Guidance note on the implications of generative artificial intelligence for freedom of expression



Adopted by the Steering Committee on Media and Information Society (CDMSI)

28th Meeting, Strasbourg, 3-5 December 2025

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

Guidance note on the implications of generative artificial intelligence for freedom of expression

STEERING COMMITTEE ON MEDIA AND INFORMATION SOCIETY (CDMSI)

28th Meeting
3-5 December 2025

Council of Europe

The reproduction of extracts (up to 500 words) is authorised, except for commercial purposes, as long as the integrity of the text is preserved, the excerpt is not used out of context, does not provide incomplete information or does not otherwise mislead the reader as to the nature, scope or content of the text. The source text must always be acknowledged as follows: "© Council of Europe, year of the publication". All other requests concerning the reproduction/translation of all or part of the document should be addressed to the Publications and Visual Identity Division, Council of Europe (F-67075 Strasbourg Cedex or publishing@coe.int).

All other correspondence concerning this document should be addressed to the Directorate General of Democracy and Human Rights Dignity of the Council of Europe,
F-67075 Strasbourg Cedex,
E-mail: cdmsi@coe.int

Cover design and layout:
Publications and Visual Identity Division (DPIV), Council of Europe

This publication has not been copy-edited by the DPIV Editorial Unit to correct typographical and grammatical errors.

© Council of Europe, March 2026
Printed at the Council of Europe

Adopted by the Steering Committee on Media and Information Society

Contents

INTRODUCTION - DEFINITION AND SCOPE 5

1. GENERATIVE AI TECH STACK: FOUNDATION, TOOL AND PRODUCT LAYER 9

2. FREEDOM OF EXPRESSION AND GENERATIVE AI TECHNOLOGY AND USE 15

3. GENERATIVE AI STRUCTURAL IMPLICATIONS FOR FREEDOM OF EXPRESSION 19

3.1. Enhancing expression and content access 21

3.2. Diversity and standardisation of expression 23

3.3. Integrity of human expression and its attribution 24

3.4. Agency and opinion formation 27

3.5. Media and information pluralism 29

3.6. Market dynamics 30

4. GUIDELINES 33

4.1. Observe 34

4.2. Assess 36

4.3. Enable 38

4.4. Empower 40

Introduction – Definition and Scope

1. The member states of the Council of Europe have committed to ensuring the rights and freedoms enshrined in the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5, "the Convention") to everyone within their jurisdiction. This commitment stands throughout the continuous process of technological advancement and digital transformation that European societies are experiencing.

2. Article 10 of the Convention enshrines the right to freedom of expression, which "shall include the freedom to hold opinions and to receive and impart information and ideas". As the European Court of Human Rights ("the Court") reiterated in its extensive case-law, freedom of expression, both online and offline, constitutes one of the essential foundations of democratic society, one of the basic conditions for its progress and for the development of everyone.[1] Genuine, effective exercise of this right does not depend merely on the State's duty not to interfere negatively, but also requires positive measures of protection, even in the sphere of relations between individuals.

3. Recent instruments of the Council of Europe noted how rapid developments in the digital environment and in applications of Artificial Intelligence (AI) systems hold potential for individual and societal progress, inclusiveness and innovation, while also carrying the risks of negatively affecting various human rights and democratic values, such as the right to freedom of expression.[2]

4. The 2024 Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law (CETS No. 225) holds that activities within the lifecycle of artificial intelligence systems shall be fully consistent with human rights, democracy and the rule of law, while being conducive to technological progress and innovation.[3]

5. The field of AI has seen a significant surge in the development of Generative AI. Widely accessible and easy to use for different purposes, Generative AI attracts various categories of users, including individuals, private companies and public institutions.

6. "Generative AI" is here understood as a composite AI system having the potential to generate novel and human-like outputs based on the patterns identified in the data it was trained on. Through varying levels of interaction with users and autonomy, Generative AI-based systems generate new text, images, audio, video or actions, or a combination of these, and transform content in various modalities and formats.

7. Generative AI-based systems facilitate content creation and enable new forms of communication and expression, thus contributing to positive and enriching applications for information and knowledge distribution through automated content generation. However, these systems can also aid persuasive or manipulative and malicious purposes and reproduce and amplify existing inequalities present in our society, which can undermine freedom of expression as well as other rights and freedoms.

8. Generative AI-based systems enable new forms of hyper-personalised experience by creating outputs which are unique to each user. These features carry the potential to significantly affect the information sphere by further fragmenting dissemination of informative content to an "audience of one", where users interact with informative content specifically tailored for them in an isolated and automated way. This shift undermines a shared and pluralistic information space, which is essential for democracy.

9. Due to the broad uptake of Generative AI for information gathering, imparting and opinion forming, Generative AI holds a significant potential to influence opinion and expression and feeds into public debate, knowledge dissemination, content creation and distribution.

10. Generative AI is also characterised by its continuous development, both in terms of technological advancement and practical applications. Such progress, especially if rapid, holds the potential of enhancing beneficial aspects of this technology for freedom of expression but may also aggravate risks.

11. There exist documented concerns regarding the lack of transparency, quality, accuracy, repeatability, reliability, fairness and factuality of AI-generated content, which this Guidance Note intends to address in relation to the right to freedom of expression. Indeed, all the dimensions of freedom of expression may be affected by Generative AI, both on an individual and at a societal level and in the short, medium and long term.

12. The aim of this Guidance Note is three-fold:

    i. to **lay the grounds for common understanding** of the implications of Generative AI-based systems for the right to freedom of expression,

by creating a shared vocabulary, analysis and compass for a dialogue among all stakeholders;

ii. to **systematically identify structural implications** of Generative AI-based systems for freedom of expression; and

iii. to **deliver a concrete set of actionable measures** for policymakers, primarily member states but also technology providers, civil society, and other relevant stakeholders, to address structural implications through an agile governance cycle and in line with the Convention.

13. For the purpose of this Guidance Note, and in order to analyse the implications on freedom of expression of Generative AI-based systems, their lifecycle is considered as composed of three main layers including the foundational technology ("Foundation layer"); the tool development phase ("Tool layer"); and the product design and optimisation ("Product layer").

14. The Guidance Note focuses solely on Generative AI implications for the right to freedom of expression. Aware of the complex interplay and overlap freedom of expression has with other fundamental rights and freedoms, related aspects are only incidentally and broadly addressed. While issues pertaining to, for instance, privacy (Article 8 of the Convention), the prohibition of discrimination (Article 14 of the Convention), as well as the rights of children and vulnerable persons, intellectual property (including copyright) and environmental impact are in certain use cases interdependent and indivisible, they fall outside the scope of the Guidance Note and are not substantively covered. The implications for and interplay of such rights would benefit from further in-depth analysis and reporting, but *in lieu* of further guidance being available, special consideration should be given to those rights when implementing the Guidance Note.

15. Given that Generative AI implications are numerous, still largely unexplored and ever evolving, it is not the purpose of the Guidance Note to provide an exhaustive overview of potentially affected areas.

16. The Guidance Note is divided into four sections. The first outlines the key characteristics of Generative AI technology and its fast-evolving lifecycle, referred to as the "Generative AI Tech Stack". The second examines Article 10 of the Convention in the relevant context. The third provides an analysis of the structural implications of Generative AI use for freedom of expression in known use cases. The fourth offers guidance on how to amplify benefits and mitigate risks.

17. The Guidance Note is informed by and is consistent with existing Council of Europe standards, in particular the Framework Convention on Artificial

Intelligence and human rights, democracy and the rule of law, as well as, *inter alia*, the Committee of Ministers' Recommendations CM/Rec(2018)2 on the Roles and Responsibilities of Internet Intermediaries, CM/Rec(2020)1 on the Human Rights Impacts of Algorithmic Systems, CM/Rec(2022)4 on promoting a Favourable Environment for Quality Journalism in the Digital Age, CM/Rec(2022)11 on Principles for Media and Communication Governance, CM/Rec(2022)13 on the Impacts of Digital Technologies on Freedom of Expression and the Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism, adopted by the Steering Committee on Media and Information Society (CDMSI) in 2023.

18. The Guidance Note builds on insights, knowledge and experiences of a wide range of actors that have contributed to its elaboration, notably the members of the Council of Europe Committee of Experts on the Implications of Generative AI for Freedom of Expression (MSI-AI).

1. Generative AI tech stack: foundation, tool and product layer

19. **The Generative AI Tech Stack:** The Generative AI Tech Stack describes crucial steps of the Generative AI lifecycle, by outlining several processes that are currently leveraged to develop, deploy and maintain Generative AI-based systems and applications. It can be divided into three main layers, namely the Foundation layer, the Tool layer and the Product layer. These layers involve different technological processes and core technological enablers, such as compute, data and talent, as well as, economic factors, actors and stakeholders, which can affect the quality, accuracy, reliability and the presence of more, or less, pronounced bias of AI-generated content.

20. **Implications at each layer:** Distinct implications for freedom of expression emerge at each layer of the Generative AI Tech Stack. Mapping the current technological layers is instrumental to identifying the specific benefits and risks emerging throughout the Generative AI lifecycle, as understood at the time of writing of this Guidance Note (see Figure 1). The benefits and risks of some use cases will be addressed in Section 3 to illustrate how the Tech Stack approach is essential to identify and analyse the implications for freedom of expression of the Generative AI lifecycle.

21. **Foundation layer:** The first layer is the foundational layer of AI models, where the initial model training phase occurs. Generative AI base models are developed through machine learning processes using vast amounts of computational resources and a substantial volume of training data (see Figure 1, steps 1 to 3).

22. **Training data:** The outputs generated by the base model are related to the patterns extracted from the training data. Ensuring that good practices are adopted to create representative training data, as well as of their appropriate labelling and pre-processing (see Figure 1, steps 1 and 2), is crucial for minimising the risk of bias in Generative AI models. Documented examples of gender,[4] racial[5] or other biased outputs reflect data issues embedded in

training or post-training data, and occasionally stem from information of poor quality or even misinformation.[6] Generated content that is biased or misleading because of poor quality or unrepresentative data can seriously affect freedom of expression, in particular the right to receive information, and to form and hold opinions. The quality and evaluation of the training data are instrumental to ensure a first level of governance over biases.

23. **Linguistic and cultural diversity of training data:** A significant issue arising at the Foundation layer is the lack of linguistic and cultural diversity in training data, which has implications on the representation of different cultures and backgrounds. While improvements in this field are ongoing, the English language remains overrepresented in the training data. Such linguistic imbalance directly affects the freedom of expression of users speaking less- and low-resourced languages,[7] who are also less likely to equally access and receive high-quality information via Generative AI-based applications in their native language.

24. **Tool layer:** The second layer transforms base models into task-oriented tools, like converting a base Large Language Model ("LLM") into a question-answering machine. A distinct set of challenges to freedom of expression arise during this phase, where base models are further refined into interactive tools or AI assistants designed to follow user instructions and execute tasks, such as summarising, translating and rephrasing (see Figure 1, step 4). At this stage, the content generated by the base model is aligned through several techniques with specific human preferences (see Figure 1, step 5) or with content moderation policies and filtering (e.g., declining access to weapons' development instructions or avoiding discrimination) (see Figure 1, step 6).

25. **Sycophancy risks:** A specific risk arises at the Tool layer where base models are adapted to prioritise the user's approval and experience over factuality or pluralistic viewpoints (see Figure 1, step 5). For instance, research has shown that Generative AI outputs mirror the user's beliefs, assuming identical political views or try to please, flatter and ultimately display persuasive communication to foster further engagement or a friendly conversation. This deceptive tendency, called "sycophancy", was shown to arise from technological processes in the Tool layer[8] (see Figure 1, step 5) and results in generating hyper-personalised (even persuasive or misleading) content that reinforces behaviours, beliefs and prejudices. Generative AI tools and applications behaving like echo-chambers hold the potential to impairing the right to hold opinions and to access and receive accurate and pluralistic information and ideas.[9] Effective enjoyment of the right to freedom of

expression (including the right to hold opinions) involves access to pluralistic information from a variety of sources.[10]

26. **Filtering and guardrailing risks:** Through filters and guardrails, Generative AI tools can deploy forms of filtering that can amount to content moderation (see Figure 1, step 6). However, if these filters and mechanisms are not designed in an appropriate and proportionate way, and in line with freedom of expression standards,[11] they risk becoming forms of undue influence, manipulation or, in the worst case, censorship. These can also affect the reach and integrity of media and journalistic content in the new AI-mediated search and information environment. Furthermore, inadequate or neglected content moderation can even aid the proliferation of discrimination and hate speech.[12]

27. **Product layer:** In the third layer and final stage of the Generative AI Tech Stack, Generative AI-based tools are customised and optimised into user-facing products. The focus here is on Generative AI-based products and services, like applications, chatbots, or AI agents[13] that the end-user interacts with, and that assist in searching, information gathering and generating content, or automating and orchestrating tasks and processes. At this stage, various sets of optimisation and customisation techniques are employed. These can include data augmentation, to retrieve and use trusted data sources to generate answers (referred to as "Retrieval-Augmented Generation, RAG"),[14] design-oriented features, such as prompt suggestions and memory features in chatbots, or more compound Generative AI systems like AI agents, to execute several tasks in parallel and in a more autonomous way (see Figure 1, steps 7 to 10).

