of the European Convention on Human Rights’ (2019) 20 Chicago Journal of International Law 255

105. Meel P and Vishwakarma DK, 'Fake News, Rumor, Information Pollution in Social Media and Web: A Contemporary Survey of State-of-the-Arts, Challenges and Opportunities' (2020) 153 Expert Systems with Applications 112986

106. Mitra B, Cramer H and Gurevich O, 'Sociotechnical Implications of Generative Artificial Intelligence for Information Access' (arXiv, 16 July 2024) <http://arxiv.org/abs/2405.11612> accessed 24 April 2025

107. ——, 'Sociotechnical Implications of Generative Artificial Intelligence for Information Access' in Ryen W White and Chirag Shah (eds), *Information Access in the Era of Generative AI* (Springer Nature Switzerland 2025)

108. Montagnani ML, Najjar M-C and Davola A, 'The EU Regulatory Approach(Es) to AI Liability, and Its Application to the Financial Services Market' (2024) 53 Computer Law & Security Review 105984

109. Nielsen J, 'IM, Not IP (Information Pollution): A Steady Dose of Realtime Interruptions Is Toxic to Anyone's Health.' (2003) 1 Queue 76

110. Northoff G and Gouveia SS, 'Does Artificial Intelligence Exhibit Basic Fundamental Subjectivity? A Neurophilosophical Argument' (2024) 23 Phenomenology and the Cognitive Sciences 1097

111. Onebunne A, 'Algorithmic Bias and Media Manipulation: A Systematic Review of AI's Role in Shaping Public Perception and Political Discourse' (2022) 16 World Journal of Advanced Research and Reviews 1239

112. 'Open Data for the World: Common Crawl Frees Petabytes of Web Data' (22 May 2024) <https://datos.gob.es/en/blog/open-data-world-common-crawl-frees-petabytes-web-data> accessed 9 May 2025

113. Orman L, 'Fighting Information Pollution with Decision Support Systems' (1984) 1 Journal of Management Information Systems 64

114. Oster J, 'Haftung für Persönlichkeitsrechtsverletzungen durch Künstliche Intelligenz' (2018) 82 UFITA 14

115. Özdemir Ş, 'Individual Contributions to Infollution (Information Pollution): Trust and Share' (2016) 7

116. Palmieri E, 'Online Bubbles and Echo Chambers as Social Systems' (2025) 54 Kybernetes 2457

117. ——, 'Online Bubbles and Echo Chambers as Social Systems' (5 March 2025) <https://www.emerald.com/insight/content/doi/10.1108/K-09-2023-1742/full/html> accessed 25 April 2025

118. Pandita R, 'Information Pollution, a Mounting Threat: Internet a Major Causality' (2014) 2 Journal of Information Science Theory and Practice 49

119. Patel K and Chudasama D, 'National Security Threats in Cyberspace' (2021) 4 National Journal of Cyber Security Law 12

121. Paulsen N and Tra My Nguyen, ‘Jeder Dritte scheitert bei der Internetrecherche | Presseinformation | Bitkom e. V.’ (*bitkom*, 27 January 2020) <https://www.bitkom.org/Presse/Presseinformation/Jeder-Dritte-scheitert-bei-der-Internetrecherche> accessed 16 May 2025

122. Pech L, ‘The Concept of Chilling Effect; Its Untapped Potential to Better Protect Democracy, the Rule of Law, and Fundamental Rights in the EU’ [2021] Open Society Justice Initiative, Open Society Foundations

123. Pentney K, ‘The Right of Access to “Reliable” Information Under Article 10 Echr: From Meagre Beginnings to New Frontiers’ (2024) 5 European Convention on Human Rights Law Review, The 230

124. Phillips W and Milner RM, *You Are Here: A Field Guide for Navigating Polarized Speech, Conspiracy Theories, and Our Polluted Media Landscape* (The MIT Press 2021)

125. Pira F, ‘Disinformation a Problem for Democracy: Profiling and Risks of Consensus Manipulation’ (2023) 8 Frontiers in Sociology 1

126. Porter B and Machery E, ‘AI-Generated Poetry Is Indistinguishable from Human-Written Poetry and Is Rated More Favorably’ (2024) 14 Scientific Reports 26133