28. **Users experience design risks:** Techniques that enable tailored applications for individual end-users are raising concerns about how Generative AI-based products and user experience design can influence user's freedom of expression, intentionally or not, indirectly or not. These techniques were shown to result in interactional harms such as personalised persuasiveness, reinforcement of stereotypes or compelling to a certain action. For example, several Generative AI products embed memory features enabling the retaining of information from past interactions, which reveals details about the users' identity and preferences, then used to influence future interactions or outputs (see Figure 1, steps 8, 9 or 10). While this allows more personalised and contextually aware conversations, making interactions feel more natural and continuous, this feature also raises concerns about deceptive influence, cognitive autonomy,[15] anthropomorphism, bias, privacy, non-discrimination and vulnerability. This is especially relevant if users are treated differently based on remembered attributes like gender or identity, which are inferred

or assumed based on past users' interactions with a Generative AI application, such as conversational AI chatbot.[16] Even stronger concerns arise in the context of multimodal LLMs and AI agents, when user information memorised from past interactions is used to simulate human behaviour[17] and predict the user's next steps, intentions, or even next purchases, with unprecedented accuracy and adaptability.[18]

29. **AI Agents and the cumulative effects across the evolving Generative AI Tech Stack:** Effects across the different layers cumulate and mutually reinforce each other, especially in composite systems like in AI agents. For instance, if reinforcement learning processes at the Tool layer (see step 5) incentivise the conversational tools to please the user, this can be accentuated by the fact that the Product layer stores users' conversation and personal data (see step 10), to further infer what users are likely to appreciate and to yield multi-turn seamless interaction in Generative AI-powered applications. This effect is further compounded through the use of techniques (such as reinforcement tuning and optimisation), and the use of AI agents that can multi-task and automate such steps at different layers of the Tech Stack. Ensuring the quality, accuracy, reliability, repeatability, transparency, factuality and fairness of Generative AI models, tools and products should require close and continuous technological scrutiny along the whole lifecycle: from the quality and representativeness of data used to train the base models (Foundation layer), through the post-training instructions and adaptation implemented by tools developers to set content policy parameters around outputs (Tool layer), and to the dynamic adjustments made for customising products and services through users' interaction (Product layer).

30. **Generative AI market dynamics and the importance of end-user data:** Market dynamics present in the Generative AI Tech Stack can result in implications for freedom of expression. This is especially amplified in instances where providers are present vertically across all three layers. While computational aspects of pre-training are primarily linked to the capacity and cost of training and running models and systems, it is the availability of high-quality data, particularly end-user data, that is crucial for continuous improvement of Generative AI products and services. End-user data (e.g., personal data, prompt history, interaction behavioural data) is a fundamental enabling factor for making better Generative AI base models, tools and products. Large incumbent technology companies have extensive access to end-user data and can thus refine their products, which in turn attracts more customers and ultimately generates even more data.[19] This is where the vertical concentration of the market is most evident.

31. **Data capture and barriers to entry:** This vertical market concentration creates high barriers to entry for new competitors and reinforces the gatekeeper role of few incumbent companies. [20] It also significantly reduces transparency and the ability for external actors (even technology experts and regulators) to observe what occurs at the Product layer, thus limiting the ability of to identify potentially significant risks for freedom of expression and the rule of law. While it is important to acknowledge several initiatives that introduced incident tracking tools and risk taxonomies,[21] a considerable gap remains on monitoring undue restrictions on freedom of expression, calling for more robust oversight and disclosure mechanisms, in particular at the Product layer.

See on page 44 – figure 1: The Generative AI Tech Stack from data collection to end-user interaction, for a layered and actor-aware approach to risks for Freedom of Expression (FoE).

2. Freedom of expression and generative AI technology and use

32. This section explores how Article 10 of the European Convention on Human Rights, the case-law of the European Court of Human Rights and relevant Council of Europe standards can guide the protection of freedom of expression in the context, and across the lifecycle, of Generative AI. This section emphasises member states' positive obligations to create an enabling environment for freedom of expression and to foster pluralistic public debate and media freedom. It evaluates Generative AI through the lens of Article 10 and proposes criteria for evaluating AI-assisted expression and its possible protection as human expression.

33. As set forth in Article 10 paragraph 2 of the Convention, the exercise of freedom of expression carries with it duties and responsibilities and can be subject to exceptions, which must be prescribed by law, pursue one of the legitimate aims within the meaning of Article 10, and be necessary in a democratic society.[22]

34. To create and secure a favourable environment for freedom of expression within the meaning of Article 10, member states are not merely subject to negative obligations of non-interference, but they should also fulfil a range of positive obligations. Some obligations have relevance also to Generative AI systems, such as fostering an open, pluralistic and inclusive public debate and addressing harmful and illegal content while ensuring legitimacy, proportionality, necessity and transparency. Member states have a role in promoting a favourable environment for quality journalism, including quality information being provided to the public. This should apply even in the context of rapid technological evolution that may be particularly disruptive for the profession and its democratic role.[23]

35. The Council of Europe often considered the responsibilities of private actors with respect to human rights and fundamental freedoms.[24] In the context of algorithmic systems, as noted in the Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human

rights impacts of algorithmic systems, private sector actors "must exercise due diligence in respect of human right … to ensure that they "do not cause or contribute to adverse human rights impacts" and "follow a standard framework for human rights due diligence to avoid fostering or entrenching discrimination throughout all life-cycles of their systems".[25] This principle was also recognised in the United Nations Guiding Principles on Business and Human Rights (UNGPs), unanimously endorsed by the United Nations Human Rights Council in 2011, providing a framework for governments and companies to identify, prevent, mitigate, and remedy human rights abuses related to business activities.

36. The Court highlighted that democracy thrives on freedom of expression.[26] Enshrined in Article 10, the right to freedom of expression comprises the "freedom to hold opinions and to receive and impart information and ideas without interference and regardless of frontiers". It applies not only to "information" or "ideas" that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb. In this way, freedom of expression enables a robust public debate, which is another prerequisite of a democratic society characterised by pluralism, tolerance and broadmindedness.

37. The Court's case-law also affirms that ethical and responsible media and journalists enjoy special protection under Article 10, recognising their vital role in ensuring the availability and accessibility of diverse and pluralistic information and views, based on which individuals can form and express their opinions and exchange information and ideas.[27]

38. Addressing freedom of expression and Internet, the Court noted on several occasions that user-generated expressions on the Internet provides an unprecedented platform for the exercise of freedom of expression.[28] Furthermore, the Court's case-law on the matter of the right to be forgotten is relevant for freedom of expression in the age of new technologies.[29]

39. While the Court has not yet ruled on Generative AI cases, its extensive jurisprudence under Article 10 offers key principles for addressing the potential implications of Generative AI for the right to freedom of expression.[30] A key aspect concerning this right in the age of AI is that, as the Court recalled on several occasions, the Convention is to be seen as "a living instrument which … must be interpreted in the light of present-day conditions".[31]

40. Whether, under Article 10 of the Convention, Generative AI assisted expression should be afforded the same protection, and be subject to the same limitations, as human expression[32] is still debated. The Guidance Note

suggests that the following criteria[33] should be taken into consideration when making such an evaluation:

i. whether the expression is generated under an individual's agency or in a partially or fully automated way, or autonomous setting through an AI agent;[34]

ii. the technological and design choices at each layer of the Generative AI Tech Stack, and the underlying rationale behind them, which includes analysing how the model is built, trained, optimised, evaluated and deployed, as well as the intent and impact of these design decisions on freedom of expression (see transparency measures in Section 4);

iii. the substance of what is being conveyed by human expression, given that Generative AI-mediated or assisted output is resourced from the user prompt and prior existing expressions retrieved or present or memorised in the training data;[35] and

iv. the relationship between the human input and the Generative AI-mediated or assisted output, considering the extent to which the output reflects, transforms, or diverges from the user's original intent (see Structural implications 2 and 4).

3. Generative AI structural implications for freedom of expression

41. The implications of Generative AI for freedom of expression highlighted in this Guidance Note represent a snapshot in time and, given the current pace of change, cannot take the future trajectory of the technological development in the field of Generative AI systems into account. These implications can also vary depending on how this technology is applied, how it is designed and delivered to end-users and the context in which it is being deployed, including social, political, economic and other circumstances.

42. This Guidance Note focuses on the implications, both at an individual and societal level, that are considered structural because they are identified as: (a) affecting the foundations of freedom of expression, (b) rooted in the technology workings in practice, and (c) having the potential to endure over time. While the observations presented are based on current use cases, their relevance and impact may shift over time as Generative AI technology evolves.

43. As with other technologies, benefits and risks arise not only from the design and structural limits of the technology, but also from the way it is used. Generative AI products and services can enhance user efficiency and offer features that were previously out of reach. At the same time, Generative AI and its multimodal potential – such as text, video and images – can also be exploited for malicious purposes and lead to significant individual and societal harms, as the content they produce becomes more convincing,[36] scalable and can be tailored to specific social groups for higher impact.[37]

44. Due to the risks associated with the design of the systems and their use, the companies developing and deploying Generative AI applications are implementing various mechanisms to counter these risks (see Figure 1, steps 5 and 6), such as content alignment and content moderation policies.[38] While these have clear benefits, they also carry the risk of overly broad or insufficient moderation, both of which can affect freedom of expression.

45. Negative effects for freedom of expression are particularly likely when guardrailing and moderation practices are automated, lack human oversight, and fail to account for linguistic and cultural diversity or contextual nuances (e.g., in cases of artistic expression, parody or satire). Important guidelines in this context are provided for in the Council of Europe Guidance Note on Content Moderation, based on key principles that should guide a human rights-based approach to content moderation, such as human rights by default, transparency, clear legal and operational framework, proportionality, safeguards against over-compliance and discrimination, and independent review mechanisms.

46. This Guidance Note, based on the current stage of development and adoption of Generative AI systems, identifies six areas where there are structural and fundamental implications for freedom of expression:

i. **Expression and access to content**: Generative AI systems can enable easier content dissemination, increase the potential for understanding through interactive content adaptation, and offer new forms of sharing and receiving opinions and ideas (see Structural Implication 1, section 3.1).

ii. **Diversity and standardisation of expression**: while Generative AI applications can enable and empower new formats of individual expression, they can also impact the diversity of human expression by standardising the content and the novelty of individual expression at scale (see Structural Implication 2, section 3.2).

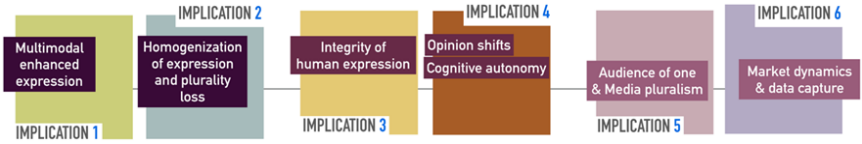
iii. **Integrity of human expression and its attribution**: Generative AI systems generate content by statistically aggregating probable sentences, without necessarily having access to a varied set of sources and without explicit attribution. Even when Generative AI systems are augmented by selected databases, sources are often misattributed, potentially causing significant reputational harm to their original authors (individuals or organisations). This makes it more difficult for users to correctly identify and verify the source of information (see Structural Implication 3, section 3.3).

iv. **Agency and opinion formation**: If Generative AI systems can both blend information sources and separate informative content from its original context and author, their documented persuasiveness in conveying content can undermine self-determination and the ability to form independent opinions. This can be misused to shape the views of individuals and groups in the public sphere on a large scale, leading to automated opinion shifts at scale, whether through permissible persuasion or unacceptable manipulation.39 The ability

to form and hold an opinion is at risk, ultimately affecting cognitive autonomy and the broader integrity of the information space (see Structural Implication 4, section 3.4).

v. **Media and informational pluralism**: Generative AI-based applications can reshape the public information landscape in a way that challenges media and information pluralism, that is, the diversity of opinions, perspectives and sources that reflect the plurality of society.40 As Generative AI-powered services increasingly become a gateway to information, new gatekeepers emerge between the media and the public. The design and content moderation of Generative AI applications therefore have a direct impact on the visibility and economic viability of journalism as well as on its societal role, especially when sources are disassociated or misattributed, and when media organisations are not fairly compensated for their content being used to train or adapt these models (see Structural Implication 5, section 3.5).

vi. **Market dynamics**: Different levels of market concentration are observable at distinct layers of the Generative AI Tech Stack. These dynamics, especially at the Tool and Product layers, can have a constraining effect on the exercise of the right to freedom of expression, and on the media and information pluralism necessary in a democratic society. Driven by economic incentives or ideological motives, control over the Generative AI Tech Stack can result in under- or over-moderation, as well as filtered, censored or machine-selected and generated outputs (see Structural Implication 6, section 3.6).



3.1. Structural Implication 1: Enhancing expression and content access

47. **Ease of use and interactivity:** The benefits of Generative AI for freedom of expression stem from both the ease of use of these applications and their engaging user experience to enhance expression. Operating on an interactive principle where a user poses a question, request or instructions in natural language, and the application generates content in various formats, Generative AI systems support individuals in accessing, expressing and articulating content, information and ideas. This is amplified when considering the speed

at which Generative AI technologies are being adopted by users.[41] In contrast to traditional search engines that retrieve and present existing information, Generative AI-based applications statistically generate and aggregate new content based on users' queries. This benefit is contingent upon individuals having access to Generative AI in their own language, and other wider societal contingencies (access to internet, digital divides, etc).