127. Romero Moreno F, ‘Generative AI and Deepfakes: A Human Rights Approach to Tackling Harmful Content’ (2024) 38 International Review of Law, Computers & Technology 297

128. Rothkopf DJ, ‘When the Buzz Bites Back’ (*University of Delaware – Global Agenda 2004*, 2003) <https://www1.udel.edu/globalagenda/2004/student/readings/infodemic.html> accessed 24 April 2025

129. Saab B, ‘Manufacturing Deceit: How Generative AI Supercharges Information Manipulation’ (National Endowment for Democracy (NED) 2024) <https://www.ned.org/manufacturing-deceit-how-generative-ai-supercharges-information-manipulation/> accessed 28 April 2025

130. Sadeghi M and Arvanitis L, ‘Rise of the Newsbots: AI-Generated News Websites Proliferating Online’ (*NewsGuard*, 1 May 2023) <https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating> accessed 24 April 2025

131. Schirmer J-E, ‘Rechtsfähige Roboter?’ (2018) 71 JuristenZeitung 660

132. Schneider PJ and Rizoiu M-A, ‘The Effectiveness of Moderating Harmful Online Content’ (2023) 120 Proceedings of the National Academy of Sciences of the United States of America (PNAS)

133. Schöndorf-Haubold B, *Das Recht auf Achtung des Privatlebens: Problemstellungen im Digitalbereich: eine rechtsvergleichende Perspektive - Deutschland* (Generaldirektion Wissenschaftlicher Dienst, Europäisches Parlament 2018)

134. Shahrzadi L and others, ‘Causes, Consequences, and Strategies to Deal with Information Overload: A Scoping Review’ (2024) 4 International Journal of Information Management Data Insights 100261

135. Sharma M and others, ‘Towards Understanding Sycophancy in Language Models’ (2024) <http://arxiv.org/abs/2310.13548> accessed 27 May 2025

136. Shoaib MR and others, ‘Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models’ (IEEE 2023) <http://arxiv.org/abs/2311.17394> accessed 28 April 2025

137. Shumailov I and others, ‘The Curse of Recursion: Training on Generated Data Makes Models Forget’ [2023] arXiv 1

138. Siebert DJ, ‘Halluzinationen von generativer KI und großen Sprachmodellen (LLMs) - Blog des Fraunhofer IESE’ (*Fraunhofer IESE*, 20 September 2024) <https://www.iese.fraunhofer.de/blog/halluzinationen-generative-ki-llm/> accessed 22 April 2025

139. Sígolo B and Casarin H, ‘Contributions of Cognitive Load Theory to Understanding Information Overload: A Literature Review’ (2024) 22 RDBCI Revista Digital de Biblioteconomia e Ciência da Informação e024027

140. Solum LB, ‘Legal Personhood for Artificial Intelligences’ (1992) 70 North Carolina Law Review 1231

141. Spennemann DH, ‘“Delving into” the Quantification of Ai-Generated Content on the Internet (Syn- Thetic Data)’ (Charles Sturt University 2025) <https://arxiv.org/pdf/2504.08755> accessed 24 April 2025

142. Sridhar MS, ‘Knowledge Dilution, Pollution, and Burden in Digital Media’ [2024] Journal of Information and Knowledge 109

143. Sun Y and others, ‘AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content’ (2024) 11 Humanities and Social Sciences Communications 1

144. Terziu A, ‘AI – the Threats It Poses to Reputation, Privacy and Cyber Security, and Some Practical Solutions to Combating Those Threats’ (*Taylor Wessing*, 18 April 2024) <https://www.taylorwessing.com/en/global-data-hub/2024/cyber-security---weathering-the-cyber-storms/ai---the-threats-it-poses-to-reputation> accessed 16 May 2025

145. Toffler A, *The Third Wave* (William Morrow and Company 1980)

146. Tomlinson B, Patterson DJ and Torrance AW, ‘Turning Fake Data into Fake News: The AI Training Set as a Trojan Horse of Misinformation’ (2023) 60 San Diego Law Review

147. ‘Truth: People Value What They Know Most - Other People’ (*Baringa*, 25 February 2025) <https://www.baringa.com/en/insights/balancing-human-tech-ai/truth/> accessed 16 May 2025