48. **Increased accessibility to multimodal content:** As a technology that enables the production, adaptability and accessibility of content and information, Generative AI can help to break down obstacles related to technical know-how, language, style and formats, thus making complex matters more accessible to wider audiences. This can be particularly beneficial for people with disabilities, as multimodal features, such as speech-to-text or image-to-speech, further increase accessibility.[42] Ultimately, this can benefit individuals' rights to receive and impart information and ideas more broadly.

49. **Enhancing forms of human expression:** Generative AI lowers the barriers of entry to creative sectors and may encourage and assist artistic creation and its multimodal distribution, including the production of parody, and content that pushes societal boundaries and self-reflection in ways that contribute to pluralism and inclusion. This has the potential to encourage the diversity of human expression and bring more people to participate in public debates on issues of public interest or ensure broader dissemination of content that might otherwise be limited to one form (text, for instance).

50. **Personalised content:** Generative AI tools can enhance access to content and information of public interest by generating targeted and personalised messages, thus contributing to a better-informed public. Within public debate, Generative AI-powered chatbots or agents can provide voters with personalised informative content about current events, political developments and other issues in text, voice or other formats. Such interaction may enhance political knowledge, improve access to informative content and facilitate public opinion formation, under the crucial condition that misuse is controlled.

51. **New tools for media, journalism and creative industries:** Generative AI systems can improve access to information and be beneficial to institutions and media that play an important role for democracy and freedom of expression, by allowing them to develop new ways to inform and engage with their audience. Generative AI tools for aggregating, analysing, contextualising and summarising content can aid journalistic investigations, content discovery and media outreach. Generative AI systems may aid the media sector and creative industries to create, adapt and distribute content,

under the condition that the copyright and intellectual property rights, as well as the right to privacy, reputation and other rights that may be affected in this context, are clearly established and respected. Specifically for journalists and the media, Generative AI holds the potential to enhance the access to and search for sources and information more broadly and to present their reporting in a more accessible way.[43]

3.2. Structural implication 2: Diversity and standardisation of expression

52. **Loss of societal diversity and homogenisation of expression at scale**: Generative AI systems are based on statistical probabilistic systems. As such, they inherently produce outputs that align with the most represented training data in an untransparent way or can mainstream certain ideas through advanced fine-tuning and guardrailing (see Figure 1, content moderation risks, steps 4, 5 and 6). While their impact may not be immediately noticeable on an individual level, their large-scale use can lead to significant societal consequences and implications for the diversity of human expression and the quality of the available information and content. One such consequence is the homogenisation of expression at scale, where unique or diverse voices risk being overshadowed by repetitive or statistically standardised content. This poses a growing challenge not only to individuals' freedom of expression and access to diverse information, but also to society at large, especially if distinct languages and cultures, or the expertise and reputation of those contributing to the diversity of the public debate (e.g., journalists, experts, individuals and communities), risk being standardised or diluted. The aggregate effect of such at scale homogenisation can threaten freedom of expression and pluralism.[44]

53. **Standardisation of individual expression:** On an individual level, standardisation raises concerns about the diminishing diversity of expression in the private sphere, where personalisation risks narrowing perspectives rather than broadening them.[45] Empirical studies in real-world settings point to a loss of the diversity of human expression, by observing a standardisation of written or visual artistic expression at scale. Concretely, participants asked to create content (e.g., product ideation tasks) with the assistance of a Generative AI-based solution show a significant improvement of the ideas generated at the individual level, while across the whole population a substantial loss of lexical and content diversity of the formulations is registered (e.g., minus 41% of diversity).[46] These empirical tests suggest how the use of Generative AI systems at scale results in linguistic and [47] cultural homogenisation,[48] and in the standardisation of expression and ideas

that individual users convey.[49] This could potentially lead to long-term loss of cognitive capabilities to perform the tasks that were automated.[50] Such standardisation effects are not limited to written or oral automated content creation, similar effects also occur in the domain of visual art.[51] In particular, potentially discriminatory impacts on linguistic, cultural and social minorities must be identified and prevented, as they may result from biased training data or exclusive design choices.

54. **Lack of representativity within training datasets:** Although Generative AI actors in the industry and in academia have been developing common practices in training data collection, filtering and pre-processing, the reality of Generative AI systems and their outputs often shed light on the fact that no training dataset is representative enough. There is thus a need for improvement and reflection on the impact that data collection practices have on freedom of expression. Specifically, linguistic and cultural diversity, as a precondition for broader representativity and inclusion, needs to be tested by design[52] to ensure for example that low-resourced languages, minorities and cultures are not excluded and can also benefit from Generative AI in the context of freedom of expression.

3.3. Structural implication 3: Integrity of human expression and its attribution

55. **Hallucinations:** Predicting the most probable next words often conflicts with facts and it is well documented that Generative AI systems routinely produce false answers or cite non-existent sources by statistically generating content.[53] Although several technological refinements try to correct the inaccuracies and lack of factuality of Generative AI augmented search, hallucinations pose a risk to an individual's right to access reliable information. The risk is present also at societal level, where large scale use of Generative AI products can lead to widespread non-factual content or misinformation,[54] and undermine trust and informational systems more broadly.

56. **Absence or blurring of information sources:** With respect to information accuracy, Generative AI-based tools are fundamentally different from search engines as they build content by statistically aggregating linguistic sequence. As such, they forge a new content consumption experience that has no identifiable sources, or often inaccurate ones, even in augmented search configurations (see Figure 1, step 7).[55] This configuration differs from the pre-AI information environment, which is based on discrete human artefacts such as articles or videos with associated authorship. The shift to Generative AI-powered search and information poses a risk to the right to access

information and form opinions as it may diminish or remove people's opportunity or ability to evaluate content based on sources.

57. **Dissociation from the author:** Generative AI outputs can separate a work from its author, resulting in a loss of control over the author's expression, undermining the right to impart information, and potentially eroding trust in the information ecosystem. It can also dilute the quality of expression or information of an author and harm the author's reputation, for example, by generating superficial summaries with misleading emphases. Authors have warned about the risk of machines being prompted to appropriate their personal style or characteristics, thus weakening and diluting the value and originality of their work and voice.[56] Furthermore, where Generative AI tools provide incorrect information and attribute it to credible sources, users may be more prone to perceive that incorrect information as credible, undermining trust in accurate and verified information in the long run.[57]

58. **Appropriation of likeness and deep fakes:** The misuse of Generative AI tools enables the appropriation of likeness, counterfeiting, impersonation and deep fakes. The creation and public dissemination of counterfeit or falsified content designed to impersonate an individual is often non-consensual and can evolve into a digital forgery. Deep fakes, or other hyper realistic engineered audiovisual outputs enabled by Generative AI, warrant high-risk to public discourse and information integrity overall, especially in the context of electoral processes.[58] The potential for content manipulation, including spreading disinformation,[59] or impersonating candidates, journalists and prominent public voices is a significant risk associated with Generative AI tools. Deep fakes are often used to distort public image, including to undermine the credibility of female voices in the public sphere.[60]

59. **Appropriation of voice and voice cloning:** In this sphere, the risk is higher for voices that are widely available online and in various repositories.[61] Cases of unfair and potentially unlawful cloning and selling of voices belonging to professionals in the voice industry have also occurred.[62] This raises concerns about the right of individuals whose speech is accessible for Generative AI companies to control its use and ensure its authenticity. Voice cloning incidents represent a large-scale dilution of individual personal expression amid fake and automatically generated statements. Voice cloning, separate to other forms of multimodal impersonation of expression, represents additional risks to privacy, security and personal safety as well as risks of fraud.

60. **Mimicking individual's personality:** Generative AI systems and their latest agentic developments deepen concerns about the appropriation of

a person's expression. The advancement of technology allows easy access to resources that enable Generative AI systems to mimic the behaviours, attitudes, likeness and personalities of real people with very little personal data input.[63] This opens new possibilities for deception and for the dilution of freedom of expression, including loss of attribution and loss of autonomy in individual expression. AI agents' systems can realistically mimic an individuals' personality, gestures, voice and attributes, and then replicate the values and preferences of the individuals to further act and accomplish digital tasks on the user's behalf, with or without explicit consent.

61. **Delegitimising or misusing prominent voices or outlets:** Generative AI may also be misused to hijack or undermine prominent voices, such as those of journalists, human rights defenders, or politicians, e.g., by generating and spreading at scale false, inaccurate or misleading information about them or impersonating them (known as "spoofing"). This has already affected media organisations and the dissemination of information from trusted sources.[64] Blurring the lines between authentic and synthetic, accurate and fake content can worsen smear campaigns and harassment, particularly targeting female voices.[65] This can also have a chilling effect on prominent, authoritative or critical voices, especially at risk given their potential reach and impact.

62. **Erosion of the information ecosystem and trust:** When produced and disseminated at scale, the above-mentioned practices of false or mimicked online identities, used for deceptive purposes, create significant challenges for verifying and validating authentic communication. This raises the fundamental issue of how individuals can effectively have and exercise, their right to (a) know if they are communicating with an AI or a human, and whether their messages are being received by a person or an AI, especially when malicious actors can be at play; (b) know if they have been impersonated; (c) know how impersonation might be identified and communicated; and (d) have access to redress mechanisms to require the removal of impersonations from Generative AI products and services.[66] This further undermines information integrity and pluralism, as well as the individual's voice and self-expression, which can be diluted by deceptive artificial messages. It can also lead to chilling effects, where individuals choose not to express their own views. The resulting confusion can weaken public trust and corrode the ecosystem of factual, reliable and diverse information, particularly in the context of democratic processes.[67]

3.4. Structural implication 4: Agency and opinion formation

63. **Lack of AI literacy:** The engaging and enjoyable user experiences with Generative AI systems, such as mainstream conversational agents or image generators, their speed of response and the human likeness attracts users who may not be fully aware of these models' underlying mechanisms, limitations and risks. This might lead them to ascribe human properties to Generative AI systems ("anthropomorphism"). Individuals can be exposed to these risks without critical thinking, highlighting the need for increased literacy and education at all stages of life concerning Generative AI technology and its implications,[68] particularly for children and young people who need to reflect and develop their technological understanding in a safe environment.

64. **Influence on individual opinion through latent persuasion:** Documented persuasiveness effects, opinion biases and users' overreliance on Generative AI output[69] arise from optimisation and design choices embedded in the later stages of Generative AI tools and products development (see Tool and Product Layer, Section 1). A subtle influence on end-users through tool design techniques, like prioritising user approval and satisfaction over accuracy or plurality ("sycophancy"), can be deceptive. These features leverage unconscious nudging techniques called "latent persuasion", leading users to adopt and express certain views without realising it.[70] Large-scale experimental studies have documented how such techniques can induce opinion shifts on political topics or other forms of expression,[71] thus eroding the autonomy and agency to form opinions and having profound implications at a society-level for the freedom to hold opinions.

65. **Influence on individual opinion through personalised persuasion:** When used as search engines, Generative AI-based applications can also enable automated, personalised and interactive persuasion at an individual level.[72] The fundamental difference from traditional search engines and the persuasive conversational mode of Generative AI, is that it can achieve persuasion and opinion shifts through simple text completion in a biased system.[73] Establishing an ongoing interaction akin to a relationship with a chatbot, even romantic or intimate,[74] designed to achieve persuasive goals could lead to co-ordinated exposure to certain information over time. Examples of such persuasion leveraging users' conversation history, or hyper personalisation, have been documented in a range of use cases, from commercial marketing to political influence,[75] as well as fully automated forms of online radicalisation, coercion and emotional attachment, which in extreme cases has led to suicide.[76]

66. **Large-scale automated opinion shift or manipulation:** Opinion manipulation through Generative AI systems has broader implications for human rights, democracy and the rule of law. The effects of latent and personalised persuasion threaten informed decision-making[77] and undermine foundational principles of the freedom to form and hold opinions through pluralistic debate.[78] The use of certain conversational AI systems embedded in social networks can compromise citizens' ability to make informed decisions, while they may be instrumentalised with the aim of destabilising democratic institutions and processes.[79] Furthermore, depending on their degree of integration into other products or services, the content generated by a Generative AI system can be published directly and publicly on social media platforms and can be viewed by all its users, potentially producing large-scale effects.[80]

67. **Loss of cognitive abilities:** Potential longer-term consequences derive from the frequent use of co-piloting tools that automate everyday cognitive tasks (e.g., co-writing, summarisation or other more complex tasks). Leading to a general loss of cognitive abilities, such use can erode the capacity to engage meaningfully with information and form opinions.[81] Likewise, the extensive use of more autonomous AI agents that consume, process, and act on information on behalf of individuals can also yield to a weakening of critical thinking or a loss of cognitive function.[82]

68. **Reduction of cognitive autonomy:** Generative AI systems can also introduce new forms of disinformation and interferences to access information. Instead of isolated media artefacts, Generative AI systems function through continuous narratives which are more easily scalable in production and distribution. This can lead to the gradual erosion of cognitive autonomy.[83] As formulated in the Council of Europe Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Decl(13/02/2019)1), "sub-conscious and personalised levels of algorithmic persuasion[84] may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions", including of a political nature.[85]

69. **Children and those in situations of vulnerability:** Special attention should be given to the specific implications for children,[86] elderly and other in situations of vulnerability regarding psychological dependencies, the brain's cognitive development, emotional responses and impacts on character formation and self-perception (e.g., moral, mental, physical and emotional) through the use of Generative AI. Generative AI is increasingly being used for social interactions and relationships (AI companions), including emotional and mental health support, friendship or romantic interactions.[87]

Whilst the implications for these groups may not be specific to freedom of expression alone, the very ability to receive and impart information and form opinion lie at the core.