149. Tucker JA and others, ‘Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature’ (Social Science Research Network (SSRN) 2018)

150. Villalobos P and others, ‘Will We Run out of Data? Limits of LLM Scaling Based on Human-Generated Data’ (arXiv, 4 June 2024) <http://arxiv.org/abs/2211.04325> accessed 22 April 2025

151. Voorhoof D, ‘Restrictions on Television Advertising and Article 10 of the European Convention on Human Rights.’ [1993] INTERNATIONAL JOURNAL OF ADVERTISING 189

152. Voorhoof D and Cannie H, ‘Freedom of Expression and Information in a Democratic Society: The Added but Fragile Value of the European Convention on Human Rights’ (2010) 72 International Communication Gazette 407

153. Walorska AM, ‘Deepfakes and Disinformation’ (Friedrich-Naumann-Stiftung für die Freiheit 2020) <https://shop.freiheit.org/#!/Publikation/897> accessed 23 April 2025

154. Wang RY and Strong DM, ‘Beyond Accuracy: What Data Quality Means to Data Consumers’ (1996) 12 Journal of Management Information Systems 5

155. Wardle C and Derakhshan H, ‘Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making’ (2017) Council of Europe DGI(2017)09 <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html> accessed 18 April 2025

156. Wei Y and Tyson G, ‘Understanding the Impact of AI-Generated Content on Social Media: The Pixiv Case’, *Proceedings of the 32nd ACM International Conference on Multimedia* (ACM 2024) <https://dl.acm.org/doi/10.1145/3664647.3680631> accessed 16 May 2025

157. Westerlund M, ‘The Emergence of Deepfake Technology: A Review’ (2019) 9 Technology Innovation Management Review 39

158. Whitman JQ, ‘The Two Western Cultures of Privacy: Dignity versus Liberty’ (2004) 113 The Yale Law Journal 1151

159. Wolf MJ, Miller KW and Grodzinsky FS, ‘Why We Should Have Seen That Coming: Comments on Microsoft’s Tay “Experiment,” and Wider Implications’ (2017) 1 The ORBIT Journal 1

160. Wong W, ‘Trends in Political Science Research: Artificial Intelligence and Voter Disinformation’ (2025) 75 International Political Science Abstracts 205

161. Xing X and others, ‘On the Caveats of AI Autophagy’ (2025) 7 Nature Machine Intelligence 172

162. Zhou J and others, ‘Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions’, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (ACM 2023) <https://dl.acm.org/doi/10.1145/3544548.3581318> accessed 26 May 2025

163. Zhou Y and others, ‘Source Echo Chamber: Exploring the Escalation of Source Bias in User, Data, and Recommender System Feedback Loop’ (arXiv, 28 May 2024) <http://arxiv.org/abs/2405.17998> accessed 23 April 2025

164. Zi-yang H, ‘AIGC Related Context: A New Communication Culture For Human’ (2024) 14 Journal of Literature and Art Studies 921

165. Żmudka K and others, ‘ChatGPT – a Tool for Assisted Studying or a Source of Misleading Medical Information? AI Performance on Polish Medical Final Examination’ (2024) 78 Annales Academiae Medicae Silesiensis 94

Case Law

1. *Ahmet Yildirim v Turkey* [2012] ECtHR 3111/10
2. *Aksu v Turkey* [2012] ECtHR [GC] 4149/04, 41029/04
3. *Animal Defenders International v the United Kingdom* [2013] ECtHR [GC] 48876/08
4. *Ashby Donald Et Autres c France* [2013] ECtHR 36769/08
5. *Ashirov and International Memorial v Russia* [2023] ECtHR 25246/07
6. *Association Burestop 55 Et Autres c France* [2021] ECtHR 56176/18, 56189/18, 56232/18, 56236/18, 56241/18, 56247/18
7. *Autronic Ag v Switzerland* [1990] ECtHR 17112/16, 57630/17, 37339/19
8. *Axel Springer Ag v Germany* [2012] ECtHR [GC] 39954/08
9. *Baka v Hungary* [2016] ECtHR [GC] 20261/12
10. *Barthold v Germany* [1985] ECtHR 27378/16, 30348/16, 59267/16, 52237/17, 57432/17, 62904/17
11. *Bédat v Switzerland* [2016] ECtHR [GC] 56925/08
12. *Brasilier c France* [2006] ECtHR 71343/01
13. *C8 (canal 8) v France* [2023] ECtHR 58951/18, 1308/19
14. *Casado Coca v Spain* [1994] ECtHR 15450/89
15. *Centro Europa 7 S.r.l and Di Stefano v Italy* [2012] ECtHR [GC] 38433/09
16. *Chauvy and Others v France* [2004] ECtHR 64915/01
17. *Danileţ v Romania* [2024] ECtHR 16915/21
18. *Delfi as v Estonia* [2015] ECtHR [GC] 64569/09
19. *Demir and Baykara v Turkey* [2008] ECtHR [GC] 34503/97
20. *E.s v Austria* [2018] ECtHR 38450/12
21. *Feret c Belgique* [2009] ECtHR 15615/07