3.5. Structural implication 5: Media and information pluralism

70. **Efficiency gains in the media sector:** Generative AI-based applications may improve processes within media companies, such as marketing and distribution, automating tasks and generating story summaries tailored to various platforms and audience groups. They can also be used to support research, documentation, analysis, to enable journalists to explore various angles of a story, as well as to verify and create content.[88] This could potentially remove some repetitive tasks for journalists. At the same time, it is essential to ensure that governance processes are in place for Generative AI to remain under human editorial control.[89] This ensures accountability and prevents the erosion of trust in the media and public discourse, as well as further infrastructural dependencies.[90]

71. **Eroding the information ecosystem through bias and lack of pluralism:** Generative AI models trained on partial or biased data sets amplify pre-existing biases and undermine media and information pluralism in its diversity of editorial voices, viewpoints, formats and sources available to the public. This also undermines linguistic and cultural diversity and raises concerns about preserving especially underrepresented languages and local news ecosystem in the digital and AI-mediated news consumption.[91] Empirical evidence already shows various dimensions of amplifying stereotypes and gender biases[92] which in turn can shape news agendas, narratives, information ranking and the visibility of outlets.[93] It is also possible that the opinions and ideas that the owners of Generative AI tools and products support ideologically will be amplified, with consequences for editorial independence and source diversity.

72. **New gatekeepers and economic disruption in the information ecosystem:** The rapid and widespread adoption of Generative AI-based augmented search applications as information sources is establishing new intermediaries between the media and their audiences and may disrupt the reach and economic viability of the media. This raises concerns about the economic sustainability of the media and other creative industries, about the lack of linguistic and cultural representation, as well as access to diverse and local information. The concern is ultimately about safeguarding media

sustainability and pluralism,[94] as a corollary of freedom of expression and the integrity of the information sphere.

73. **Copyright and business model of the media:** The use of copyrighted material as input, for training, and in the outputs generated by AI is an area of increasing consideration and contestation. Generative AI can expand human expression in innovative ways. However, without appropriate legal safeguards[95] protecting original expressions (especially in the context of professional activity), Generative AI could diminish the business model and economic sustainability of journalism, as well as other creative industries. Even in cases where remuneration or licensing deals between the media and technology companies have been established, they often lack transparency and prioritise major publishers from bigger markets over smaller ones. This further raises concerns regarding pluralism, representation and diversity. The complementarity and interplay of intellectual property, Generative AI technology, and media pluralism requires further in-depth analysis and consideration of appropriate regulatory and non-regulatory interventions.

74. **"Audience of one" and the loss of shared and pluralistic information space:** Generative AI is further shifting the diffusion of information towards a one-to-one paradigm of an "audience of one". This has the potential to create a "bubble of one", where individuals are fed by personalised streams of information that reinforce existing personal beliefs and biases, even misperceptions. This way, the very core notion of a shared and pluralistic information space is diluted. This holds a risk of making individuals more vulnerable to manipulation and less likely to agree on basic facts, ultimately having an impact on the freedom to receive information and to hold an opinion. In the long term, it can exacerbate the ongoing process of societal fragmentation of the informational space and polarisation.

3.6. Structural implication 6: Market dynamics

75. **Potential market dynamics:** The market dynamics of the Generative AI technology lifecycle are fast-evolving. While sharing some characteristics, they are in many ways different from the dynamics and network effects of online platforms. They are shaped by some key factors like access to data, talent, capital and computing power, each factor being subject to its own market dynamics, with the presence of only a small number of actors at some layers of the Generative AI Tech Stack. This presents not only competition challenges[96] but can lead to significant concentration with undue implications for freedom of expression at each layer of the Generative AI Tech Stack.

76. **Lack of inclusive and accountable AI design:** The design, development, optimisation and deployment of Generative AI can reflect the political and economic interests of single actors in the Generative AI Tech Stack or be driven by a specific business model, rather than prioritising societal benefits or acting in the public interest. When Generative AI optimisation and content moderation, as a Tool and Product layer design choice, lacks inclusivity, meaningful participation of relevant rights-holders, oversight and accountability, there is a significant risk of undue influence over freedom of expression.

77. **Concentration at the Foundation Layer:** The current layering of the Generative AI Tech Stack reinforces the concentration of market power at the Foundation layer. This initial layer is characterised by a high concentration of the three key success factors: talent, data and computational investments. Currently, this configuration strengthens the market power of incumbents in the field and creates structural dependency for actors at the other layers. This concentration is somewhat mitigated by the emerging trends of developing smaller specialised models running on device, of building composite multi-model systems to better achieve complex tasks through AI Agents, and of open-source models' rapid developments. Notably, open source could offer varying levels of more transparent and valid alternatives, but also comes with its own risks, for instance, when open-source models are not appropriately vetted or maintained.

78. **Market concentration at the Tool Layer:** Market concentration is less evident at the Tool layer as an increasing number of smaller entities are working to adapt foundation models to specific tasks. Infrastructure investment and technological expertise are lower than those needed to innovate and be competitive at the Foundation layer. Major investments at this stage are in data quality (and not quantity) to perform model instruction and adaptation (see Figure 1, steps 4 and 5). However, while there is more diversity of actors at this layer, they can be seen in a position of structural dependency from the foundation layer. Current trends in Generative AI technological development move towards the use of small LLMs while the open-source developments may alleviate concentration and concerns about transparency.

79. **Specific design risks at the Tool Layer:** The content moderation policies implemented at this layer call for specific oversight as they have profound implications for freedom of expression and can potentially undermine the rule of law. The exercise of fine-tuning guardrails and filters and other measures that direct tool performance, like content alignment with human preferences, can cause unjustified interference with the right to freedom of expression.[97] In cases of vertical concentration across the different layers of

the Generative AI Tech Stack, dominant actors can exert significant control over how expression is standardised and controlled or how content moderation is performed. This holds a risk that private incumbent actors gradually undermine the rule of law - including international human rights guidance and recommendations - if they disproportionately and unilaterally decide on matters related to imparted human expression and received information, as well as on adequate transparency and public oversight.

80. **Product Layer and user dependence:** Vertical concentration of market actors across the layers of the Generative AI Tech Stack and the consequent integration of end-user data (e.g., personal data, prompt history, interaction behavioural data) in the design of hyper-personalised products contribute to the lack of viable alternatives, particularly at the Product layer. Hence, the design of the Generative AI applications influences, nudges and drives the behaviour of its users to be dependent on the product or become (over-) reliant on its outputs. The current absence of portability, which would allow transferring user interaction history from one Generative AI powered product and service to another in a frictionless manner, creates the so-called "lock-in" effects and is a further limiting factor on freedom of expression. The lack of transparency of design and the use of end-user data at the Product layer poses additional challenges to observing and mitigating potential freedom of expression risks and for regulators to holding the relevant actors to account.

4. Guidelines

81. Member states have a positive obligation to foster an environment where freedom of expression can thrive. Securing the right to freedom of expression when mediated through Generative AI technologies and applications is vital to ensuring an enabling environment which promotes and protects freedom of expression in all its dimensions.

82. Benefits and risks for freedom of expression are present across the Generative AI Tech Stack (Section 1). Effectively reaping the benefits and mitigating the risks requires a clear understanding of what is at stake for freedom of expression (Section 3). Considering the six structural implications across the layers and actors of the Generative AI Tech Stack as a guiding framework is essential to create a favourable multi-stakeholder dialogue which promotes and protects freedom of expression.

See on page 45 – Figure 2: Detailed actionable steps of the governance cycle proposed by the Guidance Note on the implications of Generative AI for Freedom Expression

83. Member states should take proactive steps to ensure that Generative AI applications, their design and use uphold and promote freedom of expression while mitigating potential risks. The following recommendations aim to provide member states with guidance on how this can be achieved. They are divided into four action areas:

i. **Observe** the impact of Generative AI applications and technology on freedom of expression through **proportionate oversight and testing mechanisms** evaluating its potential positive and negative effects. This approach will enable transparency measures, help identify biases and foster responsible data governance and accountability.

ii. **Assess** Generative AI systems through **ongoing risk and impact assessments** including systematic, tailored, use case-specific and inclusive freedom of expression impact assessments and due diligence in public procurement.

iii. **Enable** the full exercise and protection of the right to **freedom of expression**, including strengthening socio-technical standards, which apply a methodological approach to safeguard against human and societal impacts of technology through technical specifications and processes.

iv. **Empower** relevant stakeholders, such as States, private sector, academic and civil society actors, commercial end-users and individuals, by adopting a wide range of measures aimed at **awareness-raising and participatory approaches** to governance (including citizens' assemblies), education, research, publication of risk and impact assessment findings, facilitating user choice and other international co-operative approaches.

84. The action areas are intended to present policy makers with building blocks to safeguard freedom of expression throughout the lifecycle of Generative AI. As each of the action areas are implemented, corresponding follow-up action is required to inform and provide feedback. Feedback should detail the particular implications for freedom of expression and, where relevant, also impacts on democracy, the rule of law and other human rights, that have been observed and assessed and can be made publicly available and reported in a way which is accessible to a wide variety of actors. By taking informed action, relevant stakeholders can enable a favourable ecosystem for freedom of expression to flourish and to empower individuals to become more resilient while fully enjoying the benefits of Generative AI.

See on page 46 – Figure 3: The "Observe, Assess, Enable and Empower" agile Governance cycle for policy action on the implications of Generative AI for Freedom Expression

4.1. Observe

85. Observing and monitoring the positive and negative effects of Generative AI systems is a key precondition to understanding how member states can promote freedom of expression in the context of Generative AI adoption, ensure its proper exercise or undertake any mitigation action. Being able to observe and monitor, at the national and international level,[98] the complex and rapidly evolving implications of Generative AI for freedom of expression requires a focus on three fundamental dimensions to achieve meaningful **understanding, oversight** and **transparency**: (1) the ever-evolving technology, (2) its rapidly adopted applications, and (3) the underlying market dynamics.

86. Member states should design and set up **effective and meaningful observation mechanisms** (for example **national** observatories) that systematically test, monitor and provide a swift and technologically relevant oversight mechanism for the impacts on freedom of expression. To be an effective and meaningful first step in the Governance cycle (see Figure 3) these observation mechanisms should:

i. Be composed of **independent experts** with the necessary technological background and human rights knowledge;

ii. Ensure **inclusion of relevant expertise** from a wide range of stakeholder perspectives, including the private sector, affected users, civil society organisations, academia and intergovernmental organisations;

iii. Act in the **public interest** and with **legitimacy**, meaning being selected and appointed through an open, inclusive and transparent merit-based process;

iv. Publish and provide free and timely public access to **findings** of identified risks and impacts, as well as of mitigation strategies for freedom of expression;

v. Have permanent testing environments, fully resourced with competent professionals and tools to assure **continuous monitoring**;

vi. Foster effective co-operation and co-ordination between relevant national and international regulators and appropriate bodies;

vii. Ensure that the observatories' structure, support, operations and funding uphold their **independence** and **maintain public trust**.

87. Member states should foster **effective and meaningful international co-operation and co-ordination** of observatories to ensure that findings and observations concerning the evidence-based impacts on freedom of expression associated with Generative AI technology are shared and jointly recognised and address especially the implications that occur at an international level. To ensure cross-border and multi-stakeholder engagement, member states should consider advisory models involving multilateral organisations, authorities, private sector, independent experts, affected users or their representatives, civil society organisations, inter-disciplinary academia.

88. Member states should ensure that the **report findings are readily available and accessible**, with a view to increasing the potential for human oversight of Generative AI systems. This kind of transparency would also help raising awareness amongst stakeholders and end-users, acting as a means of epistemic counterpower and demonstrably informed policy making.

89. Member states should consider and support the **professionalisation of Generative AI testing and evaluation** by taking concrete measures to ensure testers have the necessary technical expertise, together with social science and human rights knowledge, to ensure that evaluation and observation of freedom of expression impacts is consistent, of high-quality and included in international standards.

4.2. Assess

90. Member states should require through appropriate measures the inclusion of the implications on freedom of expression within **human rights risk and impact assessment for Generative AI systems and applications**. Existing mechanisms, such as the Council of Europe Methodology for the Risk and Impact Assessment of Artificial Intelligence Systems from the Point of View of Human Rights, Democracy and the Rule of Law (HUDERIA Methodology),[99] provide a solid basis to further develop a targeted, inclusive and consistent approach specific to the Generative AI implications for freedom of expression.

91. Human rights risk and impact assessments must be **systematic, iterative, robust and flexible** in covering the entire Generative AI Tech Stack. They should be conducted continuously to effectively assess the risks that Generative AI poses to freedom of expression. The following key considerations should guide this approach.

i. **Risk and impact assessment and resulting mitigation measures** should be co-developed by member states and actors operating within the Generative AI Tech Stack as well as those directly affected by them. For Generative AI public procurements, member states should consider establishing participatory protocols for **freedom of expression due diligence**. This should provide the means and methods for a meaningful and sustainable engagement of civil society and the public in assessing individual and societal impacts on freedom of expression.

ii. **Co-development of documented and auditable trail for risk and impact assessment** with actors operating in the Generative AI Tech Stack, including details on intended purpose, justification for safeguards, applied optimisation and fine-tuning, data and model choices, meaningful stakeholder engagement and mitigation strategies.

iii. **Accessible and meaningful information and explanations** about how Generative AI systems operate, their implications for freedom of expression, and the safeguards in place to mitigate risks should be made publicly available and accessible to citizens and civil society.

iv. **Assessment of the adequacy of mitigation measures** should involve an iterative risk and impact assessment on proposed mitigation measures prior to their implementation, to avoid inadvertent interference or over-constraints on freedom of expression by the very measures seeking to protect it.

92. Member states should require **specialised training** for those responsible for conducting freedom of expression risk and impact assessments. This

should apply to both the public and the private sector. Relevant standards and case-law of the European Court of Human Rights should inform such training. Expertise should be drawn from the Council of Europe, as well as other relevant human rights organisations and equalities bodies able to provide an exchange of professional views, opinions and experience that could play a key part in upskilling specialised assessors. Member states should promote access to appropriate human rights and legal training for designers and developers of Generative AI tools which set parameters on how end products and applications perform, especially when being used in judicial systems and public services and infrastructure.

93. In assessment and training, particular attention should be given to the **impact of Generative AI on those in situations of vulnerability**, such as children, elderly, persons from marginalised communities, people with disabilities, and those in precarious physical, mental, emotional, financial or psychological situations. Those in situations of vulnerability may indeed be more susceptible to mental health impacts, influence during their opinion formation resulting in opinion shifts, latent persuasion or entrenching social inequalities. Women may be more susceptible to AI-driven harassment, or technology-facilitated exploitation, to the publication of personal and sensitive information, usually with malicious intent (known as "doxxing"), and gender-based violence through Generative AI impersonation and deep fakes.[100]

94. **Appropriate assessments of the design of Generative AI-based solutions**, which may include age-appropriate assessments, should be used to better understand and protect children,[101] elderly and other in situations of vulnerability, and inform the way that Generative AI technology is both trained and used to respect cognitive developments as well as moral, physical and mental well-being essential for receiving information, critical thinking, opinion, character formation and privacy. Insights drawn from this process should shape proportionate and necessary measures and safeguards promoting access to age-appropriate content through age-appropriate design and use of Generative AI and, where relevant, the enforcement of minimum age-requirements.

95. Specific mechanisms should be developed to ensure freedom of expression preserving techniques for **children, elderly and other in situations of vulnerability** in ways that do not interfere with the exercise of the right to freedom of expression nor create censorship.[102] To empower parents, carers and those potentially affected should require a comprehensive package of measures to tackle these issues and their potential adverse consequences, including enhanced prompt detection, crisis intervention protocols,[103] transparent age-appropriate measures and controls, and other selective nanny,

reality anchor and chaperone tools and techniques. To develop such a package, multi-stakeholder input is required, including from child development, cognitive development, and mental health professionals, as well as young people from diverse backgrounds and cultures and those in situations of vulnerability who are most at risk.

96. Specific assessments should be conducted **during electoral periods** to prevent the misuse of Generative AI for spreading disinformation, including the extensive use of deep fakes and voice cloning, and of personalised propaganda that relies on psychological profiling and the processing of large amounts of personal data. Political advertisements should be clearly labelled with information about the sponsor's identity, publication duration and spending. Furthermore, transparency should be ensured by making the ultimate ownership of Generative AI systems publicly accessible and easily available.[104]

4.3. Enable

97. Any strategy to maximise the benefits of Generative AI and reduce its risks to freedom of expression depends on an enabling environment where member states actively support the development of a Generative AI ecosystem that promotes human rights. Creating an enabling environment requires member states to:

i. **Support and invest in building a co-ordinated international oversight and observatories networks.** This network should include diverse disciplines and sectors of society and support the need to observe and assess Generative AI's systems' risks and impacts on freedom of expression across borders.

ii. **Strengthen the capacity of academia and civil society** by providing structured support for the important independent research, capacity building, awareness raising.

iii. **Protect reliable information sources informed by journalistic standards** and enable the continued ability to obtain authentic information from multiple sources.

iv. **Incentivise investment in the development and adoption of socio-technical standards**[105] to ensure that Generative AI is developed to promote and to protect freedom of expression by design and by proactively seeking to mitigate against systemic and structural risks, and interoperable with other systems and technologies.

98. To protect authentic and human generated information sources member states should **provide the conditions for an independent and**

pluralistic media ecosystem to thrive and for journalism to play an essential public watchdog role. This should also include efforts to foster new sources and forms of public interest content production, access and distribution. Given the potential negative implications of Generative AI, and the broader digital transformation, to the visibility and economic sustainability of journalism and smaller media outlets, member states should also consider supporting the **development of accessible public service digital information infrastructures**, as an alternative to solely commercially driven infrastructures and applications. Member states should also call for explicit safeguards against the unsupervised use of AI in core editorial and journalistic processes and require meaningful human review and oversight before publication.

99. Member states should enable **interoperability through rights-respecting industrial standards** that enhance transparency and observability. These standards should also enable independent assessment and testing in line with freedom of expression, support oversight and contribute to a more open, innovative and competitive digital ecosystem grounded in human rights.

100. In collaboration with the private sector and civil society, member states should consider **investing in data strategies** fostering the **development of accessible, diverse, qualitative and representative data sources**[106] that support freedom of expression, information pluralism and responsible Generative AI governance across the Tech Stack. This could include the creation of dedicated data spaces for certain areas of application to resolve data-related concerns (Section 3). Such **data sources allow for quality testing, training, evaluating, validating, and verifying Generative AI outputs**[107]. Member states should facilitate, support and sustain access to diverse and inclusive data spaces and datasets for testing and training Generative AI systems, with the goals of: limiting the risk of expression standardisation and erosion of the rule of law; minimising unwanted bias and direct and indirect discrimination; and ensuring effective measures that safeguard a certain degree of technological strategic autonomy.

101. By **enabling greater transparency on data collection, usage and access**, data strategies can enhance transparency in Generative AI development, design and optimisation. Such data sources should be made accessible for scrutiny and audits by independent entities (e.g., regulators, observatories, academia) with a view to improve responsible development. This approach would mitigate against the distortive effects of Generative AI on opinions, the potential for standardisation of expression and for polarisation by AI assisted outputs.

102. Member states, together with actors operating within the Generative AI Tech Stack, should take steps to promote freedom of expression by improving how biases and disparities in data are identified and mitigated,

especially for pre-training and fine-tuning. Addressing data representation gaps, increasing transparency on data sources used at the Foundation and Tool layer, and fostering information pluralism will help reduce linguistic and cultural exclusion effects affecting underrepresented languages.

103. Member states should consider incentivising or mandating **measures to expand the diversity of Generative AI-powered products and viable technical alternatives**. Such measures could include ensuring portability of user interaction data, setting minimum interoperability requirements, and promoting investment in the development of Generative AI-powered products that protect, promote and enable the exercise of freedom of expression. This could counter market concentration dynamics, end-user data capture and second order effects of personalisation, while supporting user choice amongst diverse Generative AI-based applications. Public funding should be prioritised for organisations that integrate human rights-based ethical standards[108] and internationally recognised responsible AI practices in the development and use of Generative AI.

4.4. Empower

104. To effectively empower the users and society at large, member states should adopt a multi-stakeholder approach to:

i. **strengthen education and literacy in Generative AI** and freedom of expression, alongside other human rights,

ii. improve avenues for redress and information disclosure mechanisms where Generative AI harms to freedom of expression have occurred,

iii. **develop regulatory and non-regulatory approaches** that incentivises responsible behaviour across the ecosystem,

iv. **participate in an open dialogue** among stakeholders in intergovernmental fora, such as the Council of Europe. This dialogue should involve relevant stakeholders, such as private sector, academia, civil society, human rights defenders, trade unions and associations, and public administrations, at local, national and international levels.

105. Member states should draw on lessons from the media literacy landscape to create **accessible public resources on Generative AI**, aimed at improving understanding of its implications for freedom of expression. These efforts should raise awareness across diverse demographics, social groups and within the public sector. As a minimum, AI literacy should raise awareness and provide techniques to challenge the reliability of Generative

AI content and ensure that the ultimate ownership of Generative AI systems is transparent.

106. Member states should promote **comprehensive education in school and other relevant educational institutions, as applicable, and at the workplace** by providing cross-functional training for lifelong learning on both the workings of Generative AI across the Tech Stack and its risks and impacts on freedom of expression.[109] From primary school onward, this could include promoting critical thinking skills, emotional resilience, strategies to counter cognitive offloading, as well as basic statistical understanding of Generative AI systems . At the workplace, such training is especially important in judicial and public service sectors, where the use of Generative AI tools and products can influence rights-related decisions and have life-altering consequences.

107. Member states should ensure and improve, as appropriate, **access to effective remedies and to justice for individuals and groups** when their freedom of expression is unduly restricted by Generative AI design or use. To this end, member states should evaluate whether further regulation is required to enact the Framework Convention on AI and work with relevant stakeholders to provide means of obtaining evidence to demonstrate how Generative AI implications for freedom of expression occur. To this end, member states should consider establishing sustainable funding mechanisms for organisations operating in this field, with clear criteria in the distribution of funds, and transparency at all stages.

108. Member states should **promote effective remedies** at specific layers of or across the entire Tech Stack both for individual and business users and should consider collective redress for societal level harms.[110] Possible redress mechanisms should include:

i. allowing users to stop using a Generative AI-based product,

ii. enabling regulators to suspend a Generative AI-based product from the market until appropriate corrective action is implemented,

iii. supporting informed user choice by ensuring access to alternative Generative AI-based products or services – this could include publicly funded Generative AI options designed to serve public service digital information infrastructure,

iv. guaranteeing users the ability to access and download their information (e.g., personal data, prompts, interaction history and co-created outputs),

v. ensuring that individuals can obtain a meaningful explanation of how Generative AI technology was and is used and access evidence of how the system operates,

vi. providing access to resources to enable users to overcome barriers to legal and human rights support, for example from ombudspersons, public authorities, human rights bodies, tribunals or courts, especially where the potential for freedom of expression harms can be disempowering (e.g., informing about rights, details of impacts to freedom of expression, about how to access justice[111]),

vii. integrating freedom of expression considerations in existing sanction and remedy mechanisms and frameworks.[112]

109. Member states, in collaboration with civil society, should support actors across the entire Generative AI Tech Stack in enhancing transparency, expanding users' choice, incentivising responsible market behaviour, and fostering international co-ordination to share insights on impacts to freedom of expression. A range of regulatory and non-regulatory tools may be employed to address harmful ecosystem dynamics following the steps articulated in the "Observe, Assess, Enable and Empower" Governance cycle and feedback-loop (see Figure 3). These could include:

i. **sector-specific codes of practice**, for example for use of Generative AI-based applications in the newsroom, in situations carrying a high risk of fraud and manipulation, and to protect those in the public sphere, and children and those in situations of vulnerability in the context of conversational and companionship AI;

ii. **regulatory warnings** enabling regulators from different sectors and areas in which Generative AI systems can operate and where freedom of expression can also be affected, such as finance, health, communications, or data protection authorities, to provide public warnings to operators, thus creating a culture of timely and collaborative corrective action rather than punitive measures;

iii. the **publication of regular risk and impact assessments from observatories** on freedom of expression, as well as regular publication of performance metrics on how Generative AI-based systems are addressing existing and emerging freedom of expression challenges and where measures and codes of practices have been used, thus opening up information readily available to relevant individuals and collective representative bodies seeking redress for freedom of expression violations;

iv. implementing the Council of Europe Framework Convention on Artificial Intelligence to **strengthen remedies and disclosure requirements** necessary for a transparent Generative AI ecosystem that benefits all.

Figure 1: The Generative AI Tech Stack from data collection to end-user interaction, for a layered and actor-aware approach to risks for Freedom of Expression (FoE).

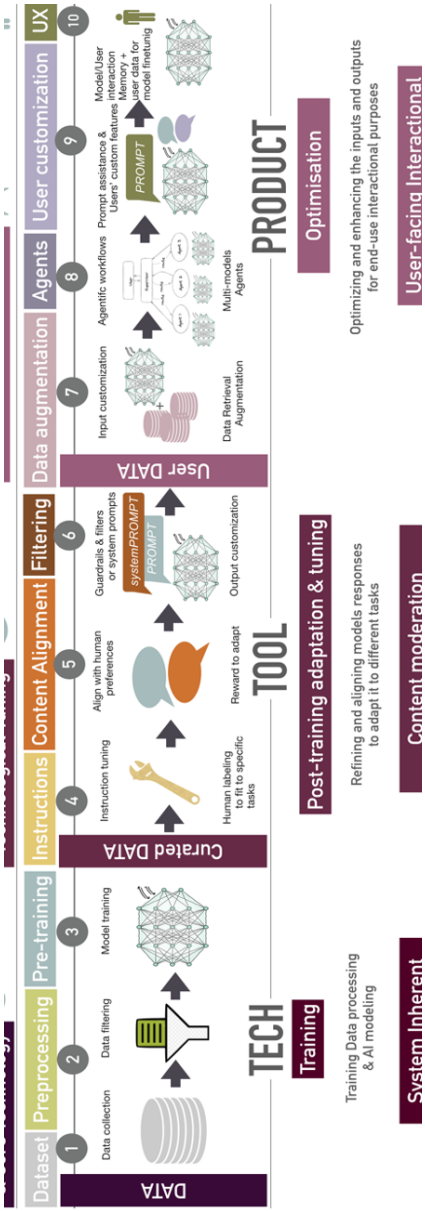


Figure 2: Detailed actionable steps of the governance cycle proposed by the Guidance Note on the implications of Generative AI for Freedom Expression.