22. *German Communist Party and Others v Germany* [1957] ECtHR 250/57
23. *Gillberg v Sweden* [2010] ECtHR 41723/06
24. *Golder v the United Kingdom* [1975] ECtHR 57496/16
25. *Gough v the United Kingdom* [2018] ECtHR 2153/15
26. *Gözel Et Özer c Turquie* [2010] ECtHR 43453/04, 31098/05
27. *Guerra and Others v Italy* [1998] ECtHR [GC] 58135/09
28. *Halet v Luxembourg* [2023] ECtHR [GC] 21884/18
29. *Handyside v the United Kingdom* [1976] ECtHR 5493/72, ECLI:CE:ECHR:1976:1207JUD000549372
30. *Ibragimova v Russia* [2022] ECtHR 68537/13
31. *Informationsverein Lentia and Others v Austria* [1993] ECtHR 24753/13, 28610/13, 57854/13, 65553/13
32. *Jersild v Denmark* [1994] ECtHR [GC] 57891/17
33. *Karatas v Turkey* [1999] ECtHR [GC] 56079/21, 57743/21, 58274/21
34. *Khurshid Mustafa and Tarzibachi v Sweden* [2008] ECtHR 23883/06
35. *Klass and Others v Germany* [1978] ECtHR 5029/71
36. *Leander v Sweden* [1987] ECtHR 9248/81
37. *Leroy c France* [2008] ECtHR 36109/03
38. *Leyla Şahin v Turkey* [2005] ECtHR [GC] 44774/98
39. *Lingens v Austria* [1986] ECtHR 9815/82
40. *Magyar Jeti Zrt v Hungary* [2018] ECtHR 11257/16
41. *Magyar Kétfarkú Kutya Párt v Hungary* [2018] ECtHR [GC] 201/17
42. *Mariya Alekhina and Others v Russia* [2018] ECtHR 38004/12
43. *M'bala M'bala v France (dec)* [2015] ECtHR 25239/13
44. *Mcvicar v the United Kingdom* [2002] ECtHR 60450/13
45. *Müller and Others v Switzerland* [1988] ECtHR 10737/84
46. *Nagla v Latvia* [2013] ECtHR 73469/10
47. *Nit S.r.l v the Republic of Moldova* [2022] ECtHR [GC] 28470/12
48. *Norwood v the United Kingdom (dec)* [2004] ECtHR 23131/03
49. *Oberschlick v Austria* [1991] ECtHR 11662/85
50. *Paksas v Lithuania* [2011] ECtHR [GC] 34932/04
51. *Pentikäinen v Finland* [2015] ECtHR [GC] 11882/10
52. *Perinçek v Switzerland* [2015] ECtHR [GC] 27510/08