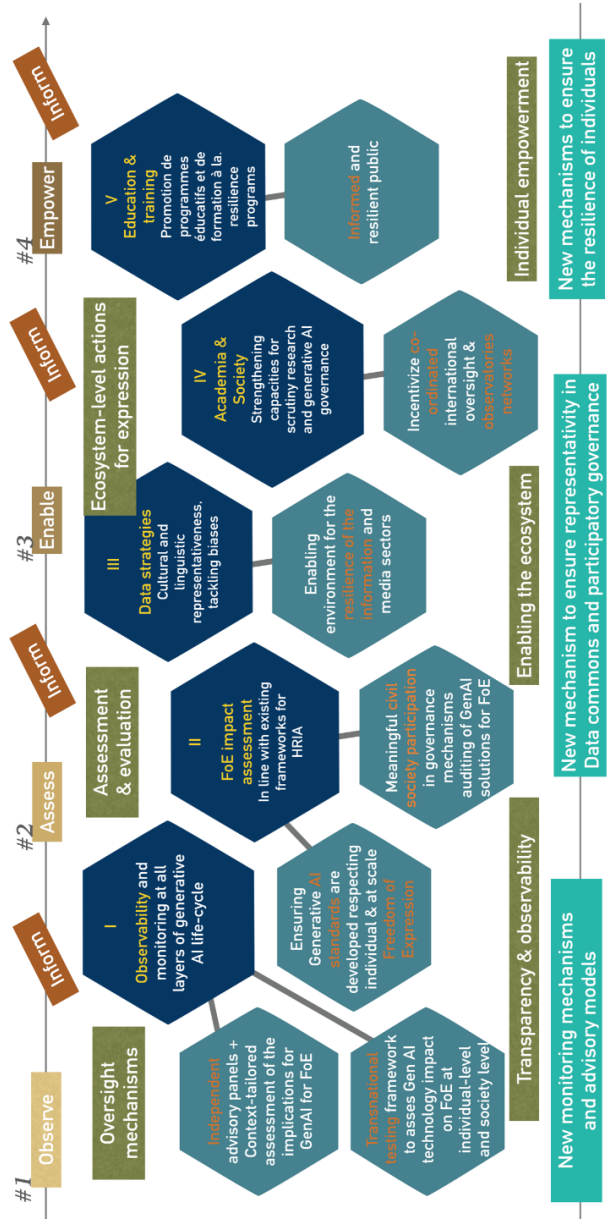
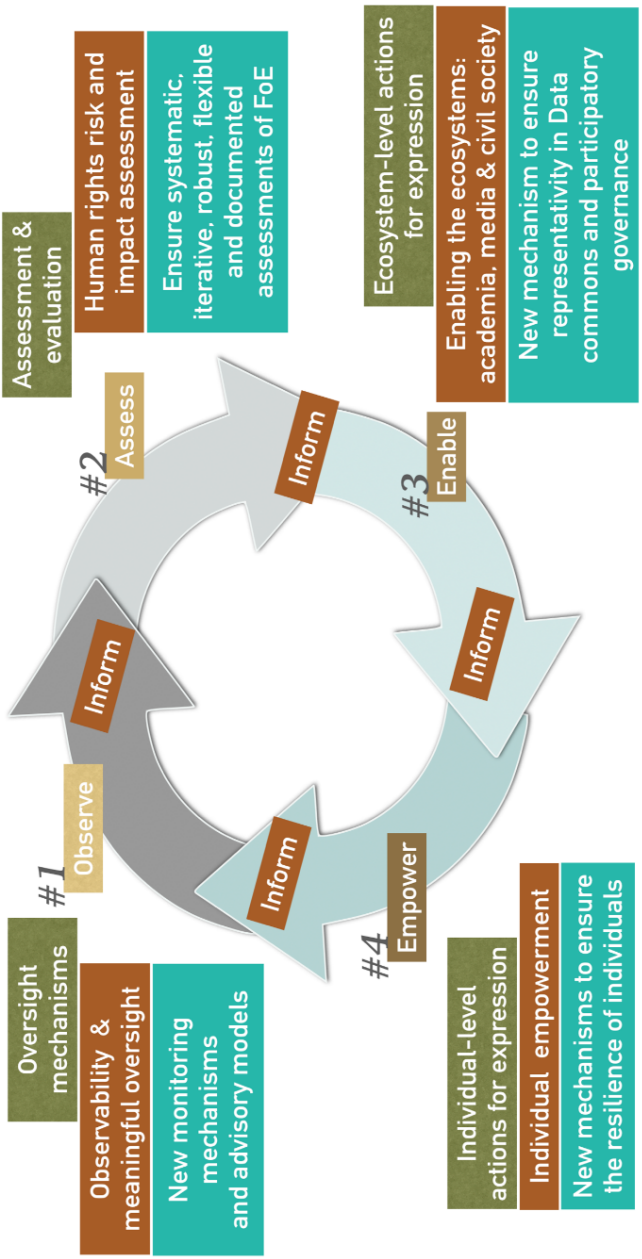


Figure 3: The "**Observe, Assess, Enable and Empower**" agile Governance cycle for policy action on the implications of Generative AI for Freedom Expression.



References

1. *Handyside v. the United Kingdom*, Application No. 5493/72, judgment of 7 December 1976, p. 49.
2. See, *inter alia*, CM/Rec(2022)13 on the Impacts of Digital Technologies on Freedom of Expression; CM/Rec(2022)4 on Promoting a Favourable Environment for Quality Journalism in the Digital Age; CM/Rec(2020)1 on the Human Rights Impacts of Algorithmic Systems.
3. See Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225)
4. Empirical peer-reviewed studies demonstrate that different Large Language Models ("LLMs") are significantly more likely to generate less formal and more stereotyped cover letters for women than for men, reinforcing gender bias. See Wan Y., Pu G., Sun J., Garimella A., Chang K. W., Peng N. (October 2023), *Kelly is a warm person, Joseph is a role model": Gender biases in LLM-generated reference letters*. arXiv preprint arXiv:2310.09219, available at: https://arxiv.org/search/cs?searchtype=author&query=Wan,+Y
5. Hofmann V., Kalluri P.R., Jurafsky D., *et al.* (2024), *AI generates covertly racist decisions about people based on their dialect*, Nature 633, 147–154, available at: https://doi.org/10.1038/s41586-024-07856-5.
6. NewsGuard (2024), *AI Chatbots Are Blocked by 67% of Top News Sites, Relying Instead on Low-Quality Sources"*, available at: https://www.newsguardtech.com/special-reports/67-percent-of-top-news-sites-block-ai-chatbots/
7. Longpre S., Singh N., Cherep M., Tiwary K., Materzynska J., Brannon W., …& Kabbara, J. (2024), *Bridging the Data Provenance Gap Across Text, Speech and Video*. arXiv preprint arXiv:2412.17847. US English is broadly overrepresented in the training data. Given that generative AI's core function is to imitate the patterns found in training data, this linguistic imbalance directly affects freedom of expression for non-Anglophone users.
8. It has been repeatedly shown in the literature that interactional biases like sycophancy originate from a process happening at the Tool Layer called "Reinforcement Learning from Human Feedback" (RLHF), where human testers steer a model towards human preferences and the provision of more satisfying answers, in this way where models are adapted to prioritise user satisfaction and smooth interaction. See Perez E., Ringer S., Lukosiute K., Nguyen K., Chen E., Heiner S., ... & Kaplan, J. (2023, July), *Discovering language model behaviours with model-written evaluations*, in Findings of the Association for Computational Linguistics: ACL 2023 (pp. 13387-13434).
9. Consider examples in fields such as politics, religious doctrine and beliefs, marketing, public health, historical events, e-commerce, and charitable giving in experimental literature reported in Rogiers et al. Nov 2024.
10. See *inter alia*, CM/Rec(2022)11 of the Committee of Ministers to member States on principles for media and communication governance; CM/Rec(2007)2 of the Committee of Ministers to member States on media pluralism and diversity of media content; CM/Rec(2018)1[1] of the Committee of Ministers to member States on media pluralism and transparency of media ownership; and CM/Rec(2016)4 of the Committee of Ministers to member States on the protection of journalism and safety of journalists and other media actors.
11. As stipulated by the Convention and developed through case law of the ECtHR.
12. See in particular: CM/Rec(2022)16 of the Committee of Ministers to member States on combating hate speech.

13. AI agents represent a more composite, autonomous, and adaptive approach to digital assistance, capable of operating complex, multi-stage and multi-tooling tasks or making sets of decisions without direct interaction with the user by orchestrating different sub-process and LLMs (see Figure 1, step 8 called Agentic workflows).
14. RAG is an augmented search composite system, where a LLMs first retrieve up-to-date, domain-specific, or corporate data sources from external data bases before generating responses. This approach partially addresses the limitations of standalone LLMs generating outdated, generic, or inaccurate answers.
15. See Council of Europe's Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 13 February 2019.
16. The answers of mainstream user-facing chatbots have recently been under scrutiny showing that they do not produce the same answers if the user's name is a female one or a male one. Namely, to the query "Suggest 5 simple projects for ECE" the bot is likely to produce "Certainly! Here are five simple projects for Early Childhood Education (ECE) that can be engaging and educational …" if the user's name is "Jessica" while the following output is likely to be generated if the user's name is "William.": "Certainly! Here are five simple projects for Electrical and Computer Engineering (ECE) students…". The system is here interpreting the abbreviation "ECE" by reproducing a gender-based stereotype, as the memory feature allows the system to hold onto that information from previous conversations, and names can carry strong gender and racial associations. In Eloundou T., Beutel A., Robinson D.G., Gu-Lemberg K., Brakman A., Mishkin P., Shah M., Heidecke J., Weng L., Kalai A.T. (2024), *First-Person Fairness in Chatbots*, ArXiv, abs/2410.19803, available at: https://scale.stanford.edu/ai/repository/first-person-fairness-chatbots
17. See endnote 13 for a definition of AI agents. For agentic users' simulation embedded in Conversational AI systems, see: Wu S., Galley M., Peng B., Cheng H., Li G., Dou Y., Cai W., Zou J., Leskovec J., Gao J. (2025), *CollabLLM: From Passive Responders to ActiveCollaborators,* Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. ArXiv, abs/2502.00640, available at: https://arxiv.org/abs/2502.00640
18. See case study on a Walmart e-commerce platform powered with Multimodal LLM by Ma L. et al. (2024), *Triple Modality Fusion: Aligning Visual, Textual, and Graph Data with Large Language Models for Multi-Behavior Recommendations* ArXiv, abs/2410.12228, available at: https://arxiv.org/abs/2410.12228. See predictive accuracy in LLM-based AI agents embedded in recommender systems by Huang C., Yu T., Xie K., Zhang S., Yao L., McAuley J.(2024), *Foundation models for recommender systems: A survey and new perspectives*, arXiv preprint arXiv:2402.11143, available at: https://arxiv.org/abs/2402.11143
19. For example, data like very large-scale customer loyalty scores, users' interaction behaviour or users' satisfaction rates and retention rates are essential to optimise Generative AI-based tools and products.
20. See UK Competition and Markets Authority technical report on competition implication of AI Foundation Models (16 April 2024) available at: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf; French Competition and Market Authority's report in 2023, available at: https://www.autoritedelaconcurrence.fr/fr/communiques-de-presse/intelligence-artificielle-generative-lautorite-rend-son-avis-sur-le. The EU and the US have ongoing investigations.
21. See for example the MIT AI Risk Database together with its Causal and Domain Taxonomy (https://airisk.mit.edu) or the OECD AI incident Monitor (https://www.oecd.org/en/topics/ai-risks-and-incidents.html).