53. *Pfeifer v Austria* [2007] ECtHR 12556/03
54. *Refah Partisi (the Welfare Party) and Others v Turkey* [2003] ECtHR [GC] 60936/12
55. *Roj Tv a/S v Denmark (dec)* [2018] ECtHR 24683/14
56. *Rtbf v Belgium* [2011] ECtHR 50084/06
57. *S and Marper v the United Kingdom* [2008] ECtHR [GC] 30562/04, 30566/04
58. *Sanoma Uitgevers B.v v the Netherlands* [2010] ECtHR [GC] 38224/03
59. *S.a.s v France* [2014] ECtHR [GC] 43835/11
60. *Savva Terentyev v Russia* [2018] ECtHR 10692/09
61. *Schweizerische Radio- Und Fernsehgesellschaft Srg v Switzerland* [2012] ECtHR 34124/06
62. *Selmouni v France* [1999] ECtHR [GC] 26856/06, 26449/08, 61815/09, 30395/10, 50999/10, 58287/10, 63120/10, 36942/11
63. *Senator Lines Gmbh v Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom (déc)* [2004] ECtHR [GC] 23765/94
64. *Shapovalov v Ukraine* [2012] ECtHR 45835/05
65. *Soering v the United Kingdom* [1989] ECtHR 14038/88
66. *Sorguç v Turkey* [2009] ECtHR 17089/03
67. *Steel and Morris v the United Kingdom* [2005] ECtHR 68416/01
68. *Stoll v Switzerland* [2007] ECtHR [GC] 69698/01
69. *Társaság a Szabadságjogokért v Hungary* [2009] ECtHR 37374/05
70. *The Sunday Times v the United Kingdom (no 1)* [1979] ECtHR 6538/74
71. *Times Newspapers Ltd (nos 1 and 2) v United Kingdom* [2009] ECtHR App nos 3002/03 and 23676/03
72. *Tyrer v the United Kingdom* [1978] ECtHR 5856/72
73. *Vejdeland and Others v Sweden* [2012] ECtHR 1813/07
74. *Vgt Verein Gegen Tierfabriken v Switzerland* [2001] ECtHR 24699/94
75. *Von Hannover v Germany (no 2)* [2012] ECtHR [GC] 40660/08, 60641/08
76. *Weber v Germany (dec)* [2015] ECtHR 70287/11
77. *Wingrove v the United Kingdom* [1996] ECtHR 58080/11

Appendix 1: Disclosure of AI Use

Name: Katharina Goth

Student Number: 0933406

Date: 26^{th} June 2025

I hereby declare that I have only used AI tools for the purposes permitted in the 'Guidelines for the use of AI in Graduation Works and the Research and Thesis Track (OST) – January 2025'.

1.Google Scholar (AI-powered version, Google LLC)

14^{th}-25^{th} April 2025:

- Purpose**:** Finding relevant literature for Chapter 1, including definitions and historical context of the information sphere, concepts like information overload, disinformation, fake news and how to define AI-generated content.
- Input**:** Search queries such as: *'information pollution', 'definition of information pollution', 'information pollution misinformation disinformation', 'information pollution structure of digital content', 'historical context information pollution', 'Floridi infosphere', AI-generated content impact media', 'definition AI-generated content legal', 'recursive training loops generative AI', 'hallucinations LLM', 'AI bias disinformation media', 'deepfakes as disinformation strategy', 'AI in propaganda and fake news', 'mass generation synthetic content', 'infollution', 'information overload', 'information fatigue syndrome', 'filter bubbles and echo chambers', 'polluted information', 'misinformation disinformation malinformation', 'information pollution mass media', 'Ai autophagy', 'AI hallucinations', 'training data degradation synthetic content', 'information integrity', 'digital public sphere', 'information pollution democratic discourse', 'information disorder', 'AI-generated content trust in media', 'information reliability synthetic text', 'AI hallucination risks', 'synthetic content overload', 'ai-generated spam', 'training data pollution', 'content overload effects', 'bias in LLM', 'disinformation AI elections', 'news ecosystem AI', 'information ecosystem AI', 'public discourse AI', 'AI hallucination citation error', 'cognitive overload information systems', 'AI-generated clickbait', 'automated content manipulation', 'AI transparency information trust', 'pollution digital discourse', 'generative AI perception bias', 'AI attention economy', 'AI-generated content risks', 'training loop AI', 'infobesity AI', 'philosophical military information sphere', 'human influence AI-generated content', 'formats AI content', 'origin information sphere pollution', 'infodemic', 'irrelevant information sphere', 'manipulation information sphere', 'decontextualisation AI content', 'consequences information pollution', 'risks AI content', 'AI-generated content image video text', 'pollution information', 'infollution consequences', 'infobesity consequences', 'definition infollution', 'definition infobesity', 'AI-generated content forms pollution'.*
- Output**:** Google Scholar produced a wide range of peer-reviewed academic articles. These include the literature I found and referenced in the bibliography. These sources helped me to structure the development and categorisation of the information sphere its

pollution.