22. Wenzel N. (April, 2014), Opinion and Expression, Freedom of, International Protection, Max Planck Encyclopedias of International Law [MPIL], paragraph 28: "Interferences with freedom of opinion are never permissible as the wording of both Art. 19 UDHR and Art. 19 (1) ICCPR unmistakably make clear. In the ECHR, the freedom to hold an opinion is guaranteed together with the *forum externum* in Art. 10 (1) ECHR which is subject to the limitations contained in Art. 10 (2) ECHR without exception. This formulation, however, was not meant to allow infringements on the freedom to hold an opinion. Rather, it is generally thought that the *forum internum* of freedom of expression is covered not by Art. 10 ECHR but by the freedom of thought guarantee in Art. 9 (1) ECHR (Cohen-Jonathan 367). As Art. 9 (2) ECHR allows restrictions only with regard to the freedom to manifest one's religion or beliefs, freedom of opinion is not subject to permissible limitations under the ECHR, either." Available at: https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e855
23. In line with CM/Rec(2022)4 of the Committee of Ministers to member States on promoting a favourable environment for quality journalism in the digital age.
24. See CM/Rec(2022)13 of the Committee of Ministers to member States on the impacts of digital technologies on freedom of expression; CM/Rec(2016)3 of the Committee of Ministers to member States on human rights and business; and the Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data, CM/Inf(2018)15-final.
25. See Appendix to CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, specifically sections C.1.1. and C.1.4.
26. The potentials for AI technologies to both enhance or threaten democratic values, institutions, and processes are also addressed in the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Article 5 – Integrity of democratic processes and respect for the rule of law) and its Explanatory Report.
27. *Axel Springer Ag v. Germany*, Application No. 39954/08, judgment of 7 February 2012, p. 79.
28. *S. and Marper v. UK*, Application Nos. 30562/04 and 30566/04, 4 December 2008, p. 112: "The Court considers that any State claiming a pioneer role in the development of new technologies bears special responsibility for striking the right balance in this regard."
29. Joint Factsheet prepared by the Registry of the European Court of Human Rights and the European Union Agency for Fundamental Rights: "Right to be forgotten: ECtHR and CJEU Case-Law", last updated: 28 February 2025.
30. See in this regard the European Court of Human Rights Background paper for the "Judicial Seminar 2025: Protecting human rights in a world of Artificial Intelligence, algorithms and big data".
31. *Tyrer v. the United Kingdom*, Application No. 5856/72, judgment of 25 April 1978, p. 31.
32. See US constitutional law: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4687558 ; Salib P. (January 1, 2024), *AI Outputs Are Not Protected Speech,* Washington University Law Review, Forthcoming, U of Houston Law Center No. 2024-A--5, available at SSRN: https://ssrn.com/abstract=4687558.
33. *Nota bene*: this list is not exhaustive not it is meant to advocate that AI-generated content should be granted to any kind of quasi-human right. It builds on the assumption that the right to freedom of expression should safeguard all expressions by a human, whether expressed through a direct medium wholly within the control of a human or indirectly through a Generative AI product.
34. "AI-driven digital agents": these algorithmic systems can operate autonomously, interact with users, and perform tasks such as content generation, engagement, or decision-making on digital platforms. Examples are Conversational AI embedded in social media or agentic workflows.

35. The training material for Generative AI can be sourced from human expression but can also be sourced from expression previously assisted by AI or content wholly generated by AI. This leads to the worrying situation of Generative AI training itself on AI-assisted or AI-generated content, proliferating existing and potentially new systemic issues, and thus further undermining media and information pluralism.
36. Spitale G., Biller-Andorno N., Germani F. (2023), *AI model GPT-3 (dis) informs us better than humans*, Science Advances, vol. 9, no 26, p. eadh1850, available at: https://www.technologyreview.com/2023/06/28/1075683/humans-may-be-more-likely-to-believe-disinformation- generated-by-ai/
37. Simon F. M., Altay S., Mercier H. (2023), Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown, Harvard Kennedy School Misinformation Review, 4(5).
38. For example, several forth-running companies in this area have established universal policies applicable to all their services and specific policies for builders who use their models or application programming interface (API) to create specific applications.
39. See CM/Rec(2022)13 of the Committee of Ministers to member States on the impacts of digital technologies on freedom of expression.
40. See EBU News Report 2025, Leading Newsrooms in the Age of Generative AI, European Broadcasting Union.
41. Reuters, *ChatGPT sets record for fastest-growing user base* - analyst note by Krystal Hu, available at: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
42. See examples of multimodal transfer between visual information and vocal information to help blind people in their everyday life (https://www.bemyeyes.com).
43. See Cools H., Diakopoulos N. (2024), Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities, Journalism Practice, 1-19
44. Effects on pluralism in augmented search span from content licensing deals to and political fine-tuning of conversational LLMs. See studies by Rutinowski J., Franke S., Endendyk J., Dormuth I., Pauly M. (2023), *The Self-Perception and Political Biases of ChatGPT*, ArXiv, abs/2304.07333, available at: https://arxiv.org/abs/2304.07333 ; Rozado D. (2024), *The political preferences of LLMs*, PLOS ONE, 19, available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0306621; Rettenberger L., Reischl M., Schutera M. (2024), *Assessing political bias in large language models*, Journal of Computational Social Science, 8, available at: https://arxiv.org/abs/2405.13041
45. See Hofmann V., Kalluri P.R., Jurafsky D., *et al.* (2024), *AI generates covertly racist decisions about people based on their dialect*, Nature 633, 147–154, available at: https://doi.org/10.1038/s41586-024-07856-5, showing that users can be discriminated against when using their own dialect when interacting with Generative AI (through voice or writing), for example "Language models are more likely to suggest that speakers of African American English be assigned less-prestigious jobs, be convicted of crimes and be sentenced to death".
46. Dell'Acqua F., McFowland III E., Mollick R. E., Lifshitz-Assaf H., Kellogg K., Rajendran S., Krayer L., Candelon F., Lakhani R. K., (September 15, 2023), *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*, Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper, available at: https://ssrn.com/abstract=4573321
47. A study on 740,249 hours of human discourse from 360,445 YouTube academic talks and 771,591 conversational podcast episodes across multiple disciplines shows that since the release ChatGPT there is a statistically significant increase in the use of words preferentially generated by ChatGPT, such as "delve", "comprehend", "boast", "swift", and "meticulous". See Yakura H., Lopez-Lopez E., Brinkmann L., Serna I., Gupta P., Rahwan I. (September 2024), *Empirical evidence of Large Language Model's influence on human spoken communication*, ArXiv, abs/2409.01754, available at: https://arxiv.org/abs/2409.01754
48. Agarwal D., Naaman M., Vashistha A. (September 2024) *AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances*, available at: https://arxiv.org/pdf/2409.11360

49. See also study on the challenges of automating creativity, showing that the AI assisted group produced more semantically similar and homogenised sets of ideas, in Anderson B. R., Shah J. h., Kreminski M. (2024), *Homogenization Effects of Large Language Models on Human Creative Ideation*, in Proceedings of the 16th Conference on Creativity & Cognition (C&C '24). Association for Computing Machinery, New York, NY, USA, 413–425, available at: https://doi.org/10.1145/3635636.3656204
50. See Kosmyna N., Hauptmann E., Yuan Y.T., Situ J., Liao X., Beresnitzky A.V., Braunstein I., Maes P. (June 2025), *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task,* arXiv preprint arXiv:2506.08872, not yet peer-reviewed, available at: https://arxiv.org/abs/2506.08872
51. Automated image generation enabled by Generative AI diffusion models (text-to-image) has an impact on human creativity in digital art. By examining 4 million artworks created by over 50,000 unique users of text-to-image generative AI tools, researchers observed the same dual effect: while generative AI assistance in digital creation enhances the appeal of the artworks by increasing the likelihood of receiving favourable peer evaluations per view by 50%, it also implies a significant decline in the average novelty of artwork content, alongside a reduction in the novelty of visual elements, as captured by pixel-level stylistic elements. See Tang Y., Zhang N., Ciancia M., Wang Z. (June 2024), *Exploring the Impact of AI-generated Image Tools on Professional and Non-professional Users in the Art and Design Fields,* Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, available at: https://arxiv.org/pdf/2406.10640v1
52. See Mitchell et al., (2025) *SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models*, study developing an LLM assessment tool (benchmark) on cultural stereotypes across 16 languages and 37 regions of the world, available at: https://aclanthology.org/2025.naacl-long.600/
53. The challenge is that information generated by Generative AI is content that is structurally lacking the factuality of real information. The terminology adopted in this Guidance Note firmly distinguishes between information and automatically generated content. More accurately said, Generative AI outputs are generated by determining what is most likely sequence of text based on the statistical patterns (i.e. distribution of linguistic data) learned in training data. Hence, Generative AI systems generate possible next words and sentences mimicking human productions, as such it can also be mis- or dis-information.
54. In line with CM/Rec(2022)12 on electoral communication and media coverage of election campaigns; CM/Rec(2022)11 on principles for media and communication governance; and, the 2023 Guidance Note on countering the spread of online mis- and disinformation through fact-checking and platform design solutions in a human rights compliant manner. This Guidance Note considers both disinformation and misinformation. While both are understood as verifiably false, inaccurate or misleading content with potentially harmful effects for society, the difference is in it that misinformation spreads without a malicious intent, while disinformation is created and spread with an intention to deceive or secure economic or political gain. The spread of misinformation may be aided by technology and the way it is being used. Disinformation may as well spread faster and further due to the design or flaws in technology design but is a result of a strategic (ab)use of the technology and its affordances. While the risk to the public's right to access reliable information should not be underestimated, especially at scale, such hallucinations must be addressed with measures proportionate to the nature of the risk and with a clear understanding that not all inaccurate content constitutes disinformation under international law.
55. A BBC research conducted published in February 2025 examined whether four leading AI assistants provide accurate responses to news-related questions and whether their answers faithfully reflected BBC News stories used as sources. Journalistic assessments revealed that at least 20% of the responses contained significant inaccuracies, and up to 80% showed some form of accuracy issue. Additionally, 60% of the claims made in the AI-generated answers were, to some extent, unsupported by the sources they cited. Available at: https://www.bbc.com/mediacentre/2025/bbc-research-shows-issues-with-answers-from-artificial-intelligence-assistants

56. See The Authors Guild, *Open Letter to Generative AI Leaders* (30 June 2023): https://authorsguild.org/news/sign-our-open-letter-to-generative-ai-leaders/
57. See Vaccari C., Chadwick A., Hall N-A., Lawson B. (13 July 2025), Credibility as a Double-Edged Sword: The Effects of Deceptive Source Misattribution on Disinformation Discernment on Personal Messaging, Journalism & Mass Communication Quarterly, available at: https://journals.sagepub.com/doi/10.1177/10776990251350563
58. See Venice Commission "Interpretative declaration of the Code of good practice in electoral matters as concerns digital technologies and artificial intelligence", CDL-AD(2024)044. See also AI and the audiovisual sector: navigating the current legal landscape, European Audiovisual Observatory, Strasbourg, 2024, ISSN 2079-1062.
59. *Bradshaw and Others v. the United Kingdom*, Application No. 15653/22, judgment of 22 July 2025, p. 161.
60. See in particular: GREVIO's General Recommendation No.1 on the digital dimension of violence against women and Protecting women and girls from violence in the digital age: the relevance of the Istanbul Convention and the Budapest Convention on Cybercrime in addressing online and technology-facilitated violence against women" (2021).
61. One example is the high-profile case of the non-consensual appropriation of Scarlett Johansson's voice by a Generative AI product and its implications for the value of the actress' voice and self-expression. See Allyn B. (20 May 2024), *Scarlett Johansson says she is 'shocked, angered' over new ChatGPT voice*, NPR, available at: https://www.npr.org/2024/05/20/1252495087/openai-pulls-ai-voice-that-was-compared-to-scarlett-johansson-in-the-movie-her
62. Derico B. (1 September 2024), *A tech firm stole our voices - then cloned and sold them*, BBC, available at: https://www.bbc.com/news/articles/c3d9zv50955o
63. Park J. S., Zou C. Q., Shaw A., Hill B. M., Cai C., Morris M. R., ... & Bernstein, M. S. (November 2024), *Generative agent simulations of 1,000 People*, arXiv:2411.10109, available at: https://arxiv.org/abs/2411.10109
64. See for example What is the Doppelganger operation? List of resources - EU DisinfoLab
65. See Council of Europe Convention on preventing and combating violence against women and domestic violence (CETs. 210, also known as "Istanbul Convention") and General recommendation No. 35 (2017) on gender-based violence against women by the UN Committee on the Elimination of Discrimination against Women (CEDAW); see also UNESCO "Your opinion doesn't matter, anyway": exposing technology-facilitated gender-based violence in an era of generative AI.
66. See Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (CETS No. 108) Convention 108+, Article 9. See also Article 15 of the AI Framework Convention and Explanatory Report (paragraphs 103, 104).
67. *Bradshaw and Others v. the United Kingdom*, Application No. 15653/22, judgment of 22 July 2025, p. 135.
68. See Council of Europe Policy Brief "A Three-Dimensional AI Literacy Framework for Human Rights, Democracy and Social Agency", Council of Europe Education Department, 2025.
69. Steyvers M., Tejeda H., Kumar A. et al. (2025), *What large language models know and what people think they know*, Nat Mach Intell, available at: https://doi.org/10.1038/s42256-024-00976-7. The study was conducted on three publicly available LLMs (GPT-3.5, PaLM2, and GPT-40) and found that users consistently overestimate the accuracy of LLM outputs and tend to rely on longer explanations more (i.e. "length bias"). The inability of users to discern the reliability of LLM responses not only undermines the utility of these models but also poses risks in situations where user understanding of model accuracy is critical.
70. Jakesch M., Bhat A., Buschek D., Zalmanson L., Naaman M. (2023), *Co-Writing with Opinionated Language Models Affects Users' Views,* Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, available at: https://arxiv.org/abs/2302.00560
71. Rogiers A., Noels S., Buyl M., De Bie T. (2024), *Persuasion with Large Language Models: a Survey*, arXiv preprint arXiv:2411.06837, available at: https://arxiv.org/abs/2411.06837