- Reflection**:** Google Scholar's AI enhanced ranking helped to find relevant and high-quality academic sources quickly. I critically assessed and selected the most useful ones.

28th April – 9th May 2025

- Purpose**:** Finding relevant literature for Chapter 2 on the scope of protection under Article 10 §1 ECHR, including human rights theory on speaker and recipient models, legal debates on AI and freedom of expression and conceptual distinctions on synthetic vs. human-attributable content.
- Input**:** Search queries such as: *'definition of expression Article 10 ECHR', 'Article 10 ECHR', 'origins ECHR', 'history ECHR', 'freedom of expression case law', 'freedom of expression ECtHR', 'freedom of information ECtHR interpretation', 'functional interpretation Article 10 ECHR', 'expression vs information ECHR', 'public discourse Article 10 ECHR', 'autonomy and freedom of expression ECHR', 'self fulfilment ECHR', 'public discourse Article 10 ECHR', 'ECtHR freedom of expression scope', 'personal scope ECHR Article 10', 'material scope ECHR Article 10', 'AI and legal personhood', 'freedom of expression AI ECHR human', 'recipient Article 10 ECHR', 'passive and active freedom of information ECHR', 'AI-generated speech ECHR', 'Article 17 ECHR', 'Article 17 abuse clause', 'Article 17 freedom of expression', 'value information public discourse', 'ECHR freedom of expression function', 'non-verbal expression ECHR', 'expression ECtHR symbolic acts', 'de Vries AI expression Article 10', 'Kaminski Norton Massaro AI First Amendment', 'protection information flow ECHR', Freedom of expression algorithms', 'Canadian charter freedom of expression AI', 'Article 10 scope of protection AI', 'challenges AI freedom of speech', 'synthetic content Article 10', 'machine speech human rights', 'ai communication legal status', 'freedom of expression attribution', 'expression through AI tools', 'freedom of expression human-machine', 'speech rights AI content', 'AI speech and democratic discourse', 'expression without authorship', 'expression Article 10 interpretation ECHR', 'purpose Article 10 ECHR', 'context freedom of expression Article 10 ECHR', 'travaux preparatoires object purpose Article 10 ECHR', 'interpretation methods ECHR Vienna', 'freedom of expression personal autonomy, self fulfilment', 'public discourse Article 10 echr case law', 'rights holder article 10 ECHR', 'AI personal scope ECHR', 'fundamental rights AI', 'Ai developer fundamental rights', 'User AI freedom of speech', 'recipient rights Article 10 ECHR', 'ECHR history', 'ECHR second worldwar', 'ECHR freedom of speech information', 'ECHR freedom of press freedom of information'.*
- Output**:** Google Scholar provided a broad range of peer-reviewed academic literature and case law, which was relevant and useful for the interpretation of Article 10 §1 ECHR. I retrieved sources on the material and the personal scope of protection, theoretical analyses of expression, information and press, ECtHR case law and insights on the role of autonomy and discourse and scholarly debates on AI-generated content. I also found literature with insights in US and Canadian law and on personhood and attribution of the content to human beings. The found literature is listed in the Footnotes and Bibliography.

- Reflection: The AI-supported search algorithm of Google Scholar helped me to get efficient access to academic sources of high relevance and quality. I critically reviewed all results, assessed their relevance and only included sources that supported the analysis of the interpretation of Article 10 §1 ECHR in relation to AI-generated content. The literature I found helped me to analyse the new phenomenon of AI-generated content within the existing framework of Article 10 §1 ECHR.

12th May – 23rd May 2025

- Purpose: Finding relevant literature for Chapter 3 on the justification of interferences with AI-generated content under Article 10 §2 ECHR with the focus set on the concepts of state interference, legality, legitimate aims in terms of the protection of the information sphere and the proportionality test (necessary in a democratic society). The research included theoretical sources, ECtHR case law, literature on democratic discourse and structural risks posed by AI and practical examples (the two use cases presented).
- Input: Search queries such as: *'Article 10 §2 ECHR', 'Article 10 §2 ECHR interpretation', 'Article 10 §2 interference', 'ECHR state interference definition', 'prescribed by law interpretation', 'prescribed by law ECHR', 'prescribed by law quality', 'freedom of expression prescribed by law', 'legitimate aims Article 10 ECHR', 'legitimate aims information pollution', 'legitimate aim exhaustive list', 'legitimate aims interpretation ECHR Article 10', 'necessary in a democratic society Article 10', 'pressing social need Article 10', 'freedom of expression proportionality test Article 10', 'national security Article 10 ECHR', 'prevention of disorder Article 10 ECHR', 'rights of others Article 10 ECHR', 'discrimination Article 14 ECHR', 'discrimination Article 10 ECHR Article 14', 'reputation Article 10 ECHR', 'Article 8 10 ECHR', 'public discourse Article 10 ECHR legitimate aim', 'public safety Article 10 ECHR', 'information sphere legitimate aim ECHR', 'legitimate aim AI ECHR', 'information pollution legal response', 'AI-generated content legal regulation Europe', 'ECtHR freedom of expression digital content', 'ECtHR dynamic interpretation', 'freedom of information deepfakes', 'deepfakes AI Act', 'deepfakes DSA', 'DSA', 'AI Act', 'evaluation DSA AI Act', 'Deepfakes Europe regulation', 'platform liability EU AI content', 'elections deepfakes AI', 'label content AI', 'watermarks AI content EU regulation', 'filter AI content Europe', '2023 Slovakia deepfake election', 'deepfake election input', 'consequences democracy AI content deepfake', 'overload filter ai content', 'political disinformation Europe regulation', 'AI Act transparency', 'AI Act watermarks AI content', 'obligations AI Act DSA AI content', 'synthetic content democratic discourse', 'satire exemption AI Act Article 50', 'pressing social need ECHR ECtHR', 'margin of appreciation Article 10 ECHR', 'freedom of expression digital age', 'public figure privacy Article 8 ECHR', 'von Hannover public figure', 'proportionality criteria Article 10 ECHR', 'necessary in a democratic society Article 10 ECHR criteria', 'national security and public order Article 10 ECHR', 'criteria justification Article 10 ECHR', 'factual statement value judgement Article 10 ECHR', 'chilling effect', 'chilling effect Article 10 ECHR', 'balance ECHR Article 10'.*
- Output: Google Scholar provided a wide range of academic literature and case law

which was often relevant to my analysis concerning Article 10 §2 ECHR. The search results included analyses of the legality requirement, legitimate aims, and the proportionality test (necessary in a democratic society) and provided academic literature on structural risks of AI-generated content and the information pollution to democratic discourse and the protection of the information sphere. The output also provided relevant information on the DSA and the AI act. It provided academic literature and reports on deepfake cases. I could use this literature to create an abstract legal analysis of the criteria of Article 10 §2 ECHR and to evaluate the two use cases concerning qualitative and quantitative/formal information pollution.

- Reflection**:** The AI-enhanced ranking algorithm of Google Scholar supported my search across legal academic sources. I carefully assessed the results for academic quality and relevance for my work and only selected sources that supported me in analysing AI-generated content and the pollution of the information sphere under Article 10 §2 ECHR.

2. DeepL Translate (DeepL SE)

14th April – 30th May 2025 + 9th June – 26th June 2025

- Purpose**:** Translation of self-written thesis sections, that I originally drafted in German into English, to comply with the thesis submission requirements at Utrecht University.
- Input**:** Own text written in German, including parts of the Introduction, Chapters 1-3 and the Conclusion of the thesis.
- Output**:** English-language versions of the self-written German draft sections. The translations were used as a linguistic structural basis for the final text. Each passage was revised, refined and adapted to ensure the alignment with academic standards, terminological precision and coherence throughout my thesis.
- Reflection**:** DeepL Translate provided a fluent and appropriate draft that supported me with the transition from German to English writing. However, the legal nuances often required post-editing, especially for terms that were specifically used under ECHR law and in ECtHR case law. The tool was useful for efficiency but did not replace critical legal judgment or linguistic revision. Moreover, the tool was helpful for support in correct punctuation in complex sentences.

I declare that the above log provides a complete and accurate overview of my use of AI tools while writing my thesis.