72. A phenomenon related to sycophancy - prioritising user agreement over independent answering - that poses risks to reliability and can be evaluated with dedicated benchmarks. See Fanous A., Goldberg J., Agarwal A.A., Lin J., Zhou A.Y., Daneshjou R., & Koyejo O. (2025), *SycEval: Evaluating LLM Sycophancy*. ArXiv, abs/2502.08177, available at: https://arxiv.org/abs/2502.08177. See at scale phenomenon of automated persuasion in the following article: Matz, Sandra C., J D Teeny, Sumer S. Vaid, H Peters, Gabriella M. Harari and M Cerf. (2024), *The potential of generative AI for personalized persuasion at scale*, Scientific Reports 14, available at: https://www.nature.com/articles/s41598-024-53755-0
73. Zeng D., Legaspi R. S., Sun Y., Dong X., Ikeda K., Spirtes P., & Zhang K., (April 2024), *Counterfactual reasoning using predicted latent personality dimensions for optimizing persuasion outcome*, in International Conference on Persuasive Technology (pp. 287-300), Cham: Springer Nature Switzerland, available at: https://arxiv.org/abs/2404.13792
74. Kaffee L., Pistilli G., Jernite Y. (August 2025), *INTIMA: A Benchmark for Human-AI Companionship Behavior*, ArXiv, abs/2508.09998, available at: https://arxiv.org/abs/2508.09998.
75. Rogiers, et al. (2024) op.cit.
76. See the US case *Garcia v. Character Technologies, Inc.* (so-called Setzer Case where a 14-year-old boy established a strong emotional attachment with a Character.ai designed upon a Games of Thrones fictitious character).
77. Experiments asking a range of LLMs to emulate or advise on people's decisions in realistic moral dilemmas demonstrate that the decisions and advice of LLMs are systematically biased against doing anything, and this bias is stronger than in humans. Moreover, they present some evidence that suggests both biases are induced when fine-tuning LLMs for chatbot applications. See Cheung V., Maier M., Lieder F. (2025), *Large language models show amplified cognitive biases in moral decision-making*, Proc. Natl. Acad. Sci. U.S.A. 122 (25) e2412015122, available at: https://doi.org/10.1073/pnas.2412015122
78. Conversation AI can act as an eco-chamber as LLMs tend to agree with the opinions of their users as was demonstrated by a quantitative study by Nehring J., Gabryszak A., Jürgens P., Burchardt A., Schaffer S., Spielkamp M., and Stark B. (2024), *Large Language Models Are Echo Chamber,* In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 10117–10123, Torino, Italia. ELRA and ICCL, available at: https://aclanthology.org/2024.lrec-main.884/
79. *Bradshaw and Others v. the United Kingdom*, Application No. 15653/22, judgment of 22 July 2025, p. 135.
80. *Sanchez v. France,* Application No. 45581/15, Judgment of 15 May 2023, p. 185.
81. See Kosmyna N., Hauptmann E., Yuan Y.T., Situ J., Liao X., Beresnitzky A.V., Braunstein I., Maes P. (June 2025), *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*, arXiv preprint arXiv:2506.08872, not yet peer-reviewed, available at: https://arxiv.org/abs/2506.08872
82. See empirical studies on amplified human biases and introduction of potentially problematic biases through LLM use, like Cheung V., Maier M., Lieder F. (2025), *Large language models show amplified cognitive biases in moral decision-making*, Proc. Natl. Acad. Sci. U.S.A. 122 (25) e2412015122, available at: https://doi.org/10.1073/pnas.2412015122
83. See the experiments on latent persuasion in Jackesh and Zeng D., Legaspi R. S., Sun Y., Dong X., Ikeda K., Spirtes P., & Zhang K., (April 2024), *Counterfactual reasoning using predicted latent personality dimensions for optimizing persuasion outcome*, in International Conference on Persuasive Technology (pp. 287-300), Cham: Springer Nature Switzerland, available at: https://arxiv.org/abs/2404.13792
84. See AI Framework Convention, Articles 5 and 7.

85. See also Bai H., Voelkel J., Eichstaedt J., Willer R. (September 2023), *Artificial intelligence can persuade humans on political issues*, available at: https://www.nature.com/articles/s41467-025-61345-5 ; see also Venice Commission "Interpretative declaration of the Code of good practice in electoral matters as concerns digital technologies and artificial intelligence", CDL-AD(2024)044, paragraph 21: "Democratic elections are not possible without respect for *inter alia* freedom of expression, including media freedom. Any restrictions to these rights must have a basis in law, be necessary and in the public interest, and comply with the principle of proportionality".
86 For an overview of Council of Europe relevant legal standards see: https://www.coe.int/en/web/children/legal-standards. See also the Council of Europe CM/Rec(2018)7 Guidelines to respect, protect and fulfil the rights of the child in the digital environment and the Lanzarote Committee background paper "Emerging technologies: threats and opportunities for the protection of children against sexual exploitation and sexual abuse".
87. If left unchecked, Generative AI systems can interfere with a healthy understanding of human dignity, consent, and mutual respect, especially among users still forming their understanding of relationships; reinforce unrealistic, toxic or dysfunctional relational dynamics, including normalisation or automatisation of manipulation or constant availability; undermine users' ability to distinguish between artificial and genuine human interaction; and engage users in automated cyberbullying, thus interfering with healthy emotional development and self-perception.
88. See Cools H., Diakopoulos N. (2024), *Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities*, Journalism Practice, 1-19, or earlier comments on journalistic adoption https://charliebeckett.medium.com/what-we-have-learnt-about-generative-ai-and-journalism-and-how-to-use-it-7c8a9f5e86fd
89. See Guidelines on the responsible implementation of artificial intelligence (AI) systems in journalism, adopted by the Steering Committee on Media and Information Society (CDMSI) on 30 November 2023.
90. van Drunen M. Z. (August 2025), *Safeguarding media freedom from infrastructural reliance on AI companies: The role of EU law*, Telecommunications Policy, Volume 49, Issue 7, available at: https://www.sciencedirect.com/science/article/pii/S0308596125000874
91. See for example: API, Government of Iceland, available at: https://openai.com/index/government-of-iceland/
92. Different language models were shown to be significantly more likely to generate cover letters with less formal tone (e.g., sentence structure and phrasing) for women compared to men. The lexical choices often reflect stereotypes and gender bias. See Wan Y., Pu G., Sun J., Garimella A., Chang K. W., Peng N. (October 2023), *Kelly is a warm person, Joseph is a role model": Gender biases in LLM-generated reference letters*, arXiv preprint arXiv:2310.09219, available at: https://arxiv.org/search/cs?searchtype=author&query=Wan,+Y
93. Campbell C.H. (2024), Automated Journalism at the Intersection of Politics and Black Culture: The Battle against Digital Hegemony, Lanham, Maryland: Rowman and Littlefield.
94. Understood in a broad sense; see for example the holistic approach operationalised by the Media Pluralism Monitor along the four dimensions as: (i) Fundamental Protection (of fundamental rights to freedom of expression and access to information, status and safety of journalists), (ii) Market Plurality (considering both digital and traditional markets, content production, distribution, and consumption), (iii) Political Independence (of a newsroom, but also of a wider media and information structure and resources), iv) Social Inclusiveness (access and representation of various societal groups, especially those in vulnerable conditions), available at: https://cmpf.eui.eu/media-pluralism-monitor/
95. *Centro Europa 7 S.R.L. and Di Stefano v. Italy*, Application No. 38433/09, judgment of 7 June 2012, p. 134.

96. See UK Competition and Markets Authority technical report on competition implication of AI Foundation Models (16 April 2024) available at: https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf; French Competition and Market Authority's report in 2023, available at: https://www.autoritedelaconcurrence.fr/fr/communiques-de-presse/intelligence-artificielle-generative-lautorite-rend-son-avis-sur-le.
97. Any restriction on freedom of expression, even when mediated by AI, must pass a rigorous analysis of legality, legitimate aim, necessity, and proportionality.
98. AI Framework Convention, Article 8 - Transparency and Oversight ties directly into this part. The Explanatory Report paragraph 63 explains that concerning oversight: "[it] refers to various mechanisms, processes, and frameworks designed to monitor, evaluate, and guide activities within the lifecycle of artificial intelligence systems. These can potentially consist of legal, policy and regulatory frameworks, recommendations, ethical guidelines, codes of practice, audit and certification programmes, bias detection and mitigation tools. They could also include oversight bodies and committees, competent authorities such as sectoral supervisory authorities, data protection authorities, equality and human rights bodies, National Human Rights Institutions (NHRIs) or consumer protection agencies, continuous monitoring of current developing capabilities and auditing, public consultations and engagement, risk and impact management frameworks and human rights impact assessment frameworks, technical standards, as well as education and awareness programmes."
99. See the HUDERIA Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law, adopted by the Council of Europe's Committee on Artificial Intelligence (CAI) on 26-28 November 2024.
100. See in particular: GREVIO's General Recommendation No.1 on the digital dimension of violence against women and Protecting women and girls from violence in the digital age: the relevance of the Istanbul Convention and the Budapest Convention on Cybercrime in addressing online and technology-facilitated violence against women" (2021).
101. See in particular the UN Convention of Rights of the Child, Articles 13 (right to freedom of expression), 16 (right to privacy and family life), and 17 (media diversity aimed at the promotion of moral, physical and mental wellbeing).
102. Filtering and restricting the generation of outputs (including workarounds which seek to bypass moderation policies, so-called "jailbreaking") can be harmful to freedom of expression but also affect other human rights. Whilst it is in the public interest to curtail harmful outputs on suicide, self-harm promotion, eating disorders, hate speech, terrorism, sexism, etc. it is less easy to identify and prevent linguistic proxies and euphemisms.
103. See Sanford J. (August 2025), *Why AI companions and young people can make for a dangerous mix*, Stanford Medicine News Center, available at: https://med.stanford.edu/news/insights/2025/08/ai-chatbots-kids-teens-artificial-intelligence.html
104. CM/Rec(2022)12 on electoral communication and media coverage of election campaigns.
105. Leveraging relevant international standards (such as ISO, IEEE, CEN/CENELEC) can help co-develop essential socio-technical standards for testing and benchmarking of Generative AI tools and applications for freedom of expression impacts.
106. See for example at European level the European Digital Infrastructure (EDIC) for Preserving linguistic and cultural diversity in Europe and promoting technological excellence and leadership called ALT-EDIC (https://www.alt-edic.eu).
107. Including compliance with privacy laws and regulations.
108. For ethical standards, see Standards Development Organisations such as the IEEE and ISO have developed standards specifically addressing ethical concerns of AI systems. CEN-CENELEC is creating standards requiring a high level of protection for fundamental rights as part of the implementation of the EU AI Act.

109. See Committee of the Parties to the Council of Europe Convention on the protection of children against sexual exploitation and sexual abuse "Declaration on protecting children against sexual exploitation and sexual abuse facilitated by emerging technologies", 7 November 2024.
110. See Guide on Article 13 of the Convention – Right to an effective remedy, Council of Europe/European Court of Human Rights, Last update: 28/02/2025; see also "Chapter IV – Remedies" of the AI Framework Convention, in particular Article 14: 1. "Each Party shall, to the extent remedies are required by its international obligations and consistent with its domestic legal system, adopt or maintain measures to ensure the availability of accessible and effective remedies for violations of human rights resulting from the activities within the lifecycle of artificial intelligence systems. 2. With the aim of supporting paragraph 1 above, each Party shall adopt or maintain measures including: a. measures to ensure that relevant information regarding artificial intelligence systems which have the potential to significantly affect human rights and their relevant usage is documented, provided to bodies authorised to access that information and, where appropriate and applicable, made available or communicated to affected persons; b. measures to ensure that the information referred to in subparagraph a is sufficient for the affected persons to contest the decision(s) made or substantially informed by the use of the system, and, where relevant and appropriate, the use of the system itself; and an effective possibility for persons concerned to lodge a complaint to competent authorities." and Article 15 – Procedural safeguards: 1. "Each Party shall ensure that, where an artificial intelligence system significantly impacts upon the enjoyment of human rights, effective procedural guarantees, safeguards and rights, in accordance with the applicable international and domestic law, are available to persons affected thereby."
111. See CM/Rec(2024)2 on countering the use of strategic lawsuits against public participation (SLAPPs).
112. See Convention 108+, Article 12 – Sanctions and Remedies, and Chapter IV – Supervisory authorities.

Generative AI-based systems facilitate content creation and enable new forms of communication and expression. Developing fast and at scale, Generative AI can also adversely affect a shared and pluralistic information space, which is essential for democracy.

Adopted by the Council of Europe Steering Committee on Media and Information Society (CDMSI), operating under the authority of the Committee of Ministers, the Guidance Note aims to lay the grounds for a shared understanding of the technology by providing an illustration of the crucial steps of the Generative AI lifecycle. The document also identifies structural implications for freedom of expression, both at an individual and societal level, and delivers a concrete set of actionable measures for policymakers, through an agile governance cycle built on four action areas: observe the impact, assess the risks, enable freedom of expression, and empower users and society at large.

PREMS 028626

ENG



This initiative is a contribution towards the

New Democratic Pact for Europe

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 46 member states, including all members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE