

1556th meeting, 8 April 2026

5. Media

5.1 Steering Committee on Media and Information Society (CDMSI)

a. Explanatory Memorandum to Recommendation CM/Rec(2026)4 of the Committee of Ministers to member States on the online safety and empowerment of users and content creators

Item to be considered by the GR-H at its meeting on 31 March 2026

Preamble

1. The Recommendation aims to provide member States with a coherent and actionable framework to promote the online safety and empowerment of users and content creators.

2. The online environment today serves as a vital public space, enabling the exchange of information and ideas, communication, civic participation, cultural expression and economic activity, to name but a few activities. However, whilst there is no doubt that the online environment is a significant boost to the exercise of the right to freedom of expression as well as to the enjoyment of other rights, it also comes with significant risks, such as harassment, hate speech, exposure to misinformation and disinformation, and algorithmic bias, that can infringe human rights, affect information integrity and deter participation. In many respects, the online environment mirrors everyday life: it offers spaces for learning, creativity and connection, but it also exposes individuals to risks and harms that need to be managed if participation is to remain safe and meaningful.

3. The Recommendation affirms commitment of the Council of Europe to addressing these risks without sacrificing the openness and accessibility of the internet. It stresses that a human rights-based approach, with the principles of legality, necessity and proportionality at its core and safety by empowerment as a central tenet, is essential to ensuring that measures they take for the safety of users and content creators do not inadvertently or disproportionately restrict freedom of expression and other rights.

4. Grounded in the Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, the Convention), the principles set out in the Appendix to the Recommendation build on the case-law of the European Court of Human Rights (the Court) on the enjoyment and exercise of human rights in the online environment. It highlights the need for States to create a free, open, and accessible internet for all, and an enabling online environment where users can enjoy their rights without discrimination. The Court has increasingly addressed the responsibilities of States in this respect, emphasising the obligations of States to protect users from harmful interference by both public authorities and private entities, as well as their obligations to refrain from imposing restrictions on rights other than those that are prescribed by law and necessary in a democratic society in pursuit of a legitimate aim. The Court has equally made clear that private entities may, in certain circumstances, be

held liable for illegal content published by users and has emphasised the duties and responsibilities that come with the exercise of the right to freedom of expression (see, for example, *Delfi AS v. Estonia* [GC], No. 64569/09, 16 June 2015).

5. The Preamble recalls a key principle in the Court's case law: the right to freedom of expression protects not only inoffensive or neutral speech but also speech that may "offend, shock or disturb" (*Handyside v. the United Kingdom*, No. 5493/72, 7 December 1976, para. 49). This principle is particularly salient in the context of online expression, where there is a great diversity of voices and where polarised or provocative speech is common. The Court has long held that robust protection for dissenting or minority views is essential in a democratic society and that this applies both in the physical environment and online, including for user-generated content, journalistic work and political expression (*Delfi AS v. Estonia* [GC], cited above, paras 131-139). While the Court has observed that a certain degree of vulgar abuse is common in many online forums, it has emphasised that this needs to be understood in context and that a certain level of tolerance is expected, especially by public figures such as politicians (see e.g., *Tamiz v. the United Kingdom* (dec.), No. 3877/14, 19 September 2017, para. 81, in which the applicant, a politician, had himself initiated the use of vulgar language). The Recommendation therefore indicates robust protection for freedom of expression as an essential component of any regulation aimed at promoting online safety.

6. At the same time, the Preamble acknowledges the existence of risks in the online environment and the potentially resulting harms, sometimes of a serious nature, to user safety, the enjoyment of rights, the functioning of democracy and other societal interests. These risks exist for all. However, certain individuals and groups may be at greater risk than others, because of their identity, their role or their own contributions to public debate. This includes but is not limited to women and girls, children those in situations of vulnerability and at risk of discrimination – including people with disabilities, national ethnic, linguistic and religious minorities, LGBTI[1] communities, as well as migrants and individuals with a migration background. All of those perceive to belong to these groups may face targeted abuse, including of intersectional nature, structural discrimination or algorithmic exclusion, which limit their ability to exercise their rights online. In particular, technology-facilitated violence against women and girls is a growing problem, with global prevalence estimated at around 85%, and women being 27 times more likely than men to experience online-facilitate violence. These risks are greater for content creators (including media, activists and NGOs), who may face abuse targeted at stopping them from speaking up. These risks threaten not only individuals but also groups, the wider enjoyment of rights in society and, ultimately, democracy.

7. The Recommendation calls for transparent and evidence-based legal frameworks and other initiatives to ensure that online risks, as well as potentially resulting harms, are assessed, addressed and mitigated in a non-discriminatory and human rights-compliant manner that safeguards against disproportionate interferences with freedom of expression and other human rights. It emphasises that such risk assessment and mitigation actions need to be undertaken in consultation with a variety of users, including content creators, affected groups and communities, platforms and other companies and all other relevant stakeholders. Assessments and mitigation actions must be tailored and intersectional, ensuring that all users can safely enjoy their rights online.

8. It is essential to ensure that users are effectively equipped to understand, navigate and respond to online risks. Empowerment is key to this. Empowerment is grounded in human dignity and autonomy of users and contributes to achieving equitable access to online technologies, enabling the full enjoyment of human rights in the online environment and fostering inclusive participation in online spaces for all. Empowerment is not only about protection but also about enabling users to meaningfully engage online. This includes access to online tools, digital literacy, representation and the ability to participate in governance processes. Achieving empowerment requires a rights-based regulatory framework in which user safety, empowerment and systemic accountability are legally mandated and subject to independent oversight.

9. Not all risks can be addressed through empowerment and the burden of addressing risks and potential harms should not fall primarily on those who are most exposed to them. Whenever it is clear or ascertained that empowerment fails or would be likely to fail in mitigating the harmful effects of online risks, States should consider alternative proportionate ways of addressing harms that flow from online risks, including the imposition of proportionate restrictions to content or its accessibility on platforms. In accordance with Article 10, para. 2, of the Convention, any limitations on the right to freedom of expression must be "prescribed by law" and "necessary in a democratic society" for the protection of a legitimate aim. Overly broad or vague measures risk stifling legitimate discourse. Article 17 of the Convention (Prohibition of the abuse of rights) also allows for restrictions on certain expressions by depriving of the protection offered by Article 10 expressive activity that is used to deflect the right to freedom of expression from its real purpose, by invoking it in order to justify, promote or perform acts that are contrary to the text and spirit of the Convention or incompatible with democracy or other fundamental values of the Convention. As per the cases so far examined by the Court, these may include, depending on the specific circumstances, incitement to violence and hatred, the promotion and justification of terrorism and war crimes, the promotion of totalitarian ideologies and the negation of the Holocaust. The Court has more generally emphasised that "[c]ertain classes of speech, such as lewd and obscene speech have no essential role in the expression of ideas" (*Rujak v. Croatia* (dec.), No. 57942/10, 2 October 2012, para. 29) and thus, fall outside the scope of Article 10 of the Convention. A similar line of reasoning might be applied to the dissemination of content that bears no reasonable relationship to the expression of ideas, such as child sexual abuse material or non-consensual private sexual materials shared with the purpose of causing distress to a person. These restrictions, however, must be narrowly circumscribed and be used only in relation to speech that is clearly incompatible with the Convention system itself. This Recommendation therefore advises States to avoid sweeping restrictions and to ensure that legal norms are clearly defined, proportionate and subject to judicial oversight.

10. The principles set out in the Appendix to the Recommendation recognise and address the central role of online platforms, especially those of significant influence that host and regulate vast portions of online public life (see Explanatory Memorandum to paragraph 11, on the definition of "platform of significant influence"). They emphasise that these actors have a responsibility to respect human rights and create an enabling online environment, ensuring the safety of their users, and that they must not operate in a regulatory vacuum. This principle is enshrined in Recommendation CM/Rec(2016)3 on business and human rights, Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, as well as in the UN Guiding Principles on business and human rights.^[2] In the context of online safety, platforms have a responsibility not to contribute to human rights abuses, including by amplifying material that carries risk of harm, to conduct their business with due diligence and to ensure access to effective remedies. They should also conduct human rights impact assessments and provide mechanisms for accountability, redress and user empowerment.

11. The preamble underlines that measures taken through content curation, removal and moderation are interferences with the enjoyment of freedom of expression and information and other rights and may disproportionately affect the exercise of these rights. In line with Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, the principles set out in the Appendix underscore the need for accountability, transparency and remedy mechanisms that align with human rights standards. At the same time, they recognise the need for measures to protect those at risk of being silenced by content carrying the risk of harm.

12. The preamble emphasises strong concern at the concentration of power among a few online platforms, the power asymmetry between these platforms and their users, and the implications of these dynamics for user safety, the human rights of users, and for democratic processes and institutions. In line with Recommendation CM/Rec(2022)11 on principles for media and communication governance, the preamble emphasises the need for a graduated approach to regulation, ensuring that platforms of

significant impact should be subject to enhanced obligations in relation to user safety and empowerment, while cautioning against disproportionately burdensome requirements on providers that do not have such an impact.

Appendix to Recommendation CM/Rec(2026)4 – Principles for online safety and empowerment of users and content creators

I. Rationale, scope, and definitions

Rationale

On paragraphs 1-5

13. Paragraph 1 of the Appendix acknowledges that the internet presents both opportunities for freedom of expression, including access to information, and risks for the safety of users, as content that can harm individuals and society may be disseminated at an unprecedented scale and speed. It recognises that online platforms, particularly a few with significant reach, driven by algorithmic content curation, have acquired a central role in both providing such opportunities and contributing to such risks.

14. Paragraphs 2 and 3 clarify that online safety cannot be understood as separated or isolated from society and the protection and promotion of human rights. Online safety is a component of a wider concept of an enabling online environment that is conducive to human rights. It cannot be equated to the absence of risks online or related to online activities, as this – if attainable at all – could come at the expense of other human rights and fundamental freedoms. The Recommendation moves from the assumption that safety is not a static concept. It varies significantly across time and space, as it may be influenced by both technological development and societal values. Furthermore, safety is inextricably connected with the specific needs of certain sectors of society, such as children, as well as individuals' perceptions and preferences as to their own acceptable level of risks and related need to not be exposed to them. As such, online safety does not differ from offline safety and should be understood in a twofold dimension. Measures taken by States to protect online safety should establish what is the acceptable level of online risks in a democratic society, and what are the appropriate protective measures to address it, either for the general public or for specific sectors of the population, without interfering disproportionately with the exercise of human rights. In addition to that, such interventions should aim to provide individuals both with the capacity to recognise and understand risks and with tools to control their online experience in a way that adapts to their preferences and choices, so to empower them in achieving their desired level of well-being. Ensuring safety in the online environment also requires addressing structural and intersectional inequalities and discrimination, including based on gender, as these can shape both exposure to online risks and the ability to seek protection and redress.

15. Measures that are aimed at reducing the availability, accessibility and visibility of certain content online may be necessary in a democratic society that affords an adequate protection of human rights, provided they are clearly defined by the law, meet a pressing social need and are proportionate to the legitimate aim pursued. Paragraphs 4 and 5 recognise that an approach to online safety based exclusively or primarily on this type of measures cannot alone foster an enabling online environment. Excessive reliance on such measures is insufficient and potentially harmful, as it can lead to arbitrary or disproportionate restrictions that undermine rights, particularly freedom of expression and privacy. In line with the general approach taken in some jurisdictions in Europe and beyond,[3] the Recommendation therefore advocates the development of a new generation of proportionate and evidence-based legal instruments, complementary to existing approaches, that enhance accountability and public oversight of platforms in relation to their design choices and general risk-management practices, while also strengthening the empowerment of users and content creators.

Scope

On paragraph 6

16. This paragraph, together with the definition of platforms, defines the scope of the Recommendation. The principles set out in the Appendix are not intended to tackle and provide solutions for all types of online risks in general. Rather, it focuses on how to address in a human rights-compliant manner risks that have a close connection with the exercise of freedom of expression. These risks can be understood in two overlapping ways: risks that arise as a consequence of exercising free expression and risks that operate to inhibit its exercise.

17. Risks that result from exercising free expression occur in the first place when individuals face retaliation for what they say or publish. For example, a woman journalist might post an investigative article on corruption and subsequently receive death threats or have her personal information leaked online. Another example is when hate speech is spread in response to someone's views, potentially jeopardising their safety. Additionally, risks also arise from expressive activity online that can harm the rights of others, such as incitement to violence or defamation, as well as societal interests, such as the integrity of information and the democratic process.

18. Risks that inhibit the exercise of free expression are those that discourage people from speaking out in the first place. This includes what the European Court of Human Rights has termed the "chilling effect" or "*effet dissuasif*", where fear of consequences silences legitimate debate. For instance, overly broad defamation laws or disproportionate penalties can deter critical reporting on matters of public interest. Furthermore, the risk of being targeted by online violence can inhibit public expression, for example by discouraging women and girls from participating in public debates or sharing their views due to fear of harassment.

19. Many risks fall into both perspectives. Poorly designed content moderation systems, for instance, may wrongly flag or suppress legitimate content, especially from marginalised voices or creators addressing sensitive topics. A human rights activist who uploads a video on human rights violations might find it removed by an automated system labelling it as 'violent content'. Such risks both result from the exercise of free expression and inhibit its future exercise. Chapter II of this Explanatory Memorandum explores the various types of risks in more detail.

On paragraphs 7-8

20. The principles set out in the Appendix are aimed primarily at States but also addresses responsibilities of platforms. Under international human rights law, States bear the obligations and the primary responsibility to guarantee and safeguard rights, but private actors, including platforms, also have responsibilities (see CM/Rec(2016)3 on human rights and business, CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries and CM/Rec(2022)13 on the impacts of online technologies on freedom of expression, as well as the UN Guiding Principles on Business and Human Rights). Such responsibilities may become legal duties under domestic law. The European Court of Human Rights has recognised that, while the duties and responsibilities of platforms may differ from those of a traditional publisher in relation to third-party content, "when internet intermediaries manage content available on their platforms or play a curatorial or editorial role, including through the use of algorithms, their important function in facilitating and shaping public debate engenders duties of care and due diligence, which may also increase in proportion to the reach of the relevant expressive activity" (*Google LLC and others v. Russia*, No. 37027/22, 8 July 2025, para. 79).

21. In line with Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, platforms are therefore expected to respect the human rights of their users. In order to discharge these responsibilities, platforms are expected to take steps to protect the safety of users and content creators. In doing so, they should act with transparency, accountability, and due diligence; and the measures they take must be human rights-compliant, ensuring that their systems and decisions do not result in unjustified restrictions on freedom of expression or other rights.

On paragraphs 9-10

22. The Recommendation pursues two distinct but interrelated objectives: (1) the protection of users' safety in online environments, and (2) the promotion of user empowerment, meaning that they are informed and in control of their online environment and are able to participate in the online sphere fully, freely and equally. Safety and empowerment are mutually reinforcing dimensions of a rights-based online environment. The objectives reflect a dual commitment to enabling individuals to exercise their rights without undue interference and to ensuring that online spaces are structured and governed in ways that uphold human dignity, security and inclusion.

23. These objectives are firmly grounded in the Convention, particularly Article 10, which protects the right to freedom of expression, Article 8, which guarantees the right to respect for private and family life (a broad notion that encompasses the protection of personal data as well as the physical and psychological integrity of a person, including aspects such as sexual orientation and the protection against serious attacks on reputation) and Article 14, which protects the right not to be discriminated against in the enjoyment of rights (see for example, *Denisov v. Ukraine* [GC], No. 76639/11, 25 September 2018, par. 95; *Beizaras and Levickas v. Lithuania*, No. 41288/15, 14 January 2020, par. 109; *Minasyan and Others v. Armenia*, No. 59180/15, 7 January 2025, par. 53). In the most serious cases, online violence may even engage positive obligations of prevention and protection under Articles 2 (right to life) and 3 (Prohibition of torture).

24. The European Court of Human Rights has consistently held that the rights enshrined in the Convention impose both negative and positive obligations on the State. The negative obligation requires that States refrain from unjustified interferences with the exercise of rights. Any restriction on freedom of expression must meet the conditions set out in Article 10, paragraph 2: it must be prescribed by law, pursue one of the legitimate aims listed and be necessary in a democratic society, meaning that it must be proportionate and respond to a pressing social need.

25. The positive obligations entail that States must take reasonable and appropriate steps to secure the effective enjoyment of Convention rights, including in private and online settings, which may require them to place and enforce obligations on private actors. The European Court of Human Rights has affirmed that States must ensure an environment in which individuals are able to exercise their right to freedom of expression without facing threats, harassment, or violence, whether from public authorities or private actors: "[T]he positive obligations under Article 10 of the Convention require States to create ... a favourable environment for participation in public debate by all the persons concerned, enabling them to express their opinions and ideas without fear, even if they run counter to those defended by the official authorities or by a significant part of public opinion, or even irritating or shocking to the latter" (*Khadija Ismayilova v. Azerbaijan*, Nos. 65286/13 and 57270/14, 10 January 2019, para. 158; See also *Özgür Gündem v. Turkey*, No. 23144/93, 16 March 2000, paras. 43, 44; *Dink v. Turkey*, No. 2668/07 and others, 14 September 2010, available only in French, par. 137).

26. Article 8 imposes a similar obligation. The Court has emphasised that "while the essential object of Article 8 is to protect the individual against arbitrary interference by the public authorities, it does not merely compel the State to abstain from such interference: in addition to this negative undertaking, there may be positive obligations inherent in the effective respect for private life. These obligations may involve the adoption of measures designed to secure respect for private life even in the sphere of the relations of individuals between themselves" (*Aksu v. Turkey* [GC], nos. 4149/04 and 41029/04, 15 March 2012, para. 59; *Minasyan and Others v. Armenia*, cited above, para. 58).

27. In the context of online safety, this dual obligation implies that States must take active measures to appropriately respond to content that carries risk of harm, including abuse, intimidation and discrimination that may inhibit the full participation of individuals, in particular those belonging to categories at

heightened risk, as identified in paragraph 16. But in doing so, States must refrain from imposing disproportionate restrictions. For example, public authorities should not order the takedown of an entire website or internet domain when only certain pages on that domain carry content that is legally restricted.

28. The positive obligations of States further encompass the duty to promote equitable access to online communication infrastructures, promote media and information literacy, and to take regulatory steps that foster the development of inclusive, rights-respecting online platforms (see CM/Rec(2016)1 on protecting and promoting the right to freedom of expression and the right to private life with regard to network neutrality and CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries).

Definitions

On paragraph 11

29. The definitions in paragraph 11 are key to understanding the scope and aims of the Recommendation.

30. The definition of "user" is intentionally broad, encompassing any individual, any group of individuals, as well as any legal entity (including media companies and civil society organisations).

31. "Content creator" is defined as a subset of users. The salient elements of the definition, which draws on the definition of "journalist" in Recommendation Rec(2000)7 on the right of journalists not to disclose their sources of information, are as follows. A content creator must:

- aim to reach an audience beyond their private circle; this means that not anyone with a social media account who regularly posts to their friends or family is a content creator;
- be engaged in the dissemination or production of content either regularly or professionally; this includes any natural or legal person whose professional activity, or a significant aspect thereof, consists in creating online content over which they exercise editorial responsibility. Such persons may operate as business entities for commercial purposes but may also be incorporated as non-governmental organisations, not-for-profit entities, citizens associations or such other legal forms as domestic law may allow. It may also include users who, without being registered as a business or otherwise and without necessarily having a commercial purpose, frequently and regularly publish content on platforms;^[4]
- produce and disseminate information and ideas, in text, audio, visual, audiovisual or other form. This is intentionally broad: "information and ideas" mirrors the text of Article 10 of the Convention, and "text, audio, visual, audiovisual, or other form" is deliberately worded open-endedly and mirrors the wording of Recommendation CM/Rec(2011)7 on a new notion of media; it is however restricted to information and ideas produced or disseminated via a platform.

32. The definition of "platform" is based on the similar definition in Recommendation CM/Rec(2022)11 on principles for media and communication governance, but it narrows it for the purpose of this recommendation. It singles out "the connection of users and facilitation of their interactions" and the purpose "to exchange ideas and information" as defining elements. Accordingly, platforms, such as marketplaces and sharing economy platforms, that connect users, but primarily for different function, are in principle excluded altogether. Nevertheless, if the exchange of ideas and information becomes prominent on platforms originally designed for other purposes, such as gaming, such platforms would fall into the scope of the definition.^[5] On the other hand, for intermediaries that disseminate media content over which they exercise editorial and curatorial control— such as news portals or video-on-demand

platforms – only those functionalities that allow user-to-user interaction, such as a comment section, would be covered. The definition explicitly includes only “publicly accessible fora”. This also means that platforms offering exclusively private communication services, including closed group chats, would be excluded, while specific functionalities such as public-facing channels and groups messaging apps are included, as evidence shows the use by extremist actors of public-facing messaging groups to organise and mobilise.[6]

33. The Recommendation establishes a new definition of “platforms of significant influence”, defining these as platforms that because of their size, reach or impact, play a substantial role in shaping the information environment globally or in particular territories. The criteria are general in nature and alternative, rather than cumulative, in scope. This means that States have a margin of discretion in choosing which ones to apply and their application is subject to the specification in domestic law of more precise and measurable criteria, provided *ex ante* to avoid arbitrary classification of platforms under this category. For example, they may rely on the number of active users, like in the EU Digital Services Act. However, they may also decide to include platforms which, despite having a smaller reach in terms of numbers, exert significant influence over public discourse by virtue of their role in originating, amplifying or coordinating narratives and can thus amplify risks (see also Explanatory Memorandum to para. 41).

34. The definitions of “platform design” and “user empowerment” build on the definitions provided in the Guidance Note on countering online mis- and disinformation.[7] The definition of “platform design” emphasises that this includes user-facing trust and safety functionalities. Examples would be recommender systems, the use of warning labels, and content moderation systems. Platform design, however, does not include choices in the moderation of lawful content that are based on the point of view or opinion of individual pieces of content. The definition of “user empowerment” provides additional specificity to the definition as provided in the Guidance Note, clarifying that this can include both online and offline measures. Examples would be media and information literacy campaigns to help users recognise misinformation, understand algorithmic bias and evaluate sources; easily usable and effective privacy and consent dashboards; customisability of design of recommender systems or other features of services; and transparency about and user control over the algorithms that drive the visibility of content.

35. The definition of “self-regulation” builds on the definition provided in the Guidance Note on countering online mis- and disinformation. It adds an explicit statement that “this includes the contractual policies and rules of platforms that affect users of their services”. The definition of “co-regulation” is drawn from Recommendation CM/Rec(2022)11 on principles for media and communication governance.

36. The definitions of “legally restricted content”, “illegal content”, “legal but regulated content” and “lawful content” for the purposes of the Recommendation and its principles need to be understood as a related set of definitions:

“illegal content” refers to any content that is prohibited under criminal, administrative or civil law. Examples include content amounting to:

- acts of child sexual exploitation and abuse, including making available child sexual abuse material as defined in Article 20 of the Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS No. 201) (Lanzarote Convention) and Article 9 of the Convention on Cybercrime (ETS No. 185) ;
- acts of a racist and xenophobic nature, as defined in the Additional Protocol to the Convention on Cybercrime concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (ETS No. 189)

- acts of violence against women, as defined in the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (CETS No. 210) (Istanbul Convention) and in line with the Recommendation CM/Rec(2026)2 on accountability for technology-facilitated violence against women and girls;
- illegal hate speech, in line with Recommendation CM/Rec(2022)16 on combating hate speech;
- incitement to violence and serious threats of physical harm;
- terrorist content, such as material inciting or instructing terrorist acts.

“legal but regulated content” means content which is lawful in itself, but whose publication, amplification or visibility may be restricted as prescribed by law in particular contexts. This may include:

- electoral silence rules or blanket bans on opinion polls during electoral periods, to protect the integrity of the democratic process (see CM/Rec(2022)12 on electoral communication and media coverage of election campaigns, para. 4.6);
- pornography, or other content such as graphic depiction of violence, self-harm material, which may be lawful but is age-restricted to protect children;
- the display of personal data or information (such as home addresses or telephone numbers), which may be de-indexed by search engines to protect privacy;
- content promoting products such as alcohol, tobacco, or gambling, which may be restricted in its advertising or placement;

“legally restricted content” covers both categories of illegal content and of legal but regulated content. In practice, this means it covers clearly unlawful expressions, such as certain forms of hate speech and incitement to violence, as well as lawful expressions subject to contextual limitations, such as age-restricted or time-limited dissemination.

“lawful content” means any content, including expressions as well as any manifestation of behaviour of users, that does not qualify as being legally restricted.

37. The definitions of “flag”, “notice” and “order” refer to the different ways in which users or public authorities may bring content that requires action to the attention of a platform. An example of “flag” would be a user reporting a social media post for hate speech or a video as containing misinformation about public health, using the platform’s built-in report button, when deemed to be contrary to either the law or the platform’s contractual rules. An example of a “notice” would be a copyright holder submitting a request to have pirated content removed from a video-sharing platform or a data subject requesting the removal of personal information under data protection laws. An example of an “order” would be a court directing a platform to remove defamatory content targeting an individual or a regulatory authority mandating the suspension of an account spreading illegal hate speech.

II. Online risks related to freedom of expression

On paragraphs 12-16

38. The European Court of Human Rights has consistently underlined that freedom of expression extends beyond neutral or agreeable speech to include views that may offend, provoke or unsettle. This principle is particularly relevant online, where debate is often sharp and diverse, and where minority or dissenting voices must be able to participate fully. While the Court has noted the prevalence of vulgar abuse in some online spaces, it has also stressed the importance of context and the higher threshold of tolerance expected, especially in relation to debates on matters of public interest (see, for example, *Handyside v. the United Kingdom*, cited above, para. 49; *Delfi AS v. Estonia* [GC], cited above, paras. 131-139; *Tamiz v. the United Kingdom* (dec.), cited above, para. 81).

39. The principles set out in the Appendix, while reaffirming robust protection for expression, recognise that the online environment also generates risks that can undermine or deter the very exercise of this right, as well as other rights. They also affirm the need to strike a fair balance to preserve the essence of all the human rights at stake. These risks affect individuals and groups, as well as the public at large, and can have serious societal consequences. When people face hostile or manipulative behaviour online, such as the coordinated harassment of journalists, doxxing of activists' personal details or the posting of non-consensual intimate images, they may be intimidated and pull back from expressing themselves on issues of public interest in the future. The fear of reputational or physical harm, legal trouble or sheer emotional exhaustion can stifle their willingness and capacity to speak openly, especially on controversial topics.

40. Broader online threats can distort the way people find and understand information. For example, when content shared through inauthentic accounts or an undeclared manipulated video are pushed to the top of search results or social media feeds, this can drown out accurate news and make it harder for people to hear a range of views. This does not just affect individuals; it has serious consequences for society. For example, false claims about vaccines can lead to lower vaccination rates and greater health risks (see, *Bielau v. Austria*, Application No. 20007/22, 27 August 2024, paras 44-45). During elections, coordinated efforts to influence social media algorithms and public perceptions around topics of public interest, as well as disinformation about the electoral process itself, could shake public trust in such processes or in the election results themselves. And when people are repeatedly exposed to divisive messages, this can deepen mistrust between groups and make it harder to agree on shared solutions to public issues.

41. Paragraph 13 breaks the risks down into three categories:

- Risks to personal and community safety and well-being;
- Risks to the democratic process, information integrity and informed public debate; and
- Risks associated with the systems deployed by providers which may interfere with the rights to freedom of expression, privacy and personal data protection and other rights.

42. Examples of risks to personal and community safety and well-being include:

- the risk of encountering or being targeted by hate speech, including sexist hate speech, hate crimes and discrimination, threats and coercion;
- the risk of encountering or being targeted by harassment, stalking, abuse and cyberbullying, both general and identity-based;
- the risk of being exposed to content that is disruptive to mental and emotional well-being, including content promoting suicide and self-harm, or content that may contribute to eating disorders or negative body image concerns, including risks associated with body dysmorphic disorder;

- the risk of being targeted by sexual harassment, sexual exploitation or intimate image abuse, including sexual digital forgery;
- risks to private life, ranging from surveillance to identity theft, fraud, blackmail and financial scams;
- the risk of being exposed to recruitment and radicalisation by terrorist and extremist groups.

43. Examples of risks to the democratic process, information integrity and informed public discourse include:

- the risk that coordinated inauthentic behaviour, including disinformation campaigns during elections, influence voters, suppress turnout or sow doubt in the final result;
- the risk that lawsuits, regulatory threats or coordinated mass-reporting campaigns pressure media outlets or platforms to self-censor or remove legitimate content, weakening the media's watchdog role and the public's access to critical information;
- the risk that content produced purely for monetary gain, including AI-generated content (colloquially referred to as "AI slop"), crowds out public interest information;
- the risk that deepfake videos or other convincing forms of impersonation circulate widely before they can be debunked, undermining confidence in verifiable evidence;
- the risk that women and girls and those in situations of vulnerability or at risk of discrimination withdraw from public discourse, thereby reducing the diversity of voices and perspectives.

44. Examples of risks associated with the systems deployed by providers include:

- the risk of exclusion, denial of access and other barriers to using public online systems;
- the risk that recommendation algorithms elevate sensational or false stories over accurate reporting, narrowing the spectrum of viewpoints people encounter and deepening polarisation;
- the risk that design choices facilitate and amplify the spread of content that carries a higher risk of being legally restricted;
- the risk of the demotion of lawful content, affecting its visibility and reducing traffic (so-called 'shadow banning').

45. Not all of these risks stem from behaviour that is criminal or otherwise illegal. While acts such as certain forms of hate speech, incitement to violence, cyberviolence or online harassment and bullying are illegal in most Council of Europe member States, other forms of potentially harmful content, such as material that may undermine mental or emotional well-being, may fall outside the scope of legal prohibitions and regulation. Yet this material can still produce significant harm to individuals or to society as a whole. The crucial issue, therefore, is not only whether the behaviour is unlawful, but what kind of response is adequate, proportionate and effective to minimise the risks of harm. Such responses should be designed in a way that strikes a fair balance different human rights, including the rights to freedom of expression and to private life, including the protection of personal data.

46. As paragraph 16 emphasises, the widespread availability and use of artificial intelligence in the production and dissemination of content may significantly amplify existing risks. The Guidelines on the implications of generative artificial intelligence for freedom of expression,^[8] explore, for example, risks, alongside opportunities, that are associated with the use of generative AI tools.

On paragraph 17

47. Paragraph 17 emphasises that certain categories of users, especially content creators, are at a higher risk than others, either because of their identity or because of their position. It also emphasises that risks online can have ramifications in the physical environment. Online harassment and threats are frequently a precursor to other forms of violence: there are countless reports of journalists, activists and politicians who have been stalked, attacked or otherwise targeted in the physical world after campaigns of online abuse.[9] This is particularly true for women in public roles, who face disproportionate levels of online harassment and gender-based targeting.[10]

48. The heightened risks to children are well-documented and have been the subject of a previous Council of Europe instruments, such as the Lanzarote Convention and Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the online environment. These Guidelines recognise that children can be exposed to a range of serious harms, from sexual exploitation to harassment and other threats to their well-being. It also makes clear that exposure to risk is not uniform: children's needs evolve with age and maturity. Effective protection must therefore be graduated, empowering and child-centred, balancing safety with the right to grow, explore, and participate meaningfully in online life. The Lanzarote Committee, the monitoring body of the Lanzarote Convention, has issued in 2022 an Implementation Report on the protection of children against sexual exploitation and sexual abuse facilitated by information and communication technologies, in 2019 an Opinion on child sexually suggestive or explicit images and/or videos generated, shared and received by children and, in 2024, a Declaration on the protection of children against sexual exploitation and sexual abuse facilitated by emerging technologies on.

49. The online risks to women and girls, especially content creators, are equally well-documented and the subject of long-standing work by the Council of Europe. The Commissioner for Human Rights of the Council of Europe has rung the alarm on violence against women and girls in the online world, describing the internet as "fertile grounds for gender-based violence against women and girls to an alarming extent, and with little accountability" and calling for action.[11] The Court has issued several judgments in which it found a violation of human rights in cases where States insufficiently protected women's right to protection against online abuse (*M.Ş.D. v. Romania*, No. 28935/21, 3 December 2024; *Volodina v. Russia* (No. 2),

No. 40419/19, 14 September 2021; *Buturugă v. Romania*, No. 56867/15, 11 February 2020). The Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO), which oversees the implementation of the Istanbul Convention, issued in 2021 its General Recommendation No. 1 on the online dimension of violence against women, defining various forms of cyberviolence (such as non-consensual image sharing, online stalking and threats) and urging States to take steps for the protection of women and girls online and for the prevention and prosecution of online violence. Furthermore, the Committee of Ministers has adopted Recommendation CM/Rec(2026)2 on accountability for technology-facilitated violence against women and girls, providing guidance to member States on enhancing legal, institutional and regulatory responses to such violence.

50. Women and girls, children, and those in situations of vulnerability and individuals and groups at risk of discrimination, including people with disabilities, national ethnic, linguistic and religious minorities, LGBTI communities, as well as migrants and people with a migration background, are at risk of identity-based online targeting. In its 2024 Annual Report, the EU Fundamental Rights Agency (FRA) emphasised the need for regulation of online spaces, referring to these as "an area of high risk for fundamental rights – particularly for vulnerable and marginalised people", in particular the risk of "the 'othering' of particular groups through disinformation or the spread of online hatred and by creating barriers for vulnerable demographic groups", concluding that "addressing these risks in the course of creating the online environment of the future is central to building a more inclusive Europe." [12]

51. Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors highlights the risks faced by journalists and notes that "violations are increasingly taking place online". Threats to online safety have serious consequences for the ability of journalists to report on issues of public interest and for the public's right to be informed. A 2017 Council of Europe study found that 37% of journalists interviewed had self-censored potentially critical reports because of a risk of harm. A 2020 follow-up study highlighted the experiences of twenty journalists.[13] One of the journalists interviewed for the study was Daphne Caruana Galizia. She was murdered only days after she was interviewed. Women journalists are far more likely than their male counterparts to face online abuse. A 2021 UNESCO report demonstrates the extent of attacks against women journalists and the impact on their well-being, their work and press freedom at large.[14]

52. In one of the most prominent cases of online abuse targeting women journalists, *Khadija Ismayilova v. Azerbaijan* (cited above), the European Court of Human Rights examined the covert surveillance of a female investigative journalist whose bedroom and other private spaces were secretly filmed and the footage leaked online. The Court described the abuse as "grave and an affront to human dignity" and "a serious, flagrant and extraordinarily intense invasion of her private life." It further underlined that "the applicant is a well-known journalist and there was a plausible link between her professional activity and the aforementioned intrusions, whose purpose was to silence her" (para. 116).

53. Politicians, researchers, educators, scientists, activists and others who frequently contribute to debate on matters of public interest are also frequently targeted with abuse aimed at stopping them from participating in public debate.[15]

54. People belonging to several of these groups are even more likely to face intersectional risks. When various risk grounds intersect, individuals are exposed to more complex forms of discrimination, exclusion and violence leading to unique lived experiences and vulnerabilities.[16] The Council of Europe Commissioner for Human Rights provided several examples of such intersecting forms of abuse in a 2022 statement on online violence against women and girls.[17]

On paragraph 18

55. Paragraph 18 makes clear the existence of online risks as such do not always justify the introduction of measures that interfere the exercise of freedom of expression and other human rights, in particular by restricting or regulating content. Measures taken by States should respond to a pressing social need and be proportionate, a consideration that implies the unavailability of other appropriate measures that do not restrict rights or are less restrictive (*Glor v. Switzerland*, No. 13444/04, 30 April 2009, para. 94), such as for example media literacy and other empowerment initiatives. The same approach should be followed by platforms, for example in the context of content moderation (see Explanatory Memorandum to Recommendation CM/Rec(2022)11, para. 12.2).

56. To ensure that the requirements of necessity and proportionality are met, such measures should be introduced only when there is evidence of an actual or probable risk of harm. This evidence-based approach is reinforced throughout the Recommendation and its principles. The Preamble highlights the need for transparent and evidence-based legal frameworks to assess and address online risks in a manner consistent with human rights. Paragraph 5 calls for evidence-based regulatory and co-regulatory frameworks; paragraph 45 requires that content rules be grounded in transparently gathered evidence; and paragraphs 51, 64 and 66 similarly oblige regulatory authorities and legislators to rely on evidence when designing interventions.

On paragraph 19

57. Paragraph 19 highlights that measures intended to improve online safety can themselves threaten human rights. A key concern is the imposition of disproportionate restrictions on freedom of expression, particularly when States or regulators incentivise platforms to act pre-emptively, leading to overremoval or “collateral censorship”. The French *Loi Avia* (law No. 2020-766 of 24 June 2020 on combating hateful content on the internet) illustrates this risk: although aimed at tackling hate speech, its core provisions were struck down in 2020 by the Constitutional Council as a disproportionate threat to free expression, lacking judicial oversight and safeguards against overremoval (decision n° 2020-801 DC, 18 June 2020). In the context of broadcast regulation, the European Court of Human Rights has emphasised that an animal rights broadcast ad could not be banned simply because viewers might find it “unpleasant” (*Verein gegen Tierfabriken Schweiz (VgT) v. Switzerland (No. 2)* [GC], No. 32772/02, 30 June 2009, para. 96).

58. Similar risks arise where recommender systems or content-ranking tools are, for reasons of safety, designed and run in ways that invisibly reduce the reach of certain viewpoints, disproportionately affecting minority voices or controversial but lawful perspectives. For example, research published by the EU Fundamental Rights Agency in 2022 found that offensive and hate speech detection algorithms produce biased results and can have a discriminatory impact.[18]

59. States and private actors should therefore ensure that safety measures are necessary, proportionate, transparent and evidence-based, with independent scrutiny and effective remedies. The objective is not to set safety and rights in opposition, but to ensure that safety measures strengthen the protection of rights as a whole.

III. General principles for an enabling online environment

Principles for States

On paragraphs 20-23

60. The emphasis on creating an enabling online environment reflects a long-standing principle in the jurisprudence of the European Court of Human Rights and in documents of the Committee of Ministers: under Article 10 of the Convention, States have positive obligations to promote conditions in which freedom of expression can be exercised effectively and by all. In its case-law on attacks against journalists, the Court has repeatedly underlined that such conditions require effective protection for those most at risk. In *Dink v. Turkey* (cited above) and similar cases, the Court found some States in breach of their obligations where they failed to shield journalists from threats, harassment and violence, noting that “positive obligations imply, inter alia, that States are required, while establishing an effective system of protection for authors and journalists, to create an environment conducive to the participation in public debates of all those concerned, allowing them to express their opinions and ideas without fear” (para. 137).

61. Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors emphasises that an enabling environment requires that, “as a minimum, the safety, security and protection are guaranteed effectively in practice for everyone, in particular journalists and other media actors, and there is an expectation that they can contribute to public debate without fear and without having to modify their conduct due to fear” (principles, para. 18).

62. The same principle holds true for online safety: States have a duty to promote conditions for everyone to contribute to public discourse without fear of intimidation or reprisals. The existence of a positive obligation to “create a safe and enabling environment for everyone to participate in public debate” is recognised by the Committee of Ministers in the preambles of Recommendations CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, CM/Rec(2022)16 on combating hate speech and CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression. Recommendation

CM/Rec(2016)5 on internet freedom calls on States to “create an enabling environment for Internet freedom” and Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries affirms that legislation applicable to internet intermediaries “should create a safe and enabling online environment for private communications and public debate”. Furthermore, Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems calls on States to “ensure that algorithmic design, development and ongoing deployment processes incorporate safety, privacy, data protection and security safeguards by design” (Section B, para. 3.2).

63. The present Recommendation and its principles build on this *acquis*, stressing that safety, inclusiveness, pluralism and user empowerment are essential to a rights-compliant online environment. While the Recommendation focuses on measures relating to the online environment, its principles recognise that achieving this goal requires a mix of regulatory and educational measures. Policies in this field should therefore form part of a comprehensive and coordinated strategy that tackles the underlying societal conditions and inequalities that give rise to online abuse and shape users’ exposure to it. Such a strategy should include measures to promote equality, social cohesion and democratic values; measures to reinforce the rule of law and public safety; and measures to empower users to make informed choices about their online experience. In practice, this may encompass:

Education and awareness: integrating online citizenship and online safety into school curricula; campaigns that build resilience to disinformation; programmes that encourage respectful online behaviour;[19]

Media and information literacy: national strategies to strengthen critical engagement with online content; public support for fact-checking initiatives; partnerships with civil society to improve access to trustworthy information;[20]

Community empowerment: funding initiatives to support groups disproportionately targeted online (e.g. women, minorities, LGBTI persons, journalists); funding hotlines or support services for victims of online abuse; funding training for NGOs and community leaders on online safety; [21]

Support for quality journalism and pluralistic media: ensuring independent public service media; offering transparent and fair systems of financial support for investigative journalism; protecting journalists from legal harassment and abusive litigation (SLAPPs);[22]

Law enforcement and accountability: effective criminal investigations into technology-facilitated violence and abuse; specialised police units trained to address cybercrime sensitively and effectively; safeguards, including protection from criminal liability of intermediaries for acts of their users,[23] to ensure prosecution and judicial processes respect human rights; and accessible complaint and redress mechanisms for users.[24]

64. Paragraph 22 affirms that a safe and enabling online environment depends on preserving the technical safeguards that protect rights, including end-to-end encryption of private communications. In *Podchasov v. Russia* (No. 33696/19, 13 February 2024, paras 76-79), the Court has observed that measures that weaken or bypass encryption (for example, decryption mandates, key-escrow ‘backdoors’ or functionally equivalent measures) risk amounting to general and indiscriminate surveillance which is considered disproportionate and in breach of Article 8 of the Convention.

65. Paragraph 23 underscores the need to ensure that risks of exclusion and marginalisation from online spaces for specific categories as a consequence of either measures taken by States or lack thereof, are duly considered and taken into account. Both State regulation of how platforms protect and empower

users online and the failure to act where protection is needed can unintentionally create barriers to accessibility and inclusion. For example, safety tools such as age-verification or identity-verification requirements may be introduced to protect children, but if poorly designed they can exclude adults without official identification documents or those lacking sufficient digital literacy. Similarly, automated content moderation systems aimed at tackling hate speech or disinformation may disproportionately silence minority voices, women or persons with disabilities who rely on particular language patterns or assistive technologies. Conversely, a failure to intervene, for example by not requiring platforms to have accessible reporting channels for abuse, can leave groups that are already at heightened risk without protection. States therefore have a responsibility to ensure that measures designed to promote safety do not unintentionally entrench disadvantage, and that inaction does not perpetuate unequal exposure to online risks.

66. Paragraph 25 builds on the principle that transparency is essential to democratic accountability, both in the activities of States and in the operations of influential private actors such as online platforms. The Council of Europe Convention on Access to Official Documents (CETS No. 205) establishes that the public has a right to know how public authorities exercise their powers, subject to exceptions that should be narrowly interpreted and balanced against the overriding public interest in disclosure. The protection of whistleblowers plays an important role in this framework. As Recommendation CM/Rec(2014)7 on the protection of whistleblowers makes clear, individuals who disclose information in the public interest – for example, about unlawful surveillance requests, unsafe moderation practices or systemic risks to users – should be protected from retaliation.

Principles for platforms

On paragraphs 26-29

67. Because of their central role in facilitating and shaping online expression, all platforms should integrate safety and empowerment considerations into their core service design and governance choices. It is not sufficient for platforms to take steps only reactively: safety and empowerment must be embedded by design, including in the development and deployment of artificial intelligence systems, recommender tools and content-moderation mechanisms. They should pay specific attention to the higher risks that may be faced by women and girls, children, and those in situations of vulnerability and individuals and groups at risk of discrimination, including people with disabilities, national ethnic, linguistic and religious minorities, LGBTI communities and people with a migration background. These provisions build on similar principles previously set out across a range of instruments. For example, Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries stresses that intermediaries should embed respect for human rights into their design and decision-making processes, and should carry out regular assessments of the human rights impact of their services. Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems requires that human rights, democracy, and the rule of law be safeguarded throughout the design and lifecycle of AI and algorithmic systems. Assessment and mitigation of risks and adverse impacts is also a core feature of the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225).

68. In accordance with the principle of proportionality and the graduated approach to platform regulation, such responsibilities and the means to uphold them take different forms depending on platforms' size, reach or impact. Platforms of significant influence have heightened responsibilities and may thus be subject to specific legal requirements, such as detailed in Sections IV and V of the Appendix. While safety-related responsibilities for smaller platforms should be more limited in scope, so as not to overburden them and stifle innovation and competition, they remain important. Such platforms may be expected, for example, to adopt clear and accessible community standards, ensure that users have simple tools to report abuse and provide timely responses to complaints. In general, when subject to regulatory

oversight, all platforms should be able to show that they have acted with due diligence in taking risks into consideration and incorporating reasonable measures in their design that are adequate to their risk level and capacity.

69. Safety measures should not be pursued at the expense of media pluralism, diversity of voices or the open and inclusive nature of public discourse. For example, rules that incentivise platforms to automatically remove large volumes of flagged content may disproportionately affect minority voices or controversial but lawful perspectives, particularly when automated moderation tools misinterpret cultural or linguistic expressions. Similarly, obligations requiring platforms to prioritise trusted sources of information may unintentionally marginalise local or independent media, reducing the diversity of viewpoints available to users. Efforts to counter the publication of legally restricted content should avoid imposing overbroad restrictions that could be exploited to suppress legitimate criticism of governments or powerful actors.

70. To prevent these risks, interventions should be transparent, proportionate and grounded in international human rights law. Transparency requires clear rules on how content is moderated and ranked, as well as accessible remedies for users whose speech is unfairly restricted. Proportionality requires that restrictions be narrowly targeted to address clearly defined harms, rather than imposed through broad measures that risk suppressing legitimate debate. And grounding in human rights law ensures that protective measures are consistent with, *inter alia*, the rights to freedom of expression, privacy and non-discrimination.

71. Paragraph 29 underlines that platforms that operate at scale in a country or region must understand and respond to local contexts. This means, for example, knowing and recognising that in a given country or region, certain minority groups may be disproportionately targeted by coordinated harassment campaigns; that disinformation may be spread in local languages that automated moderation systems do not detect; and that hate speech may take the form of coded terms unfamiliar to staff based outside the country. They should also take account of gender-specific risks, including how context-specific social norms and online

behaviour can expose women and girls to particular risks of harm. To address these risks effectively, platforms should employ or contract staff with knowledge of local political, cultural and social dynamics; they should designate accessible points of contact for users and regulators; and they should ensure that safety teams are fluent in the official languages of the jurisdiction. This is a matter of effectiveness: without local expertise, moderation systems miss abuse, remove legitimate content in error, or fail to respond to urgent risks such as incitement to violence. Good practice examples include platforms establishing dedicated election integrity teams in countries during sensitive political periods or contracting local NGOs to provide expertise on safety threats and human rights.

Principles for content creators

On paragraphs 30-32

72. In line with the text of Article 10, paragraph 2, of the Convention, content creators, like everyone who exercises their right to freedom of expression, have duties and responsibilities. The level of these duties and responsibilities depends on factors such as the format of the content and its relevance to public debate. For example, a journalist producing investigative reports on political corruption bears a higher responsibility for accuracy and impartiality than a fashion influencer or a meme creator. However, all content creators are expected to respect human rights and dignity.

73. The European Court of Human Rights has consistently held that, in so far as the protection afforded by Article 10 is concerned, the role of bloggers and popular users of the social media disseminating content on matters of public concern may be assimilated to that of "public watchdogs", traditionally played by the press (*Magyar Helsinki Bizottság v. Hungary* [GC], 2016, para. 168). Such

protection is subject to the condition that they comply with the duties and responsibilities traditionally connected with the function of journalist. Content creators who publish on issues of public interest, claim professional expertise or reach significant audiences bear a heightened duty to act in good faith and uphold principles of accuracy, fairness and integrity. For example, researchers may be under an enhanced obligation to ensure accuracy when communicating to the general public.[25] This reflects the case-law of the European Court of Human Rights on the "duties and responsibilities" of those contributing to public debate (*Stoll v. Switzerland*, No. 69698/01, 10 December 2007, para. 104; *Jersild v. Denmark*, No. 15890/89, 23 September 1994, para. 31; *Savva Terentyev v. Russia*, No. 10692/09, 28 August 2018, para. 79). In *Bielau v. Austria* (cited above), a doctor had been fined by the Disciplinary Council of the Austrian Medical Association for publishing scientifically untenable claims about vaccines, a decision which was upheld by domestic courts. The Court stressed that "[r]estricting the freedom of expression of doctors may be called for in cases of categorical and untrue public information on medical questions, in particular if that information is published on a website, to protect the health and well-being of others" (para. 44). Noting that the doctor promoted "self-healing and homeopathy" and used the statements to advertise his services, the Court treated him as a content creator with a professional background whose expression could be restricted when he failed to meet these duties and responsibilities.

74. The responsibilities outlined in paragraphs 30–31 align with emerging regulations on the category of content creators often referred to as "influencers".[26] Attention is drawn to the need for those content creators whose activity focuses on current events and matters of public interest, sometimes known as "journalist influencers" or "newsfluencers", to align with journalistic standards.

75. Several European States have adopted or are developing rules for social media personalities with sizeable followings who shape public opinion. For example, France has passed legislation that, *inter alia*, bans the commercial promotion by influencers of certain high-risk products (e.g. surgery, cryptocurrencies, nicotine) and requires transparency about advertising and photo modification.[27] In Spain, "users of special relevance" of video-sharing platforms are under an obligation to register and are subject to the audiovisual media rules on protection of minors and audiovisual commercial communications. The development of self-

and co-regulatory codes of conduct is also promoted.[28] In countries without specific legislation, influencers are typically covered under consumer protection law and, if they meet the definition and qualify as providers of such services, under the rules on audiovisual media services. The latter often transpose and implement the EU Audiovisual Media Services Directive, which prohibits incitement to hatred, protects children, and regulates advertising, sponsorship and product placement.[29]

76. In this context, regulators and advertising self-regulatory organisations across Europe are issuing regulations and guidance documents to clarify the legal and ethical duties of influencers, especially around commercial content, transparency and the protection of children. For example, the Belgian audiovisual media regulator has adopted a Content Creator Protocol covering commercial content, child protection and hate speech; the Estonian regulator has issued labelling guidance and is developing rules for influencers as on-demand audiovisual media service providers; the Norwegian regulator has set mandatory labelling rules; in Italy audiovisual regulatory guidelines apply to influencers deemed to be audiovisual media service providers. In Sweden and France, the respective advertising self-regulatory bodies formally apply the 2024 International Chamber of Commerce Advertising and Marketing Communications Code, which also covers influencers.[30] The overall trend is to strengthen transparency and user protection, often in the context of audiovisual media services regulation, with growing expectations of good practice and compliance.

77. Paragraph 32 outlines the responsibilities of parents and legal representatives of children who, within the age limits and other regulatory boundaries that may be established by domestic law, act as content creators. Conversely, it does not cover the case of parents or legal guardians who, in acting as content creators, use or exploit the image of their children (so-called "sharenting"). The need for parents and legal representatives to act in the best interests of their children is a well-established principle in international and Council of Europe human rights standards. The UN Convention on the Rights of the Child, as well as Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the digital environment, underline that the child's dignity, safety and development must guide all actions concerning them. Where children act as content creators, parents or legal representatives bear a particular duty of care, not only to comply with applicable labour, advertising and data protection rules but also to safeguard the child's well-being. In practice, this means ensuring that children are not pressured into producing content, that their education, rest and play are not undermined by online activities, and that they are protected from risks such as overexposure, harassment or commercial exploitation. For example, parents should be attentive to how much personal information a child reveals online and should avoid publishing content that could be prejudicial to the child's well-being and reputation later in life. The aim is to balance children's opportunities to participate, create, and express themselves online with robust safeguards that protect their rights and long-term interests. Children, especially when they act as content creators, should benefit from digital, media and information literacy to empower them to make informed choices, access appropriate remedies and support if their content is misused or results in victimisation of any kind.

78. National practice in this area is emerging. For example, in France specific legislation regulates the commercial exploitation of the image of children under the age of sixteen on online platforms.^[31] The law, which also covers the case of children being the subject, rather than the author, of the content, introduces a procedure for oversight by a public authority of the non-occasional production and dissemination of content, to be distributed on platforms, in which an underage child is the main subject. Following a declaration by the parents or legal guardians, the competent authority may address directives to them to safeguard the health and well-being of the child, as well as the fulfilment of their right to education. Additionally, it protects their

earnings and enshrines the right to be forgotten, meaning that platforms will be obliged to take down content on the child's request.

IV. Principles for legal frameworks on online safety and user empowerment and their implementation and enforcement

Common principles

On paragraph 33

79. The positive obligation of States to address the risks of harm online arises from their broader duty to protect human rights, including the rights to private life and freedom of expression, all of which can be impacted by online activities. Increasingly, these risks are being addressed through national or supranational legislative frameworks, such as the EU Digital Services Act or the United Kingdom Online Safety Act.

80. In the context of online safety, legal and regulatory frameworks can play a vital role in holding internet intermediaries accountable, as their infrastructure and services can be misused for the rapid spread of large amounts of content that carries risk of harm, as well as for inauthentic behaviour such as the use of fake accounts or artificially boosting the visibility of content to expand reach. They also constitute a necessary condition for measures adopted to address online safety concerns to be compatible with human rights of both users and, to the extent relevant, internet intermediaries. Therefore,

legal frameworks should address both the risks posed by specific types of content available online and the role of intermediaries in enabling, amplifying, facilitating and preventing these risks. However, States must also ensure that such frameworks do not result in overcompliance or discriminatory implementation.

81. Paragraph 33 lists the three complementary types of regulatory approaches that States may take to online safety rules. The first category, under letter (a), comprises rules that limit certain types of expression or a manifestation of behaviour, as well as their dissemination, based on their content, and provide for the legal consequences of their production and dissemination. Such content-specific rules are normally not exclusive to the online environment but constitute the extension to online expression of limitations that would also apply in other settings or media. Examples include broadly different types of rules such as those prohibiting child sexual abuse material, incitement to violence, hate speech, defamatory statements or misleading advertisement. Content rules, however, may also be tailored to the online environment, especially when there is a need to provide for specific safeguards in the online space or consequences relating to the nature of the medium used. The second category, under letter (b), comprises rules that indicate the exceptional cases and conditions upon which platforms may be held liable for user-generated content that they store.

82. The principles applicable to these two types of interventions are recalled and detailed in the dedicated subsections of the Appendix. However, paragraph 33 already stresses that only those forms of expression and types of content that have been clearly defined as legally restricted could be subject to restrictions under content rules. It also implies that the conditions under which intermediaries can be held liable for specific instances of legally restricted expression generated by users should be defined by the law.

83. Paragraph 33, letter (c), comprises rules that follow a systemic rather than content-based approach to responsibilities of intermediaries in fostering an enabling online environment. They focus on the processes through which internet intermediaries rank, moderate and remove content rather than on the content itself, as already established in numerous Council of Europe standards (CM/Rec(2022)11 on principles for media and communication governance, Principle 12, and its Explanatory Memorandum (CM(2022)44-addfinal), para 12.4; CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, para. 1.6; Guidance note on countering the spread of online mis- and disinformation, para. 23). This type of legislation should address, in particular by promoting accountability and enhancing empowerment of users, the systemic duties and responsibilities that platforms should have with regard to their own systems and processes, including how they may contribute to risks of online harms.

On paragraph 34

84. Internet intermediaries perform a variety of functions and services (CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, Preamble, para. 4). Platforms in particular have become an inherent part of people's information and communication practices, exerting significant influence over how users produce, access and engage with information and media content. When platforms moderate and rank content, including through automated processing of personal data, they exert forms of control which influence users' access to information online in ways comparable to media, or they may perform other functions that resemble those of publishers (CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, Preamble, para. 5). Nevertheless, despite their power and role, platforms continue to operate as intermediaries. This role, even when it involves the managing of content generated by their users and its curation and selection through algorithmic systems, necessitates a differentiated approach in their governance as opposed to the governance of media actors (CM/Rec(2022)11 on principles for communication and media governance). The key distinction lies in liability: platforms are not generally liable for specific pieces of content generated by their users, unlike media outlets that exercise editorial responsibility and can therefore be held accountable for the content they publish. The absence of editorial

responsibility, however, does not relieve platforms of all duties and responsibilities. As noted by the Court in *Google LLC and others v. Russia*, "when internet intermediaries manage content available on their platforms or play a curatorial or editorial role, including through the use of algorithms, their important function in facilitating and shaping public debate engenders duties of care and due diligence, which may also increase in proportion to the reach of the relevant expressive activity" (cited above, para. 79; see also the Explanatory Memorandum to the section on Intermediary liability rules below). The necessity to differentiate the responsibilities of platforms from those of traditional media, however, does not mean that certain rules that have the same objective cannot be applicable to both platforms and the media. For example, both platforms and the media should have the obligation not to disseminate illegal content. An example of this is the set of obligations imposed on video-sharing platforms by the EU Audiovisual Media Services Directive, Article 28b, to protect minors and the general public from certain types of content that are also restricted for providers of audiovisual media services. Even though actual responsibilities of platforms and the media in this respect differ, the rules applicable to both players share the same objective of preventing harmful consequences for users.

On paragraphs 35-36

85. Paragraphs 35 and 36 reaffirm the fundamental principle of a human rights-based governance framework for internet intermediaries, as set out in the Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression (para. 1.1). State-driven responses to online risks should clearly identify, in their legal frameworks, which content should be addressed through measures that interfere with the right to freedom of expression of users, such as the removal or demotion of content or the suspension or termination of accounts of specific users. This flows directly from the requirements of Article 10, paragraph 2, of the Convention, in particular that any limitation to the exercise of freedom of expression shall be provided by the law. Content that is not legally restricted, especially when such restrictions would disproportionately affect the rights enshrined in Article 10 of the Convention, may nevertheless raise concerns for its potential impact on other human rights and, consequently, on the safety and well-being of users. States should address such content exclusively through other measures that do not interfere with the right to freedom of expression of users and rather focus on mitigating the risks of harm. These notably include the measures for platform accountability and user empowerment set out in the relevant parts of the Appendix.

86. Any policy or action by public authorities that interferes with the right to freedom of expression should be prescribed by law, pursue a legitimate aim and meet the requirements of legal certainty, necessity and predictability (CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, para 1.2). Broad systemic duties and responsibilities should not be construed as requiring the removal or limitation of lawful content, nor should they be implemented in such a way that would impose on platforms content-based restrictions outside the cases provided for by the law. For example, platforms may be required to put in place systemic measures to mitigate the risk of children accessing content that is legally restricted to them, such as age-assurance systems or setting up their recommender systems to

prevent children from being exposed to content recommendation that could pose a risk to their safety and security, particularly when encountered repeatedly. However, they should not be required to remove such content for the generality of their users in order to protect children. Similarly, mitigation measures that may be required by regulatory authorities against the risks to an informed public discourse posed by content that is lawful but purposely misleading should focus on content-neutral criteria, such as the existence of certain patterns in its dissemination of content that may be indicative of a coordinated inauthentic behaviour.

87. Platforms often have internal rules and content policies prohibiting certain types of content even if that content is not legally restricted under domestic law. These policies and rules are typically outlined in the platforms' terms of service agreements or community guidelines which users agree to upon joining,

and they constitute contractual commitments between the platform and its users. While States should not exploit these internal rules to impose de facto removal obligations for lawful expression, they should hold platforms accountable for how these rules are applied and enforced. State oversight in this context should focus on procedural safeguards including transparency, protection against arbitrary decisions and the availability of appeal mechanisms, in order to ensure the protection of rights and freedoms, including, for example, that content is not removed or restricted in ways that unjustifiably interfere with freedom of expression.

On paragraph 37

88. Blocking or banning access to an entire website, domain, or online platform is considered one of the most severe forms of interference with the right to freedom of expression. Such measures not only restrict access to specific unlawful content but often suppress a vast amount of lawful expression. Recommendation CM/Rec(2016)5 on Internet freedom states that any such measure taken by State authorities or any request by State authorities to carry out such actions must comply with the conditions of Article 10 of the Convention regarding the legality, legitimacy and proportionality of restrictions.

89. Wholesale blocking of an entire online platform represents an interference that should be seen as a last resort measure in very exceptional cases. Any content-related concerns should be addressed through evidence-based and proportionate measures, such as targeted removal of illegal content, or by imposing systemic duties and responsibilities to address legitimate concerns, such as the protection and well-being of children. The European Court of Human Rights has addressed this issue in several cases. In *Ahmet Yıldırım v. Turkey* (No. 3111/10, 18 December 2012), the Court found that the decision by Turkish authorities to block the entire Google Sites platform due to one allegedly unlawful page was a disproportionate and unnecessary restriction, which rendered vast amounts of information inaccessible, thus directly affecting the rights of internet users and having a significant collateral effect. Similarly, in *Cengiz and Others v. Turkey* (Nos. 48226/10 and 14027/11, 1 December 2015), the Court ruled that the wholesale blocking of YouTube – based on a few videos deemed illegal – violated the applicants' rights to receive and impart information.

90. Restrictions on access to a service, domain, or website should be considered a measure of last resort, ordered by a judicial authority or another independent public authority whose decisions are subject to judicial review, such as independent regulatory authorities, and applicable only in the most severe cases. This may be the case, for example, when a service hosts exclusively illegal content or is used for the commission of criminal offences involving threats to the life or safety of individuals. Such restrictions should be specific, temporary and imposed only when other measures have failed to produce the desired effect and when the infringement in question continues to cause serious harm. Preliminary steps could be provided for in legislation, such as empowering relevant State authorities to order the cessation of the infringement or to adopt interim measures aimed at preventing the risk of serious harm. Another circumstance in which access to a service may be subject to a general restriction arises where a platform persistently fails or refuses to comply with legal obligations or regulatory requirements that are consistent with human rights, within the framework of platform accountability legislation, provided that such non-compliance results in actual harm. Such harm may include, inter alia, the widespread dissemination of illegal content, such as child sexual abuse material or non-consensual sexual images, or the enabling of access by children to pornographic material.

91. States should provide effective procedural safeguards, including independent judicial oversight, opportunities for appeal or review by both operators and affected users and clear limitations on the discretion of public authorities. The Recommendation CM/Rec(2016)5 on Internet freedom also outlines transparency obligations (para. 2.2.5). It recommends that States publish information about websites that have been blocked or from which information was removed, including details on the legal basis, necessity and justification for such restrictions, the court order authorising them and the right to appeal. Further

guidance on the content of such reports to be published by the States can be found in the Explanatory Memorandum to Recommendation CM/Rec(2022)16 on combating hate speech (CM(2022)43-addfinal - [1434/4.4], para. 107). Meaningful explanations could also be provided by app stores or internet access providers when users search for an app that has been removed or attempt to access a blocked website. Such notices should clearly state the reasons for the restriction and include guidance for affected users on how to seek remedy.

On paragraph 38

92. As required by Article 10 of the Convention and re-stated in the Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, any request, demand or other action by public authorities addressed to internet intermediaries that interferes with human rights and fundamental freedoms shall be prescribed by law. States should not exert pressure on internet intermediaries through non-legal means. The same principle should apply to actions targeting content creators, particularly where such measures could negatively affect their freedom of expression. Any such powers granted to public authorities, including law-enforcement authorities, should be clearly defined to protect against arbitrary application.

93. Imposing measures on internet intermediaries and content creators that affect the availability of content without a clear legal basis is particularly concerning during times of crisis and extraordinary circumstances that could lead to a threat to public security or public health, such as conflicts, civil unrest, acts of terrorism, natural disasters or public health emergencies. Recommendation CM/Rec(2024)7 on the effective protection of human rights in situations of crisis provides a general framework for States on how to exercise their powers in such situations, and recalls that "Member States should safeguard freedom of expression and the public's access to accurate and reliable information in situations of crisis" (para. 16). Requests and measures in response to such crises should be taken only where, and to the extent that they are strictly necessary and urgent, taking into account that such measures should be proportionate to both the severity of the situation and their potential implications on the rights and interests, including the freedom of expression. Furthermore, such measures should apply only for a limited period of time as prescribed by law. Wherever appropriate, crisis-related measures should be based on heightened due diligence obligations for platforms, as envisaged under the Crisis response mechanism in Article 36 of the EU Digital Services Act. Examples include adapting content moderation processes, e.g. by increasing resources for content moderation, intensifying cooperation with trusted flaggers, implementing awareness-raising measures and promoting trusted information (see, for example, EU Digital Services Act, Preamble, recital 91).

94. States should ensure that legislation contains safeguards against any misuse of power, such as administrative authorities issuing orders or applying other forms of regulatory pressure on internet intermediaries without a legal basis. Legislation should also provide intermediaries and users with the right to an effective remedy in cases where such pressure is exerted.

On paragraph 41

95. This paragraph reinforces the graduated and proportional approach to platform regulation emphasised by the Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries, according to which the scope and approach taken to fulfil platforms' responsibilities may vary based on the risk of harm and potential severity of their impact on human rights (para. 2.1.2). Micro and small platforms typically operate with limited resources and exert minimal influence over public discourse. Subjecting these actors to the same regulatory requirements as platforms operating globally, with advanced technological capabilities and larger economic resources, risks stifling growth, discouraging new entrants and undermining the broader goals of a diverse and competitive digital ecosystem.

96. Platforms of significant influence are defined by their capacity to shape the information environment and to affect the enjoyment of human rights. As such, they should be subject to stricter due diligence obligations and more robust regulatory oversight, such as risk assessments and heightened transparency reporting requirements.

97. By providing alternative criteria to define the different categories of platforms, paragraph 41 leaves to States a wide margin in deciding the appropriate criteria to graduate the responsibilities of different platforms. It allows the use of quantitative criteria such as the user base, as it is the case in the EU Digital Services Act for the designation of very large online platforms and search engines (Art. 33). The regulatory framework may also use qualitative criteria for this purpose, such as the level of platforms' societal impact or risk level. The criteria upon which the risk level of different internet intermediaries is assessed should be specified clearly, reviewed periodically, measured precisely and communicated transparently by the competent State authority (CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, para 1.4).

98. The general exemption of intermediaries from liability for user-generated content and its exceptions, as detailed in paras. 54-56 of the Appendix, should always apply to all internet intermediaries, including platforms irrespective of their size, reach or impact. However, when liability exemptions exceptionally do not apply in the specific circumstances examined by the Court in the case *Delfi AS v. Estonia* [GC] (cited above, paras 115-116, see also Explanatory Memorandum to paragraph 55), the type and size of the provider may be decisive factors in limiting such liability and in avoiding disproportionate requirements on them (*Pihl v. Sweden* (dec.), No. 74742/14, 7 February 2017, para. 31; *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, No. 22947/13, 2 February 2016, para. 82).

On paragraph 42

99. Paragraph 42 reiterates a principle enshrined in the Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries. States should not impose any action that may lead to the obligation of intermediaries to systematically monitor user activity or to actively seek facts or circumstances that would indicate illegal activity, for instance by requiring the use of automated systems to proactively scan user data or content for legal infringements. However, internet intermediaries should be able to voluntarily carry out their own-initiative investigations in relation to illegal content. This proactive conduct should not, in itself, result in their liability.

On paragraph 43

100. In the digital age, the regulation of online safety cannot be effectively achieved through isolated national measures. Given the cross-border nature of the internet, States should cooperate to ensure that their legal and regulatory frameworks are firmly grounded in international human rights standards and principles and are as coherent as possible. This is especially crucial in regions where countries share linguistic, cultural or political contexts, as they often face similar online risks and societal impacts.

101. A common approach to regulation and cooperation in the enforcement of rules is essential both from a market perspective and in terms of legal certainty. Aligned rules can promote better compliance by digital platforms operating across multiple jurisdictions. Fragmented regulatory approaches, on the other hand, may result in inconsistent enforcement and uneven protection for users. For example, diverging laws can lead to forum shopping, where platforms choose to operate under the most lenient jurisdictions, thereby undermining protective standards.

102. Active engagement within relevant intergovernmental bodies, transnational networks and other international organisations should also be fostered in order to ensure alignment with relevant European and international instruments (see also CM/Rec(2022)16 on combating hate speech, para 63).

103. Existing treaty frameworks for online criminal behaviour and cooperation in the investigation and prosecution of such crimes, such as the Convention on Cybercrime and its two Additional Protocols (ETS No. 189 and CETS No. 224), provide examples of enhanced intergovernmental cooperation. These instruments aim to ensure common approaches to illegal content and effective cooperation in their cross-border implementation, while upholding human rights.

104. States may also establish mechanisms for cross-border regulatory coordination which may include protocols for joint investigations, shared technical standards and, where appropriate, coordinated enforcement actions.

Content rules

On paragraph 44

105. In accordance with Recommendation CM/Rec(2022)11 on principles for media and communication governance, the promotion of human rights and fundamental freedoms in communication "entails aligning rules for the offline and online environments" (principle 6). The Explanatory Memorandum thereof clarifies that this requires "taking into account differences but avoiding stricter regulation of content disseminated via platforms" (para. 6.2). As underlined by the Venice Commission, for example, "the establishment of a specific regime only applicable to electronically distributed versions of the written media generates a different legal treatment between identical content. ... Any distinction between legal regulations applicable to printed press, to online press and to the broadcasting media should be justified." (CDL-AD(2020)013, Albania - Opinion on the draft amendments to the Law n°97/2013 on the Audiovisual Media Service, para. 26). The online sphere should therefore not be subject to arbitrary or disproportionate restrictions that would not be permissible offline. However, online platforms and digital communication present unique challenges and introduce new risks due to the scale and speed of dissemination, the potential for anonymity, or algorithmic amplification. These factors may require tailored governance measures to mitigate harm, as discussed below.

On paragraph 45

106. Content rules are aimed at restricting the publication of certain types of expression or manifestation of behaviour or their dissemination or accessibility. Therefore, they directly affect the ability of individuals and groups to receive and impart information, including through digital technologies. In line with the principles articulated in Recommendation CM/Rec(2022)11 on principles for media and communication governance, paragraph 45 highlights the requirement of transparent and evidence-based rule-making, which is essential for meeting the necessity and proportionality tests under Article 10, paragraph 2, of the Convention. Such evidence should confirm the existence of clear interferences with rights that need to be addressed through content-specific restrictive measures. As stated in paragraph 1.1.4. of the Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, adoption of legislation or regulations should be preceded by human rights impact assessments.

107. Transparent evidence-gathering implies that States, when defining policies, legal and regulatory frameworks containing content rules, should seek the participation of a range of non-State actors, including civil society and the research community, as well as support for independent research aimed at understanding and adequately addressing particular risks.

108. A proportionate approach to restricting content requires that States define the types of content to be restricted narrowly and with legal clarity, choose remedial actions that are tailored and minimally intrusive, and apply restrictions only within the necessary geographic scope, in order to avoid extraterritorial effects.[32]

On paragraph 46

109. Article 10, paragraph 2, of the Convention requires any interference with the exercise of freedom of expression to be “provided by the law”. This paragraph of the Appendix addresses the principle of legality, from the point of view of the “quality” that such law must possess. According to the Court, a norm cannot be regarded as a “law” unless it is formulated with sufficient precision to enable citizens to regulate their conduct and that they must be able – if need be with appropriate advice – to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail. At the same time, the law can leave space for interpretation and absolute clarity is not required (*Perinçek v. Switzerland*[GC], No. 27510/08, 15 October 2015, para. 131; *Sanchez v. France*[GC], No. 45581/15, 15 May 2023, paras 125-126). Only the type of content that has been defined by law in a sufficiently clear and precise manner should be subject to restrictions that interfere with freedom of expression.

110. The law must be accessible and foreseeable, meaning that individuals must be able to understand what is prohibited and what the consequences are. Such legal clarity and predictability benefit all parties involved. State authorities responsible for ordering the removal or blocking of legally restricted content should be able to clearly and unambiguously determine whether specific content falls within their mandate, which enables timely and effective enforcement. Intermediaries are better positioned to comply with legal obligations when they understand exactly what is expected of them. This reduces the risk of overremovals taken as a precautionary measure. Finally, users benefit from greater transparency and legal protection, while their rights and freedoms are less likely to be unjustifiably restricted. By contrast, vague or overly broad definitions open the door to subjective interpretations, which can result in arbitrary enforcement and disproportionate and unjustified hindrances to freedom of expression (CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, para 1.3). For example, the Venice Commission has stated that “the prohibitions on the dissemination of information based on vague and ambiguous ideas, including ‘false news’ or ‘non-objective information’, are incompatible with international standards for restrictions on freedom of expression ... and should be abolished” (CDL-AD(2019)016, Joint Report on Digital Technologies and Elections, para. 90).

111. The Court has emphasised that the scope of the concepts of foreseeability and accessibility depends to a considerable degree on the content of the legal instrument, the field it is designed to cover and the number and status of those to whom it is addressed. Content rules governing the general conduct of users should therefore be formulated with a higher degree of clarity and precision than those addressed to professional content creators or platforms; professionals are expected to be better aware of their responsibilities under the law and to take special care in assessing the risks that such activity entails (*Chauvy and Others v. France*, No. 64915/01, 29 June 2004, paras 43-45). In particular, for criminal law provisions (such as those related to hate speech) the Court requires that the scope of the relevant offences must be defined “clearly and precisely”, so to avoid a situation where the State’s discretion to prosecute for such offences becomes too broad and potentially subject to abuse through selective enforcement (*Savva Terentyev v. Russia*, cited above, para. 85; see also *Altuğ Taner Akçam v. Turkey*, No. 27520/07, 25 October 2011, paras. 93-94). Similarly, content subject to the most severe restrictions, such as removal or blocking, should be defined with the highest level of legal clarity and precision.

On paragraph 47

112. The principle of proportionality requires restrictions to be tailored to the seriousness of the harm caused, or likely to be caused, by different types of content. Domestic law, including when so required by obligations arising out of the Convention or other international human rights instruments, should require platforms to remove or block the most serious forms of harmful content, such as child sexual abuse material, terrorist or violent extremist propaganda, or speech inciting imminent violence or hatred, which should be prohibited under criminal law. These measures are justified by the serious harm, or imminent risk thereof, caused by the material concerned. The removal process should be in line with the principles of legality, necessity and proportionality, and comply with international human rights standards, and the principles outlined in Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet

intermediaries. Moreover, in accordance with paragraph 2.3.6 thereof, when platforms remove or block content that is illegal, they should ensure that evidence is retained for effective criminal law investigations. Restrictions, including blocking or removal, may also be justified in cases where content violates the provisions of administrative or civil law not limited to the online environment, following the principle that what is illegal offline should also be illegal online. This would be the case, for example, with copyright infringements or with content amounting to illegal hate speech that does not reach the threshold for criminal liability (see CM/Rec(2022)16 on combating hate speech, para. 3). Finally, restrictions and specific content-related rules may be imposed on certain regulated services, such as audiovisual media services, as

provided for in the specific sectoral legislation (see also Explanatory Memorandum on paragraph 53). However, States need to make sure that content moderation responses are adapted to the type and nature of legally restricted content in question, as well as the specific problem they are trying to solve, taking into consideration that legally restricted content varies greatly in terms of their characteristics and the gravity of their consequences,^[33] and that there are safeguards against unjustified restrictions on freedom of expression in place.

113. Certain types of content, while not illegal in all circumstances, may still pose potential harm depending on the context. In these cases, risks related to some types of legally restricted content might be effectively mitigated by less intrusive measures. Such content, whose identification by domestic law remains subject to the principle of legality and legal certainty, qualifies as legal but regulated content. Examples include: age-restricted access for violent or pornographic content which helps preventing children from being exposed to it without removing content from adult audiences; transparency and targeting rules for political advertising or commercial promotion of certain products and services (e.g. gambling or alcohol); and reduced visibility in search results or recommendation feeds reducing the spread and prominence, or demonetisation which removes financial incentives for it, of otherwise lawful, content that produces risks of harm, such as some forms of disinformation. These measures do not suppress the content but limit the harmful effects of its unregulated distribution. In this respect, attention should also be paid to the existence of accountability legislation that establishes systemic duties and responsibilities for platforms as part of an empowering platform governance framework, as described in paragraphs 57 and 58 of the Appendix and paragraphs 149 to 152 of this Explanatory Memorandum.

114. Restrictions on access to legal but regulated content must always be assessed on a case-by-case basis. They should not be presumed to be less intrusive or less impactful than measures applied to illegal content. For example, de-ranking lawful content can make it virtually invisible. Such actions may have a chilling effect on expression comparable to outright removal. Therefore, before imposing any restriction on content, State authorities should carry out a thorough, evidence-based assessment of both the expected benefits and the possible unintended consequences. This includes evaluating how the proposed measure might impact freedom of expression, media pluralism, access to information and other human rights. Authorities must also consider the broader effects on communities and the online environment. Once a measure is implemented, its effectiveness and proportionality should be regularly monitored and reassessed to prevent overreach, abuse or collateral overremoval of lawful content. In line with the principle of proportionality, the least restrictive and most targeted measure capable of achieving the legitimate aim should always be preferred.

115. States should periodically review their content-related laws and regulatory frameworks to ensure their clarity, effectiveness and alignment with current societal, technological, and legal developments. The review should relate both to their overinclusiveness and underinclusiveness. These reviews should be informed by transparent, independent evidence-gathering and inclusive consultations involving civil society, researchers, intermediaries and affected communities. Moreover, States should support, including financially, independent research into online risks of harm, content risks and the evolving digital landscape, in order to adapt policy responses to emerging challenges and better target existing measures.

On paragraphs 48-49

116. In line with the principle of legality and legal certainty, public authorities, when enforcing content-specific restrictions, including by requiring intermediaries to restrict access or remove content, should only do so for content that falls within the categories of legally restricted content under domestic law, and within the limits conferred by it. The law should also provide safeguards against the selective, discriminatory or arbitrary use of any powers that require the restriction of online content. Before and after ordering the enforcement of a restriction, public authorities should carefully evaluate the possible impact of such action,

including on freedom of expression, and should apply the least intrusive measure available (CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, para. 1.3.1.).

117. Paragraph 49 reaffirms the principle set out in paragraph 1.3.2. of Recommendation CM/Rec(2018)2, according to which State authorities should obtain a formal order by a judicial authority or other independent administrative authority, such as an independent regulator, whose decisions are subject to judicial review, when demanding intermediaries to restrict access to content. Notices received from other public authorities should not be automatically considered as legally binding orders obliging intermediaries to comply. In these cases, however, failure to take action may also have consequences for the regime of liability of intermediaries (see Explanatory Memorandum to paras 54-56). Whenever domestic law gives a public authority the power to issue "notices" that trigger legal consequences other than applying a restriction, such as the obligation to give priority to the examination of the legality of a given piece of content, these should be treated as "orders" for the purpose of this recommendation. The mandate and the scope of authority granted to public authorities empowered to issue such orders should be clearly defined in law.

118. The use of the term "in principle" indicates that there may be situations in which the seriousness or imminency of harm may call for swift interventions by authorities other than those mentioned in paragraph 49. Paragraph 1.3.2. of Recommendation CM/Rec(2018)2, for example, indicates that the above requirement does not apply in cases involving content that is clearly unlawful, such as child sexual abuse material, or in cases where expedited measures are required in accordance with the conditions prescribed in Article 10 of the Convention. The cases in which public authorities other than judicial or independent ones may order restrictions to content should be defined by the law.

119. Any order requiring restrictions should clearly state its legal basis and provide reasons that are relevant and sufficient to assess its conformity to the law and, when relevant, indicate how the impact of the measure on the freedom of expression of the content producer or issuer has been considered in line with the requirements of Article 10 of the Convention. A judicial remedy which is accessible, sufficiently prompt and capable of providing redress should be available to the intermediary that is the addressee of the order, as well as to any user whose freedom of expression, including the right to receive information, is affected by the measure.

On paragraph 50

120. Platforms may impose restrictions on lawful user-generated content through their own contractual policies and rules. In some cases, such content-related restrictions may be introduced as measures intended to prevent or mitigate risks, based on the obligation to carry out risk assessments as detailed in paragraph 60 of the Appendix. However, it should be made clear that any such restrictions on lawful content derive solely from the platforms' own policies and rules and are not required by the law.

121. The contractual rules and policies of platforms should be developed in dialogue with users and user communities, and with their meaningful input. This is essential to ensure that they provide effective protection against harms they face and do not go further than necessary in restricting users' rights and freedoms. This includes reflecting how specific risks are manifested, as well as how the platform's own

design and functionalities, including the use of automated systems, may influence users' exposure to risks of harm. The UNESCO Guidelines for the governance of digital platforms identify human rights due diligence as one of their core principles, mandating "meaningful engagement with a variety of stakeholders to identify specific risks for groups in situations of vulnerability and marginalization".^[34] The Guidelines emphasise that digital platforms should also be open to expert and independent input on how these assessments are structured.

122. States should ensure that platforms regularly and systematically review and assess the impact of their policies, systems and practices, including how they address new and evolving risks. These assessments should also be conducted prior to introducing design changes or new policies. They should

ensure participation of relevant stakeholders, particularly those most affected by potential risks. Furthermore, they should be transparent and easily accessible, in line with paragraph 88 of the Appendix.

On paragraph 51

123. This paragraph directly tackles the risks resulting from platform's inaction towards users and content creators who clearly abuse their rights to publish information and to submit notices. It covers two distinct cases, concerning users and content creators who: act in egregious violation of the law, by systematically disseminating illegal or other legally restricted content; and systematically abuse their right to submit notices, in particular when it can be observed that they do so in order to harass or silence other users.

124. The recommended action includes restrictive measures such as demotion and demonetisation of all their content and, in the most serious cases, suspension and termination of their present and future accounts. Such action, however, has very serious consequences for the right to freedom of expression of the targeted user and the right of other users to receive information and carries the risk of collateral censorship. Therefore, the utmost care is called for in their application. It should be reserved only to cases in which the intent to spread legally restricted content is evident, the action is systematic and other less restrictive measures have failed to produce a change in the user behaviour.

125. The procedural safeguards against arbitrary application, as detailed in paragraphs 90 to 93 of the Appendix, apply: they include the right of the affected users to be duly informed and have access to fair and effective out-of-court remedies.

On paragraph 52

126. As reaffirmed by the Court in *Google LLC and others v. Russia* (cited above, para. 90), Article 10 of the Convention also protects the right to not be compelled to express oneself (*Gillberg v. Sweden* [GC], No. 41723/06, 3 April 2012, paras 85-86) and the right to be silent (*Kobaliya and Others v. Russia*, Nos. 39446/16 and 106 others, 22 October 2024, para. 84). Therefore, any legal provision or administrative action that has the effect of compelling platforms to host specific content constitutes an interference with their rights under Article 10 of the Convention and requires justification in accordance with its paragraph 2: the measure must be prescribed by law, based on a clear, accessible and foreseeable legal framework; it must pursue a legitimate aim; it must be necessary in a democratic society, which requires demonstrating a pressing social need, supported by relevant and sufficient reasons, and a proportionate response to the aim pursued.

127. The concept of a "pressing social need" demands that the measure respond to a genuine and weighty public interest. Such a need must be supported by relevant and sufficient reasons, including clear and substantiated evidence demonstrating that the measure is necessary to serve that public interest. Examples might include providing information during a public emergency or mandating content whose unavailability would significantly reduce media pluralism and undermine the diversity of viewpoints essential for democratic discourse. However, such issues should primarily be addressed through less

intrusive measures, in particular systemic duties and responsibilities of risk mitigation and user empowerment placed on platforms, rather than by imposing *ad hoc* obligations to publish specific content or information.

On paragraph 53

128. Recommendation CM/Rec(2011)7 on a new notion of media recommends that States adopt a broad notion of media, encompassing “all actors involved in the production and dissemination, to potentially large numbers of people, of content ... while retaining ... editorial control or oversight of the contents”. Recommendation CM/Rec(2022)11 on principles for media and communication governance reiterates that “media and communication governance should aim to ensure that the media, individual journalists and others comply with content obligations in accordance with Article 10 of the Convention and with professional standards” (principle 10).

129. New media actors, such as bloggers, vloggers, influencers, podcasters, citizen journalists and other independent digital voices, are having an increasing impact on online content and the flow of information. They typically operate independently, retaining editorial control over their content, but outside formal editorial oversight typical of traditional media. Given the audience shift towards online sources for information and entertainment, content creators operating outside the framework of traditional media outlets

have a growing relevance in how people receive and access information, including on news, and a wider impact in shaping values and opinions. Their reach, visibility and authenticity-based relationship with their audience give them considerable capacity to affect democratic attitudes, values and political opinions, as well as health, personal attitudes and career decisions. While they can have positive effects, they can also produce harmful effects by disseminating misinformation and disinformation, hate speech, or discriminatory content. In particular, when operating on platforms that are particularly popular among children, they can exert a significant influence over them and their well-being, raising concerns for otherwise lawful messages such as unrealistic portrayals, the promotion of unhealthy habits, and risks to mental health. Therefore, their behaviour and practices have a direct impact on the quality and safety of the online environment.

130. As already noted, certain categories of content creators may be subject to content-related legal obligations, especially in the framework of audiovisual media and consumer protection legislation. Other professionals, when they act as content creators, may also be subject to existing self-regulatory ethical and professional frameworks. This may be the case for lawyers, doctors or professional journalists, but also other categories such as researchers, teachers and, more generally, civil servants (see Explanatory Memorandum to paragraphs 30-32). In this framework, paragraph 53 addresses two recommendations to States.

131. Firstly, States should ensure that all professional content creators, as defined above (see Explanatory Memorandum to paragraph 11, second indent), are under an obligation to disclose information on how they monetise their content. This notion should be understood broadly, and not be limited to monetisation through views, membership and subscriptions, sponsorships and partnerships or advertisement and product placement, but should encompass all commercial strategies and practices that seek to obtain economic gain from the content, such as redirection to other platforms or websites to sell products and services.

132. Secondly, States are called to promote the development of self-regulatory frameworks, especially for those categories of content creators that fall outside the scope of existing regulation and self-regulation. One example may be that of creators who, while not being registered as a media service

provider or a professional journalist in accordance with domestic law, effectively act as such. Well-designed self-regulatory frameworks can benefit both content creators and their audiences, contributing to a safer and more trustworthy online media landscape.

133. Self-regulatory frameworks can play a key role in helping content creators uphold ethical and professional standards, including those aimed at protecting children, improve the reliability of information and build public trust. These frameworks should be transparent, with publicly accessible rules and procedures, inclusive of diverse voices and content types, and grounded in human rights. Rather than focusing on restricting lawful expression, such mechanisms should emphasise accountability, fairness and harm prevention. This includes clear principles on issues like transparency of sponsorships, responsible content amplification, respectful engagement and correction of errors. Content creators should guarantee that the content they provide complies with the relevant content obligations to protect vulnerable groups, especially children, from harm and self-regulatory initiatives or internal compliance procedures should provide for the provision of age ratings, independent classification of content prior to dissemination and handling complaints (see Explanatory Memorandum to CM/Rec(2022)11 on principles for media and communication governance, para. 10.5).

134. Useful guidance in this respect can be found in the Conclusions of the Council of the European Union on support for influencers as online content creators. The document outlines various measures including strengthening media literacy skills among content creators, facilitating policy dialogues with their representative organisations and influencer agencies, supporting the development of self-regulatory bodies or mechanisms such as ethical codes and involving content creators in the development of media policy measures that may affect them.

135. Self-regulatory frameworks should be promoted without prejudice to the possibility for States to include content creators under statutory media regulation: if independent self-regulation mechanisms are lacking or ineffective, or if the public interest requires a stronger involvement of the State as guarantor of these interests, States should not be prevented from adopting appropriate and proportionate co-regulatory frameworks (Explanatory Memorandum to CM/Rec(2022)11, para. 10.5).

Intermediary liability rules

On paragraph 54

136. In *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary* (cited above, para. 86), the European Court of Human Rights has cautioned against expanding excessively the liability of intermediaries, which could have a chilling effect on free expression. More recently, in *Google LLC and others v. Russia* it has observed that heavy penalties imposed on a platform for failing to comply with broadly framed takedown orders “placed an excessive burden on intermediaries ..., effectively compelling them to act as censors of political speech on behalf of the State authorities, an approach incompatible with the Court’s approach to freedom of expression” (cited above, para. 79). Inspired by these concerns and wording, paragraph 54 cautions against the risk that intermediaries, if held disproportionately liable for user-generated content, may pre-empt possible liability by removing content whenever the slightest doubt about its lawfulness arises. This practice, often referred to as overblocking or collateral censorship, does not come in response to the legally restricted nature of the content itself, but rather from the intermediaries seeking to minimise legal risks, avoiding fines or reputational harm. To prevent overblocking, States should ensure that liability frameworks are proportionate, clearly defined and targeted, and that internet intermediaries can operate in a regime of sufficient legal certainty.

On paragraph 55

137. Paragraph 55 reiterates the principle enshrined in para. 1.3.7 of the Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries.[35] Internet intermediaries facilitate the transmission, access, or storage of content on behalf of users. In recognition of this role, they may not, as a general rule, be held liable for third-party content they do not create but merely provide access to, transmit or store. State authorities may hold intermediaries liable with respect to individual pieces of user-generated content they store only when two conditions are met: (1) the intermediary has specific knowledge that the content is legally restricted; (2) the intermediary fails to act promptly to restrict access to the impugned content. States should ensure that conditions, including any time frames if and when appropriate, for the removal of legally restricted content or the enforcement of other restrictions are established by law.

138. Intermediaries may become aware of the legally restricted nature of the content through sufficiently substantiated notices by users, professional user groups or public authorities, but also as a result of their own investigations or while complying with their legally mandated duties on online safety and accountability. Notice-based procedures should be transparent, accessible and effective. State authorities should ensure that such procedures are not designed in a manner that incentivises the takedown of lawful content (CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, para. 1.3.7). Legislation should also make clear that internet intermediaries cannot be held liable merely on the basis of general awareness that their service is being used to store or disseminate legally restricted content or because they voluntarily undertake, in good faith and diligently, actions aimed at detecting and acting against legally restricted content they may host.

139. The action taken by the intermediary after acquiring knowledge should consist in assessing promptly and in good faith whether the content is legally restricted and, if so, restrict access to it or remove it, as appropriate. Legal frameworks should also clarify that internet intermediaries will not be held liable for choosing not to remove content based on a good-faith, fact-based and legally sound assessment, even if the content is later qualified by competent authorities as being in breach of criminal, civil or administrative law (see also Explanatory Memorandum to CM/Rec(2022)16 on combating hate speech). When the law sets specific time frames for action, these should always relate to a specific category of content and take into account its nature and severity. Content whose dissemination carries the most imminent risk of harm, such as child sexual abuse material or direct incitement to violence, should be subject to strictest conditions, including shorter time frames for removal. Inappropriately short timelines for acting on notices concerning content falling into broadly defined categories that do not carry an imminent risk of harm may incentivise overremoval, including the takedown of lawful content.

140. In assessing, under Article 10 of the Convention, whether an internet portal operator may be held liable for the failure to remove comments posted by a third party, the European Court of Human Rights has identified four criteria with a view to striking a fair balance between the right to freedom of expression and the rights of the aggrieved person or entity, namely: the context and contents of the comments; the possibility to hold the authors of the comments liable; the measures taken by the intermediary to prevent illegal content and the conduct of the aggrieved party; the consequences for the aggrieved party and for the intermediary (*Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, cited above, para. 60 ff; *Delfi AS v. Estonia* [GC], cited above, para. 142 ff.).

141. Based on these criteria, the Court has not ruled out the possibility, in exceptional and specific cases, to hold certain internet intermediaries liable "if they failed to take measures to remove clearly unlawful comments without delay, even without notice from the alleged victim or from third parties" (*Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, cited above, para. 91). In *Delfi AS v. Estonia* [GC] (cited above), the Court held that Article 10 of the Convention did not prevent ordering a large professionally managed internet news portal run on a commercial basis – which publishes news

articles of its own and invites its readers to comment on them – to pay damages for anonymous extreme comments, which amounted to illegal hate speech or incitement to violence, to an article posted on its website. The Court also underlined that the case did not concern “other fora on the Internet where third-party comments can be disseminated, for example an Internet discussion forum or a bulletin board where users can freely set out their ideas on any topic without the discussion being channelled by any input from the forum’s manager; or a social media platform where the platform provider does not offer any content and where the content provider may be a private person running the website or blog as a hobby” (para. 116). In subsequent cases, having regard, *inter alia*, to the absence of illegal content of extreme gravity, such as illegal hate speech or any direct threats to physical integrity, in the user comments in question, the Court found that liability of internet portals for third-party comments without actual knowledge of their illegality was not compatible with Article 10 of the Convention (*Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary*, cited above, para. 91; *Pihl v. Sweden* (dec.), cited above, para. 32; *Tamiz v. the United Kingdom* (dec.), cited above, para. 84; *Høiness v. Norway*, No. 43624/14, 19 March 2019, paras 73–74).

142. The case law of the Court does not imply that imposing liability on this basis would be necessary to strike a proper balance between the different rights at stake, nor that its reasoning allowing the removal of the liability exemption would apply to other types of intermediaries. It merely indicates that in certain exceptional cases, imposing liability on a specific type of intermediary who does not have actual knowledge of the illegality of specific pieces of user-generated content may not violate the rights of that intermediary under Article 10 of the Convention. The lack of diligence by the intermediary in adopting reasonable and appropriate content moderation measures is one of the necessary conditions for such a finding. Therefore, when legal frameworks are adopted to provide public oversight of platforms’ diligence in content moderation, in line with the present Recommendation, the adequacy of the preventive measures taken should be addressed through those platform accountability mechanisms.

On paragraph 56

143. Due process obligations must be in place to safeguard against arbitrary or disproportionate requests or action by public authorities. As stated in Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, legislation should clearly define the powers granted to public authorities, particularly when exercised by law-enforcement authorities (para. 1.2.2). Transparency obligations are another critical safeguard against overremoval. This could involve the responsibility of State authorities, when issuing content-related requests to intermediaries, to clearly state the legal basis and provide substantiated reasons for their requests. Additionally, States should publish information on the number, nature and legal basis of content restriction requests submitted to intermediaries (*ibidem*, para. 1.2.3).

144. A key safeguard in this context is ensuring that users are provided with sufficient information to challenge the actions undertaken by intermediaries in response to removal, blocking, or other content-related orders. First, intermediaries should be required to provide clear, easily accessible and meaningful information wherever a decision is made to restrict content, whether based on a court or administrative order, a notice submitted by an individual user or a user group, or the enforcement of platform’s own terms and policies. Explanations should be communicated directly to the affected users, in concise and simple language understandable to them, and should include the legal basis on which the decision is grounded.

Second, affected users should be clearly informed about all available redress options, including the platform’s internal complaint-handling system, external appeal mechanisms, independent regulatory bodies and the possibility to directly appeal the decision to judicial review (see also Explanatory Memorandum to paragraphs 89–92 below). The provision of such information may however be delayed, as provided by the law, when necessary, in the context of ongoing judicial proceedings.

145. States should refrain from creating conditions that pressure internet intermediaries into removing or blocking content more extensively than legally required. As noted above, ambiguities in the definitions of content subject to restrictions, especially when combined with additional limitations, such as unjustifiably short removal time frames, often do not allow for in-depth assessment in complex cases. Provisions should therefore be in place to discourage pre-emptive and hasty removals of potentially lawful content.

146. Internet intermediaries should be allowed to apply provisional measures in certain unclear and complex cases that require detailed factual investigation or legal analysis, for example by referring them for further assessment by independent bodies, including regulatory authorities and co-regulatory or self-regulatory bodies. Provisional measures may also include de-prioritisation or contextualisation of the content until a final decision is made by the appropriate platform body or other mechanism. De-prioritisation means giving lower priority to content and thus limiting its dissemination, whereas contextualisation involves published content with a note indicating it could constitute legally restricted content.

147. The Appendix reaffirms an important principle set out in Recommendation CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, which is that States remain ultimately responsible for ensuring the protection of human rights. They cannot transfer or delegate this obligation to private entities, be it internet intermediaries or third parties to which they have delegated certain responsibilities in content reporting or dispute resolution. All regulatory frameworks should include effective and independent oversight mechanisms in order to make sure that any delegated duties align with human rights standards.

148. Certain issues may be effectively managed through co-regulatory frameworks, particularly where industry adopts codes, adherence to which may be used to demonstrate compliance with legal obligations. For example, voluntary codes of conduct under the EU Digital Services Act form a set of commitments that may serve as a mitigation measure tailored to specific systemic risks, such as those to civic discourse and election integrity, for which information integrity is key, or specific challenges of tackling legally restricted content, such as illegal hate speech. However, any co-regulatory mechanism must be grounded in a legal framework established by the State, which defines its scope, ensures accountability and includes safeguards against arbitrary decisions by non-State actors. The Guidance note on content moderation outlines essential principles in this regard that should be used by States as a guide in designing and enforcing their co-regulatory approaches. It also points to pitfalls of self- and co-regulatory approaches driven by government pressure, concluding that "any policy interventions that have the purpose of minimising risk...should also have clear targets, adjustment mechanisms and supervision, meaningful protection for freedom of expression, as well as tools to identify counterproductive impacts" (pp. 26-28).

Platform accountability and user empowerment rules

On paragraphs 57-58

149. An enabling online environment is one that supports meaningful user participation, protects individuals from harm, and provides users with agency and control over their online experiences. Such an environment is essential for the realisation of digital rights, including freedom of expression, access to information and participation in public discourse. However, structural issues within online platforms such as opaque algorithms, unsafe design features and inconsistent content moderation can create conditions that undermine rather than enable these rights, especially for marginalised and vulnerable groups.

150. Given the growing influence of online platforms in shaping how people access information and exercise their rights, it is crucial to ensure that users are equipped to understand, mitigate and respond to online risks. As emphasised in the Guidance note on countering online mis- and dis-information, the

integrity of online information requires a holistic strategy. The impact of “low-quality” content, such as misinformation, ultimately depends on whether users (a) are regularly exposed to, and recognise the importance of, high-quality content; (b) are capable of distinguishing between high-quality and low-quality content; (c) act responsibly towards others in sharing and discussing different kinds of information they may encounter; and (d) enjoy strong safeguarding protections for their human rights, know how to exercise them, and are confident they can make a positive difference, as well as protecting themselves, by exercising them. Legislation should therefore place user empowerment at the core of online safety and platform accountability frameworks. Empowerment measures should provide users with tools, processes and systems to manage their online experience, have an understanding of how content is presented, curated and recommended by platforms, the ability to make informed decisions and challenge content moderation outcomes, as well as mechanisms to safeguard their rights.

151. Users should be empowered to actively participate in the online environment. However, this empowerment should not come at the cost of overburdening them with excessive responsibilities to safeguard their own rights (CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, paragraph 1.7). In addition, empowerment duties should not be viewed as a substitute for broader responsibilities of platforms to address online risks of harm (see also CM/Rec(2022)11 on principles for media and communication governance, Principle 15). Certain key empowerment duties should therefore be legally mandated and subject to oversight and should not be left to the discretion of platforms, including through measures applied as part of self-regulatory arrangements.

152. There is a growing global consensus that protecting rights online requires proactive and structural measures, not just reactive content moderation. As previously elaborated (see Explanatory Memorandum on paragraph 33), legal frameworks should address the systemic duties and responsibilities that platforms should have with regard to their own systems and processes, rather than imposing liability for hosting specific pieces of legally restricted content generated by users, in cases other than those referred to in paragraph 55 of the Appendix. This means that legislation should establish overarching obligations addressing how the structural and operational features of platforms, such as their business models, terms and conditions and their enforcement, design settings, algorithmic and advertising systems, content curation and moderation processes, or transparency practices contribute to various risks of online harm. Embedding systemic duties and responsibilities into the platform governance framework aligns with the principle of fostering an enabling online environment, as this approach places responsibility on them to proactively identify, assess and mitigate such risks, both current and evolving, including through measures that empower users and support their agency.

On paragraph 59

153. States should require online platforms to embed user safety into the core architecture of their services from the outset, rather than addressing risks only after harms occur. This means ensuring that harm prevention is integrated into the design, development and deployment of platform features, aligning system architecture with users’ safety and rights. Furthermore, design should be subject to regularly implemented risk mitigation measures, in order to address risks of harm arising from design features that facilitate the amplification of content or behaviour. Algorithmic amplification, particularly through recommender and advertising systems, can significantly extend the reach of content that carries risk of harm, including legally restricted content. Likewise, user interface design features such as auto-play, nudges or dark patterns can manipulate users or create risk-prone environments. Core elements of a user safety by default and by design approach may include technical design choices which limit the amplification of legally restricted content through recommender systems, default to stricter privacy settings for minors and disable profiling-based content or ad targeting unless opted into by users. For instance, Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries which, regarding the use of personal data, calls for “privacy by default” and “privacy by design” principles to be applied at all stages with a view to prevent or minimise the risk of interference with the rights and fundamental freedoms of users

(para. 2.4.3). For example, Sections 9-13 of the United Kingdom Online Safety Act imposes safety by design obligations on regulated platforms, including default settings for children, design assessments for features like algorithmic feeds, and obligations for user empowerment through settings and controls.

154. The design of online platforms, including user interfaces, content architecture, and system interactions, plays a central role in shaping how users access, engage with and contribute to information online. Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression calls for the digital infrastructure of communication to be designed to promote human rights. In line with the UNESCO Guidelines for the governance of digital platforms, "digital platforms should ensure non-discrimination and equal treatment in their design processes, including addressing biases, stereotypes, and discriminatory algorithms or content moderation practices that affect women and girls, as well as groups in situations of vulnerability and marginalization, including indigenous communities" (para. 93).

155. Design-based obligations should take into consideration the need to uphold freedom of expression and a pluralistic information environment in which diverse voices and perspectives are present and accessible online. Online services should be designed to minimise restrictions on lawful expression or to marginalise minority or dissenting opinions. Such risks may stem from the design of algorithmic systems used to recommend, rank, prioritise or otherwise curate content; user interface elements optimised for maximised user engagement; or default settings that limit visibility or discoverability of certain types of content or user accounts. To counter these risks, platforms should provide clear, accessible and granular controls that allow users to influence how content is curated and recommended to them, such as adjustable recommendation settings, transparency over why specific content is shown, or the possibility to opt out of algorithm-driven recommendations that shape the discoverability of content in their feeds or search results and content curation based on user profiling.^[36]

On paragraphs 60-63

156. Due to their scale, reach and structural role in shaping public discourse, platforms of significant influence should be subject to enhanced responsibilities regarding the impact of their design and operational decisions. They should therefore be required to review the impact of their services on freedom of expression and other human rights, such as the right to private life, the right to non-discrimination and the rights of the child. Furthermore, the assessments should consider any impacts on democratic and electoral processes, as well as the risks for dissemination of legally restricted content.

157. Risks may be related to the design of services and the way they function. Furthermore, risks might stem from platforms' content moderation processes. For example, certain design features such as user registration mechanisms might offer inadequate protection against the creation of fake accounts used for harassment or fraud. Likewise, platforms' functioning can represent a risk for the dissemination and amplification of content, both through algorithmic prioritisation of the content that has the most potential to engage audiences and the misuse of platforms' advertising policies which allow for monetisation of such content.

158. Effective incorporation of user safety by design requires concerned platforms to undertake robust risk assessments throughout the implementation, maintenance and updating of design decisions. These may include pre-launch risk assessments (e.g. abusability testing) to minimise risks of harm before implementing design decisions, as well as implementing ongoing monitoring mechanisms to ensure any risks arising can be effectively mitigated. To that end, States may introduce regular intervals at which online platforms are obliged to carry out risk assessments.

159. Systemic risks manifest differently across user groups. Effective risk management and mitigation therefore necessitate input from those affected, along with contextual evidence that reflects their specific circumstances. States should therefore ensure that, when undertaking risk assessments, platforms of

significant influence engage in open, transparent and meaningful consultations (CM/Rec(2022)13on the impacts of digital technologies on freedom of expression, para. 3.5.) with affected stakeholders, particularly those whose rights and experiences are directly impacted by platform design and governance practices. Particular attention should be paid to ensuring the participation of women, given the prevalence of gendered online harms and the specific risks they face in online environments. Consultations with stakeholders are crucial in assessing elements such as whether platforms' content moderation policies and systems provide effective protection against risks of harm, including the subtle and evolving ones; whether they accurately respond to context-specific instances, e.g. by reflecting the impact of language, cultural or regional differences; whether they have real-life harmful consequences. Meaningful consultation ensures that platforms incorporate the lived experiences and vulnerabilities of different user groups and can better identify and mitigate the unintended risks of harm arising from their design and content governance practices. For example, the Guidelines on the protection of minors adopted in 2025 by the European Commission pursuant to Article 28.4 of the Digital Services Act, provide that the design and functioning of recommender systems for minors, as well as any element of the platform that concerns their privacy, safety and security should be tested with minors and their feedback taken into account – including consulting with minors of different ages, from a diverse range of cultural and linguistic backgrounds, and those with disabilities.^[37]

160. The consultative process is not only key to ensuring effective protection from harm but also to prevent the implementation of mitigation measures that go beyond what is necessary and proportionate in a democratic society, potentially restricting lawful expression or pluralistic discourse. Concerned platforms should be required to document how stakeholder input was considered and to demonstrate that the outcomes of these consultations were given due weight in shaping their risk mitigation measures, in line with Recommendation CM/Rec(2022)13on the impacts of digital technologies on freedom of expression, paragraph 3.5, which stipulates that internet intermediaries should provide full information on the process, content and outcome of consultations, disclosing all the feedback they receive and explaining whether and how it is taken into account.

161. The publication of risk and human rights impact assessments by platforms is increasingly recognised as a necessary element of transparent and accountable platform governance. The EU Digital Services Act, for example, requires very large online platforms and search engines to publish systemic risk assessments they conduct, along with the measures taken to mitigate those risks (Article 42.5). Public documentation of risk and human rights impact assessments is an important tool for user empowerment, since it allows users to better understand the risks associated with platform services. It also allows civil society and researchers to scrutinise and evaluate platform practices, regulators to assess compliance with legal requirements, and the broader public to engage in informed debate on the societal role and governance of platforms of significant influence.

162. Paragraph 63 addresses the positive obligation of States to foster an enabling environment for meaningful participation and scrutiny. In addition to ensuring legal guarantees of transparency, consultations and access to information about risk assessments and mitigation measures, States could also create, facilitate and support multistakeholder advisory or oversight forums. This could include providing resources and capacity-building to ensure involvement of civil society, researchers and independent experts in evaluating compliance and identifying priority risk areas across platforms. Such bodies should be able to have a meaningful influence, e.g. by issuing recommendations, initiating investigations or requesting additional data from platforms.

163. Platforms may redact or withhold specific information on their risk and human rights impact assessments where full public disclosure of certain internal platform data or operational procedures may create security vulnerabilities or be exploited by malicious actors, leading to increased risks of harm for users, for instance when it can be reasonably expected that such information could be exploited by malicious actors to coordinate harassment or abuse. However, such redactions should be strictly limited to

what is necessary to prevent these adverse effects and should be subject to regulatory oversight to prevent misuse. Regulatory bodies should have access to the full, unredacted information in order to assess the legitimacy of redactions and to ensure that exceptions are not applied arbitrarily or excessively.

On paragraph 64

164. Effective supervision and enforcement of legislative frameworks on platform accountability and user empowerment require that regulatory authorities operate with a high degree of independence, integrity and institutional capacity. States should ensure, both through legislation and in practice, that such authorities are able to perform their functions free from political, commercial or platform influence, whether direct or indirect. For instance, authorities should utilise independent technical infrastructure and data analysis capabilities, avoiding overreliance on platforms for compliance monitoring and evidence gathering. This may involve developing in-house technical expertise or leveraging specialised external partners such as independent auditors and researchers (see paragraphs 86–87 of the Appendix). However, in order to effectively exercise their supervisory and enforcement powers, regulatory authorities should have the power to request from platforms any information necessary for the performance of these functions. Platform accountability legislation should therefore foresee a legal obligation for platforms to provide such information for the purposes of regulatory oversight, as well as proportionate sanctions in case of failure or refusal to act on such request by regulatory authorities.

165. It is of utmost importance that the independence of regulatory authorities includes their financial independence and autonomy, in order to ensure they have adequate financial, technical and human resources to carry out their duties effectively. Recommendation Rec(2000)23 on the independence and functions of regulatory authorities for the broadcasting sector offers valuable guidance in this regard, which should be appropriately adapted and applied to the context of digital platform regulation. In addition, Recommendation CM/Rec(2022)11 on principles for media and communication governance sets out key governance standards that are directly relevant, particularly Principle 3, which stresses the importance of independence and impartiality in regulatory and governance bodies; and Principle 4, which highlights the need for evidence-based and impact-oriented governance choices. These standards reinforce the obligation of States to establish and support regulatory authorities that are functionally autonomous, sufficiently resourced and empowered to act on the basis of expertise and evidence. Regulating digital platforms, in particular, requires a wide range of specialised skills, as well as expert understanding of the principles of platform regulation. This includes the ability to assess how risks manifest differently across platforms, services and user groups. To address complex and evolving challenges in the digital environment, multidisciplinary expertise is essential.

On paragraph 65

166. Promoting an enabling environment that is conducive to online safety in a democratic society is a collective endeavour, in which civil society and other non-State actors acting in the public interest play a fundamental role. While the Appendix focuses on the respective responsibilities of States and platforms, it also entrusts specific roles and tasks to such actors, as elaborated in Section V, subsections on transparency, procedural rights and collective action of users. Paragraph 65 recognises the need to ensure that these actors perform their functions in the public interest, remain independent from both State authorities and platforms, and operate under a regime of transparency and accountability, with appropriate safeguards against potential abuse of their position. To that end, they may be subject to certification processes and reporting requirements. In order to ensure that these entities genuinely serve the public interest and carry out their delegated duties accurately and objectively, States should create a sustainable environment for their independent functioning, including through reliable sources of funding to

compensate them fairly, which ensure their effectiveness and independence. Among the possible forms for such compensation that are not dependent on the State budget, States may consider the introduction of levies or fees imposed on platforms.

V. Measures for online user empowerment

General provisions

On paragraphs 66-67

167. Paragraphs 66 and 67 reflect a user agency-focused approach to online risks, in which users are granted more influence over their online experience. This ensures that their individual autonomy is respected and their ability to manage their own exposure to risks is enhanced. The approach recognises that many harms encountered in digital environments are context-dependent and can often be mitigated, wholly or in part, by empowering users to make informed choices about the content and interactions to which they are exposed. It also contributes to strengthening users' resilience and equips them to take a more active and participatory role in the future development of digital governance, making them active participants in the digital environment. To this end, legal frameworks should actively support – and, where appropriate, require – mechanisms that enable user-centred curation and moderation of online content and behaviour. Recognising that most online platforms already operate mixed models of content moderation with some level of agency granted to users, user empowerment tools should extend beyond traditional binary approaches to include graduated options. Such mechanisms may include the ability to adjust the visibility or prioritisation of content according to individual preferences, or to restrict exposure to certain types of content through customisable tools. For example, the United Kingdom Online Safety Act introduces 'user empowerment duties' that oblige certain services to offer adults tools to filter or restrict content that is legally restricted only for children, such as abuse or glorification of self-harm, without removing such content altogether. Users should also have the option to rely on trusted third parties to support or manage these settings on their behalf.

168. User empowerment cannot serve as a substitute for a systemic, human rights-based approach to content governance and risk mitigation. Rather, it should be designed and implemented as part of a broader framework in which platforms remain fully accountable for ensuring a safe online environment. The redistribution of agency should thus be viewed as a shared responsibility, requiring sustained commitment from all actors to avoid reinforcing power imbalances or shifting the burden of safety onto users alone (see Explanatory Memorandum on paragraphs 57-58 above).

On paragraphs 68-69

169. Paragraphs 68 and 69 reaffirm the principle of proportionate and graduated responsibility of platforms (see Explanatory Memorandum on paragraph 41), with a specific focus on user empowerment duties.

170. Obligations placed on platforms under applicable legal frameworks should represent their minimum requirements, leaving significant scope for platforms to voluntarily implement additional measures. The Appendix supports the principle that platforms should be encouraged to adopt empowerment measures even where they are not strictly obliged to do so by law, recognising that voluntary action can play a complementary role to regulatory and co-regulatory frameworks in fostering an enabling online environment.

171. The role of States in this context could be to engage in regular dialogue with online platforms in order to enable and facilitate the setting up of collaborative multistakeholder forums between different actors including platforms, user groups, civil society actors, the academic community, etc. In addition, States can play an important role in supporting various digital literacy activities involving online platforms, which can encourage proactive measures aimed at improving users' digital and media and information

literacy, including their understanding of how platforms present and recommend content (see CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age, para. 1.4.5, and UNESCO Guidelines for the governance of digital platforms, paras 80-84).

Empowerment by design

On paragraph 70

172. Paragraph 70 states a very general principle: the design choices of platforms should give users control over their online experience to the maximum extent possible. This implies that tools are designed and implemented in a way that facilitates their use. The maximisation of user empowerment, however, does not relieve platforms of their duties, as applicable, to assess and mitigate the risks of human rights impacts of their systems.

On paragraph 71

173. Recommender systems, i.e. algorithmic tools used to rank, filter and suggest individual pieces of content to users, are among the principal areas in which platform design choices can give rise to systemic risks.

174. By shaping how individuals access information and engage with content, recommender systems represent a core design component of online platforms. Ranking and recommendation mechanisms optimised for engagement often prioritise content with high predicted interaction potential and display it to users who are most likely to engage with it. This may contribute to the amplification and rapid dissemination of borderline content, including hate speech and disinformation. Recommender systems optimised primarily for engagement may adversely affect the discoverability of content by limiting users' exposure to a diversity of sources and perspectives, thereby undermining media pluralism and the effective exercise of the right to receive information.

175. Platforms should be obliged to make available a range of tools that give users meaningful agency to choose the types of content they wish to see and customise platform algorithms in line with their preferences, values and sensitivities. This should include enabling users to actively set their own choices, opt out of default settings or decline specific features or design elements imposed by the platform. For example, the United Kingdom Online Safety Act explicitly prescribes the inclusion of user tools as a mandatory design requirement for the service. It imposes a user empowerment duty on large user-to-user services (e.g., those designated as "Category 1 services") to allow adults to customise their experience by choosing whether to block, filter or receive warnings about content which encourages suicide or self-harm, encourages eating disorders, involves abuse or hatred based on protected characteristics (Section 16). Article 38 of the EU Digital Services Act subjects very large online platforms and search engines that use recommender systems to an obligation to introduce at least one alternative option which is not based on profiling. These may include chronological newsfeeds, recommendations of the most popular content or content that is trending at a particular moment. The risk management system or protection of minors under the Digital Services Act might impose further empowerment measures concerning recommender systems.^[38]

176. Platforms should provide users with tools that enable them to effectively protect their privacy, including the possibility to choose from a set of privacy setting options. As stated in Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems, default options should lead only to the collection of data that are necessary for and proportionate to the specific legitimate purpose of the data processing, while tracking settings should be set as default in opt-out mode (para. 2.2).

On paragraph 72

177. The Guidance Note on countering online mis- and disinformation offers valuable recommendations on platform design and user empowerment solutions that facilitate user agency and informed decision-making about the content with which they engage. Among other things, these may include the provision of supplementary information to users, age-related alerts, trigger warnings and additional content from official and independent authoritative sources such as professional news organisations and public service media.^[39]

178. Content labels are an important design feature that provides indications as to accuracy of information and enables users to assess its contextual reliability and integrity. A prime example is the use of labels or flags placed by independent fact-checkers. Fact-checking labels play an important role in empowering users to make informed decisions about the content they view, interact with and share. By providing valuable context, fact-checking reduces the perceived credibility of disinformation, making users less likely to share such content. The Guidance Note on countering online mis- and disinformation emphasises that "States and platforms should recognise the centrality of fact-checking as a key practice for the health of public debate", as well as that "platforms should collaborate with fact-checking organisations to ensure that their users are exposed and encouraged to share high-quality information on matters of public debate, and to challenge and debunk mis- and disinformation when they encounter it" (para. 9).

179. The important role of third-party labelling and content moderation arrangements has also been recognised in regulatory efforts to combat disinformation, most notably through the Code of conduct on disinformation,^[40] a co-regulatory instrument under the EU Digital Services Act. The commitments under the Code include use and integration of fact-checking in signatories' services, through measures such as collaboration with independent fact-checkers and use of mechanisms such as labelling, information panels or policy enforcement to help increase the impact of fact-checks on audiences.

On paragraph 73

180. Platforms of significant influence, through their recommender and content moderation systems, fundamentally steer public discourse, influencing both the visibility and the accessibility of information. Systems that are designed, operated and controlled exclusively by the platform providers themselves, may offer users limited meaningful choice over the mechanisms that shape their online experience, which could raise important concerns in relation to freedom of expression, pluralism and democratic participation. Paragraph 73 recommends that States explore, through multistakeholder and evidence-based processes, as per general principles, the possibility of introducing in their regulatory frameworks a duty for such platforms to allow the deployment of tools developed by third parties to perform these functions. Such tools, sometimes referred to as "middleware", have the potential to offer greater diversity of expression as well as a positive impact on circulation of information.^[41] Among other things, they could allow for different regional and linguistic markets to be more adequately served, as well as new services to emerge that are exclusively geared towards the public interest. For example, users could choose a recommender prioritising verified journalistic sources or one curated for local community relevance.

181. Users' agency in this respect could be enhanced by providing them with an option to entrust the decisions about their user experience to third-party services of their choice. This would allow them to select from a variety of third-party providers and algorithms that best align with their preferences and needs, thereby potentially broadening user options and enhancing individual choice over what content is seen. Allowing personalisation of both content curation and content moderation could help users align their online environment with their own preferences and values, whether that means stricter filtering of content, prioritisation of verified news sources, greater emphasis on local or community-relevant content, or moderation practices sensitive to cultural and linguistic contexts. However, it is also recognised that reliance on third-party service providers of content tools and services may raise concern about pluralism,

as they may enhance echo chambers, information bubbles and rabbit hole effect, entail delicate issues about the sharing of responsibility and liability between the third-party tool provider and the hosting platforms, and raise technical issues.

182. Therefore, while encouraging States to consider this approach, paragraph 73 also provides some preliminary regulatory boundaries. Integration of third-party tools and services should be enabled only under non-discriminatory and fair conditions, in accordance with principles established by Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, in particular paragraph 2.3. Additionally, providers of such tools should be subject to standards of transparency, accountability and compliance with human rights obligations, including the right to privacy and freedom of

expression. Safeguards must be put in place to protect user data, ensure interoperability without compromising security and prevent abusive or discriminatory third-party systems from gaining access. In this respect, transparency requirements should extend to the disclosure of the nature and extent of involvement of third parties in the content moderation process to ensure effective oversight.

On paragraph 74

183. The need for platform design to proactively promote user empowerment and safety for persons with impairments is grounded in a framework of Council of Europe standards (CM/Rec(2016)5 on internet freedom, para. 2.1.1) as well as broader international human rights law (UN Convention on the Rights of Persons with Disabilities, Articles 2 and 21).^[42] The right to freedom of expression enshrined in the Convention includes equal access to the internet for all, without discrimination. It is further highlighted in Recommendation CM/Rec(2019)6 on the development and promotion of digital citizenship education, which states that persons with disabilities as digital citizens are entitled to access to safe and inclusive online environments and to digital literacy education tailored to their needs. Persons with impairments often face accessibility barriers that prevent them from fully benefiting from safety and empowerment measures on digital platforms. These barriers may arise from inaccessible interfaces or from the lack of compatibility with assistive technologies. To address these risks, platform design should ensure that persons with impairments are able to use and deploy third-party accessibility tools – such as screen readers, captioning systems or text-to-speech applications – so that built-in accessibility features can be effectively supplemented with independent tools tailored to individual needs. This is particularly important for online safety and empowerment functions, such as complaint and flagging mechanisms, which are only effective if users can access and engage with them.

On paragraph 75

184. Children are especially vulnerable in the digital environment, as they may not yet have the capacity to fully recognise online risks or take appropriate steps to protect themselves. It is therefore essential to implement effective safeguards that uphold their rights, ensure their safety, and support their healthy development online. The requirement for effective age assurance systems is essential to protect children from exposure to products, services and content that are illegal (such as child sexual abuse material) or legally restricted for their age group: pornography and dating sites, as well as any other specific age-restricted content such as online gambling, online sale of tobacco and alcohol, commercial communications for products and services not intended for children, as well as other types of content that may be prejudicial to their physical, mental or moral development.

185. Recent legislative initiatives reflect a growing recognition of the need for robust age assurance mechanisms in the online environment. The United Kingdom Online Safety Act imposes duties on the regulated providers (of user-to-user services) to implement “highly effective age assurance” to prevent children from encountering certain types of legally restricted content, including pornography and content that encourages, promotes or provides instructions for either self-harm, eating disorders or suicide. The

related Guidance on highly effective age assurance, adopted by Ofcom, the independent regulator entrusted with regulatory functions under the Act, outlines the criteria for highly effective systems (technical accuracy, robustness, reliability and fairness) and requires that such systems be proportionate and privacy-preserving.^[43] Article 28.b of the EU Audiovisual Media Services Directive requires video-sharing platforms to establish and operate age verification systems to protect minors from audiovisual content that may impair their physical, mental or moral development. Article 28 of the EU Digital Services Act, on the other hand, mandates online platforms accessible to children to “put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors”. In the Guidelines on the protection of minors, adopted pursuant to this provision, the European Commission clarified that age assurance can be one such measure but stressed that platforms should conduct a prior assessment to determine whether the measure is both appropriate and proportionate and whether the same level of protection could be achieved by relying on other “less far-reaching measures” (para. 31). The Guidelines take a risk-based approach to age assurance. For instance, they consider that age verification is necessary to protect children from high-risks,

such as the sale of alcohol, tobacco, nicotine-related products or access to any type of pornographic content. Conversely, where the risk to children’s safety and security is medium, age estimation would be appropriate and proportionate. Under Article 35, very large online platforms and search engines are under the additional obligation to take targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate.

186. Paragraph 75 recommends a stricter approach to age assurances when platforms predominantly offer content or services that are legally restricted to minors, such as pornography or gambling. This expression is meant to cover platforms where a large part of user interaction concerns, for example, pornographic content, despite the hosting of other non-legally restricted content. Stricter requirements for age verification of users for this kind of platform, and specifically platforms in which pornographic content is shared, are already required by the legislation of several Council of Europe member States.^[44]

187. Age assurance tools should be designed not merely to restrict children’s access but to uphold children’s rights by striking an appropriate balance: protecting them from legally restricted content and services while avoiding the risk of excessive barriers that would exclude them from legitimate content and services. For example, simple self-declaration of age is widely recognised as insufficiently reliable, as it can be easily misused, and therefore cannot be considered an adequate form of age assurance. Some methods that may appear technically effective for age assurance can, in practice, interfere with children’s rights in ways that are not immediately apparent. In particular, age estimation methods that use biometric data such as voice or facial features (AI facial analysis or voice recognition) or profiling of users based on their online behaviour, give rise to considerable concern regarding data privacy and security. For this reason, requirements for the deployment of age assurance systems should be accompanied by clear and practical guidelines on their security, transparency and inclusiveness, with specific attention to those methods using artificial intelligence. Additional safeguards should include independent oversight and the imposition of transparency and accountability obligations on platforms. Platforms should be required to publish clear and accessible explanations of how their age assurance systems operate, specifying what data is collected, how decisions are made and how users can seek correction in cases of error. Furthermore, platforms should be obliged to provide regular transparency reports on the performance of these systems, including information on accuracy, error rates, and measures taken to mitigate risks to children’s rights.

188. Age assurance systems may also negatively affect the rights of adults, for example by requiring the disclosure of excessive personal information or restricting access to content and services. Moreover, excessive reliance on systems that require advanced digital skills or the provision of official identity

documents may inadvertently exacerbate the risks of exclusion of individuals that are already at risk of marginalisation in society or in online spaces. Therefore, the safeguards outlined above should take into account the need to prevent any disproportionate or unnecessary interference with the rights of children and adults alike and pay specific attention to their effects on those at risk of marginalisation and discrimination.

On paragraph 76

189. The deployment of parental tools should be understood as both a risk-mitigation and an empowerment measure. These tools enable parents and guardians not only to manage children's exposure to potentially legally restricted content and services but also to actively guide their digital engagement and consumption habits. However, such tools – as well as any measure aimed at mitigating risks of harm to children – must place the best interests of the child as the primary consideration, in line with Article 3 of the UN Convention on the Rights of the Child. Furthermore, the level of parental control should be proportionate to a child's age and maturity, ensuring that children are granted increasing autonomy as they develop. This approach is consistent with Principle 2.2 on the evolving capacities of the child in Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the digital environment, which recognises that policies and practices should respond appropriately to the differing needs of younger children and adolescents.

On paragraph 77

190. Paragraph 77 introduces a measure aimed at enabling content creators to signal to any platform hosting their content, as well as to users, that content may not be suitable for an audience under a certain age. This action may be required under domestic legal frameworks, professional or ethical standards, other self-regulatory regimes, or it may be taken on a voluntary basis by the content creator in order to assume responsibility for protecting children. This requirement should be understood as a measure which empowers content creators to contribute to a safer online environment for children. For example, a broadcaster that is required under applicable media legislation to use age labels for content not suitable for children of certain ages should be able to apply the same labels to identical content distributed via its social media channels. Likewise, content creators who are not professional media actors but wish to exercise professional responsibility should have access to functionalities that enable them to do so.

On paragraph 78

191. The deployment of age assurance systems, parental controls, content labelling and other tools should not be viewed as sufficient in themselves or as solutions to the risks children face online. The provision of parental tools by platforms must not be interpreted as a shift in responsibility from platforms to parents, nor should the provision of labelling tools relieve platforms of their own responsibilities regarding content that is legally restricted or in breach of contractual rules and policies to protect children. While these tools can support parents and guardians in guiding and protecting their children and content creators in acting responsibly and complying with their own duties and responsibilities, platforms remain ultimately responsible for ensuring the mitigation of risks. In particular, this responsibility includes preventing the hosting and dissemination of illegal content, such as child sexual abuse material.

On paragraph 79

192. The ability of users to move their online identity, data and content between services is a key aspect of their digital autonomy and, thereby, empowerment. Limitations on portability of online profiles lock users into single services in order for them to preserve their visibility and audience, which also reduces opportunities for pluralism in the digital environment. For content creators in particular, the ability to transfer their profiles and audiences across platforms helps to safeguard freedom of expression and

artistic creativity. It prevents undue dependence on a single platform's policies or algorithms and contributes to a more diverse and resilient digital public sphere. States should therefore require platforms to adopt design choices and technical standards that support profile portability.

193. Facilitating portability is also consistent with broader European standards, including the rights of individuals under data protection law to access and transfer their personal data, as well as the objectives of interoperability and openness in the governance of digital services. The EU Digital Markets Act,^[45] for example, grants data portability rights to all users with the aim of promoting fairness and contestability in digital markets by curbing the gatekeeping power of dominant platforms, especially through improving data access and portability for users.

Transparency

On paragraph 80

194. Transparency requirements are essential for understanding how platform policies and practices influence freedom of expression, as they equip users with the knowledge needed to interpret how their online experience is shaped and to make conscious choices about their engagement with content. In line with Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, platforms should provide adequate transparency in the design and implementation of their terms of service and their key policies, such as information regarding content removal, recommendation, amplification, promotion, downranking, monetisation and distribution, particularly with respect to their consequences for freedom of expression. The obligation to explain algorithmic systems for content organisation and curation

requires platforms to clearly inform users how content is ranked, prioritised and personalised, for example how their behaviour and interactions with content or other users affect their user experience through personalised recommendations and content ranking. Based on this understanding, users should be able to modify the parameters used for content organisation and curation and thereby make informed decisions on discoverability of content in their feeds or search results. The purpose of this transparency requirement is therefore twofold: first, to support informed user autonomy and the ability to influence what information is presented to them; and second, to reduce risks of harm by increasing user control over algorithmic content organisation and curation.

195. These explanations should be contained in the platform's terms of service. To be meaningful, they should be understandable to the target user audience, which may include making them available in local and minority languages and providing contextual examples.^[46] Furthermore, they should be as specific as possible in explaining the criteria that determine the order in which content appears, as well as the reasoning behind the weight or importance given to those criteria (CM/Rec(2018)20 on the role and responsibilities of internet intermediaries, para. 2.2.3). User autonomy and control can be further supported by explaining how the parameters for content organisation and curation can be modified.

On paragraphs 81 and 82

196. Transparency of content moderation practices and decisions is a core mechanism of platform accountability, particularly given their significant implications for freedom of expression, access to information, and the diversity of public debate. The publication of transparency reports is widely reflected in international standards and legal frameworks. Recommendation CM/Rec(2018)20 on the role and responsibilities of internet intermediaries in paragraph 2.2.4. calls for the "regular" publication of transparency reports providing simple, easily accessible and meaningful information on all content restrictions, including on the basis for such decisions, for example, court orders, user requests or enforcement of platforms' own policies. Platforms should therefore be required to provide transparency

reports at regular intervals. The EU Digital Services Act requires all platforms to publish annual transparency reports detailing content moderation actions, while very large online platforms and search engines are required to submit them every six months (Articles 15 and 42).

197. The reporting requirement should include both quantitative and qualitative information. Statistics should indicate the number of reports received, the actions taken and the outcomes of such actions, broken down according to qualitative categories. Operational principle No. 1 ("Numbers") of the Santa Clara Principles^[47] outlines a detailed set of transparency expectations in this respect, such as providing information on the source of flags (State actors, trusted flaggers, users, automation), the grounds for action (breach of law or platforms' own rules and policies), and whether content moderation decisions were taken with human oversight or by automated processes.

198. Automated content moderation systems lack the capacity to understand context and nuances required for accurate assessment, which could lead to both overremoval and underremoval of content. These risks are particularly acute when such systems operate without adequate human oversight or when safeguards to prevent discriminatory or biased outcomes are insufficient or lacking. In addition to providing statistics on content moderation decisions taken in automated processes, transparency reports should therefore include qualitative explanations related to the use of algorithmic moderation. These reports should disclose which algorithmic systems are relied upon for content moderation, the categories or types of content to which they are applied (e.g. hate speech, copyright enforcement, terrorist content), and provide

clear descriptions of the decision-making process, including the key criteria and logic employed.^[48] Reports should also include meaningful assessments of system performance, such as accuracy and success rates, error margins, and the safeguards in place to protect users' rights. Meaningful performance metrics may require separate reporting of false positive and false negative rates (rather than just overall accuracy), breakdown of performance by content type and language and contextual accuracy metrics, recognising that automated systems cannot distinguish context. The obligation to report on accuracy indicators, error rates and safeguards associated with automated moderation systems is reflected in the EU Digital Services Act (Article 15). In the United Kingdom, the Final transparency guidance on Online safety transparency reporting, developed by Ofcom, indicates that the regulator may require service providers to report on the use of automated systems in content moderation, including accuracy metrics or quality assurance mechanisms that are applied.^[49]

199. One essential safeguard is the ability of users to obtain a human review of any automated content moderation decision. Meaningful human review requires more than mere token involvement and should be conducted by individuals with genuine authority and adequate resources. Thus, reviewers must have appropriate authority and capability to change a decision made through automated means and have access to all relevant data such as the context of flagged content and the algorithmic reasoning which triggered the automated decision. Realistic time allocations, which allow for thoughtful review and adequate personnel deployment are also critical to effective oversight.

On paragraph 83

200. Advertising is a fundamental part of how online platforms finance their services. However, digital advertising often takes diverse forms, some of which are designed to blend seamlessly with regular content, making it difficult for users to distinguish between editorial information and sponsored material. This blurring of lines undermines users' ability to critically assess what they see online and increases the potential for manipulation. A lack of transparency in advertising facilitates hidden commercial practices that exploit consumer vulnerabilities. These risks are heightened in the case of targeted advertising, which relies on profiling users through the collection of personal data and tracking of online behaviour. Such

practices can have serious consequences for users' data protection rights while also generating wider societal harms. They can also amplify harms for democratic and electoral processes, especially in light of concerns about the spread of disinformation and political microtargeting.

201. Transparency about the source of advertising is a crucial tool for user empowerment. It enables users to clearly recognise when content is paid for, by whom and the amounts spent on advertising. Such information should be made easily and visibly available so that users can unambiguously identify content as an advertisement, distinguish it from other forms of content and understand the person or entity on whose behalf the advertisement is placed. A common practice is the use of labels directly attached to advertisements. For example, the EU Code of Conduct on disinformation envisages such measures for political advertising. Furthermore, it requires signatories to maintain a publicly accessible repository of political and issue-based advertisements. Article 39 of the Digital Services Act also mandates that very large online platforms and search engines maintain publicly accessible ad repositories containing data on displayed adverts, the advertiser and beneficiary, targeted audience and aggregate ad spending.

202. Advertising transparency also requires that users can understand why they are seeing a particular advertisement. This includes clear and accessible explanations of whether targeting techniques have been used and the parameters applied, such as demographic, geographic, contextual, interest-based or behavioural criteria (EU Code of Conduct on disinformation, measure 9.2). By providing this information in an understandable format, platforms support informed user choice and mitigate risks of covert influence and manipulation.

On paragraph 84

203. Platforms are increasingly offering opportunities for content creators to monetise directly, for instance through participation in revenue redistribution programmes. As the creator economy grows in scale and influence, there is a growing recognition of the need for transparency in how platforms monetise user-generated content, particularly with respect to who benefits, under what conditions and based on which allocation models. Opaque monetisation practices can create risks for both users and content creators. The lack of clarity about content that is monetised, by whom and under what conditions limits the ability of users to assess potential commercial incentives behind the content they consume. At the same time, content creators depend on platform monetisation systems to reach audiences and generate income. If the criteria that determine how resources are distributed are opaque or unpredictable, creators may face undue dependency or pressure to adapt their content to algorithmic or policy preferences. A further concern is undue demonetisation, often carried out by automated systems, which can significantly affect creators' livelihoods.

204. Transparency about monetisation is also closely tied to the integrity of democratic processes. One of the risks of platform monetisation practices is that they create financial incentives for the publication or amplification of legally restricted content and activities online. In particular, the monetisation of disinformation poses serious risks to public trust and democratic discourse, as opaque advertising systems often channel revenue to websites that deliberately spread disinformation. Transparent disclosure of monetisation practices, including the criteria used to allocate resources, is therefore essential to ensure that legally restricted content is not financially rewarded, while safeguarding fair opportunities for legitimate creators and supporting a healthier digital information ecosystem.

205. The European Commission's guidance on the interpretation and application of the Unfair Commercial Practices Directive clarifies that influencer marketing (including paid posts, affiliate content, retweets or tagging the trader or brand) must clearly disclose when promotion is paid and that such disclosure should be prominent. This obligation also applies when influencers endorse their own products or business.^[50] At the international level, the OECD's work on platform workers and the digital economy

highlights the need for fair and transparent remuneration systems, including in content-creation contexts and urges governments to ensure that creators are not subjected to opaque or arbitrary changes in monetisation policies.^[51]

On paragraph 85

206. Platforms of significant influence should provide tools that enable content creators to disclose how their content is monetised, i.e. how it generates revenue. This may include disclosing the identity of their business partners, ensuring visibility of their advertising arrangements, such as sponsorships, affiliate partnerships and participation in platform-specific monetisation programmes. For users, transparency about these financial relationships is crucial as it allows them to understand potential commercial biases, evaluate the credibility of content and make informed decisions about what they consume. These tools could be in the form of visible labels or metadata tags on posts or videos, or creator dashboards accessible to users, showing revenue sources or business relationships tied to specific content.

207. Article 28.b of the EU Audiovisual Media Services Directive, for example, requires video-sharing platforms to have a functionality for users who upload user-generated videos to declare whether such videos contain audiovisual commercial communications, in recognition that platforms cannot always be aware of such arrangements. Emerging self-regulatory arrangements across Europe provide good practice examples for disclosing the identity of sponsors. For instance, in Ireland, the Competition and Consumer Protection Commission and the Advertising and Standards Authority have issued joint Guidance on

influencer advertising and marketing,^[52] whose measures apply to content creators regardless of whether they are registered as audiovisual media service providers with the media regulator.

On paragraphs 86-87

208. Independent scrutiny is essential for understanding online risks, including those arising from platforms' design, operation and policies, as well as for assessing the effectiveness of any measures taken by platforms to mitigate those risks. Without legally guaranteed access to platform data, research relies on voluntary, inconsistent and selective cooperation by platforms or access through third-party tools. Globally, researcher access to data remains *ad hoc* and challenging. The obligation for platforms to grant researchers access to data is increasingly being advocated at the international level, including in the OECD Principles on artificial intelligence^[53] and the UNESCO Guidelines for the governance of digital platforms (Principle 5). Both instruments emphasise that such access must be secure, ethical and compliant with data protection laws.

209. Because meaningful scrutiny of platforms requires access to granular and often sensitive data to evaluate online risks, researchers should not be restricted from accessing and processing personal and confidential data when it is necessary for the purpose of research, provided that all due data protection safeguards are in place. Safeguards should address both the protection of users' privacy and personal data (for example, through anonymisation of datasets) and the protection of businesses' proprietary information and trade secrets.

210. As detailed in paragraph 6.6. of the Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, platforms should be required to grant access to individual-level data to researchers who have been vetted by an independent scientific institution and fulfil the criteria with respect to their qualifications, expertise and independence from commercial and political interest, and who have received approval by the ethical review board of their affiliated university. States may also establish systems in which platforms are obliged to provide access to researchers vetted by an independent authority on the basis of the same criteria, including those not affiliated with universities.

211. Access to public data should be subject to less stringent requirements and may also be granted to other stakeholders conducting research, such as civil society or media organisations, provided that they demonstrate adherence to data protection and security standards, confirm that their work is conducted independently and in the public interest. In addition, platforms should allow independent researchers access to publicly available data without technical restrictions. Such access should not be limited to interfaces administered by platforms, such as the application programming interfaces (APIs), but should also permit the collection of data through “scraping”, i.e. automated collection of data from the user-facing interfaces of websites or apps.

212. Paragraph 87 refers in particular to the obligation of platforms to allow independent researchers to conduct research on the platforms and to interact directly with platform systems for the purpose of examining the impact that platform services may have on human rights and other risks arising from their design, functioning, policies and rules, as well as for assessing the effectiveness of measures aimed at empowering users. For example, researchers might experimentally run certain types of legally restricted content to test its visibility or reach, or to evaluate how the platform’s moderation systems respond to such content.

Procedural rights

On paragraph 88

213. Contractual policies and rules of platforms – such as terms of service and community standards – define the boundaries of permissible user behaviour, content and interactions. Therefore, they should be made publicly available in clear and plain language that is understandable and easily accessible to the average user of the service. Furthermore, they should be available in the languages commonly spoken by

the platform’s users and affected communities to ensure accessibility and comprehension. They should also explicitly explain the potential outcomes of their enforcement. The grounds for the most severe enforcement decisions, such as account suspension or content removal, should be indicated with particular clarity, avoiding vague or ambiguous wording.

214. Special attention should be given to the needs of children, who represent a significant share of platform users and are especially vulnerable to opaque or overly complex contractual provisions. In line with the Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the digital environment, contractual rules relevant to children should be formulated in simple, age-appropriate language that enables them to understand both their rights and responsibilities when using online services. A service can be considered relevant for children if it is directed at, or predominantly used by, them.

215. Significant changes to terms of service or community standards can have wide-ranging implications for users’ rights and activities, including their ability to create, share or monetise content. Platforms should therefore be obliged to notify users in advance of any substantial modifications, accompanied by meaningful explanations that highlight not only the nature of the changes but also their practical consequences. Explanations of updates should be clearly signalled, for example through direct notifications in users’ newsfeeds or by providing “redlined” versions of updated contractual terms that highlight specific changes in a way that allows users to compare the previous and updated versions.

On paragraphs 89-92

216. In enforcing their policies and rules, platforms may take actions that may restrict users’ and content creators’ freedom of expression. These actions can include removal of content such as posts or videos, suspending or permanently deleting user accounts, demonetising content, downranking it in recommendation systems, or demoting and suppressing the visibility of content, users or groups of users. Each of these actions can significantly affect users’ ability to communicate, access information or

maintain their online presence and audience. It is therefore essential that safeguards are in place to ensure that content is not removed or suppressed unjustly. Shadow banning is particularly problematic because of its opacity: it restricts a user's visibility without their knowledge, depriving them of the ability to understand or challenge the platforms' decision. Council of Europe instruments (in particular CM/Rec(2018)2¹⁵⁴ on the role and responsibilities of internet intermediaries and CM/Rec(2022)13¹⁵⁴ on the impacts of digital technologies on freedom of expression, Section 4), as well as international standards and regulatory framework¹⁵⁴ call for transparency in content moderation, requiring platforms to notify users whenever restrictions are imposed on their freedom of expression, as well as to provide clear, accessible avenues for appeal. The absence of meaningful feedback about content moderation decisions can undermine users' trust in the moderation system, discourage their participation in platform governance, and have negative consequences for user empowerment.

217. Notifications sent to users whose content or account has been affected by a content moderation decision need to explain the grounds for the decision, i.e. it should indicate whether it was based on a violation of legal provisions or the platforms' own rules and contain a reference to the specific legal provision, internal policy or rule allegedly violated. Users should also be informed of what triggered the procedure, whether it was automated detection, flags or notifications from other users or trusted flaggers, or a request or order from State authorities. Explanation of the decision-making process refers to indicating whether the decision was made through automated means or human review. Furthermore, notifications should provide a clear explanation of the process available to appeal the decision. The EU Digital Services Act, under Article 17, requires platforms to provide users with a clear, detailed and timely statement of reasons whenever they remove, restrict or otherwise moderate content, accounts or services. It also obliges platforms to ensure that these statements include the legal or policy basis for the action, type of restriction applied (e.g., removal of content, account suspension, downranking, demonetisation or disabling of access), and information on available avenues for redress. Under Article 20, users and content creators may

challenge such decisions through a user-friendly, free-of-charge internal complaint-handling system, which platforms are required to process in a timely manner and respond to with clear and reasoned feedback.

218. When platforms of significant influence take action with regard to content, their duty to notify should extend beyond the original creator of that content. Notification should also reach any identifiable user directly affected or concerned, provided that such users have opted in to receive these notifications. For example, this could include a person mentioned in the content, such as an individual who is the subject of a defamatory post and who may have a legitimate interest in knowing whether the content remains available or has been removed. Other users may also include individuals or organisations who flagged the content, since they have initiated a moderation request and therefore have a clear interest in the outcome of the platform's decision.

219. In addition to the availability of internal appeal systems, users should also have the right to seek remedy through external appeal mechanisms. States should provide for the conditions necessary for such independent appeal or oversight mechanisms to function, ensuring their legitimacy and accountability, as well as safeguards of their independence, transparency and equitable access.¹⁵⁵ These mechanisms should also be able to provide effective remedies, including restoring content or accounts when decisions are found to be unjustified. Such mechanisms may take the form of alternative dispute resolution bodies such as the out-of-court dispute settlement bodies (EU Digital Services Act, Article 21), independent ombudspersons, or other independent third-party review mechanisms. The availability of external mechanisms should not absolve platforms of their obligation to maintain a robust internal appeal system, nor should it limit users' right to seek judicial redress (CM/Rec(2018)2 on the role and responsibilities of internet intermediaries, paras 1.5.2 and 2.5.5).

Collective action of users

On paragraphs 93-94

220. User reporting complements platforms' own monitoring efforts by enabling the identification of content in breach of platforms' own rules at scale, which is particularly important given that systematic proactive monitoring is rarely feasible or proportionate. These mechanisms are especially valuable for detecting context-dependent violations, such as harassment or hate speech, where human input is often necessary for accurate assessment. Crucially, accessible and effective flagging tools empower users to play an active role in shaping their online environment and foster greater understanding of platform policies and enforcement practices. To strengthen this participatory function, flaggers should receive meaningful feedback, for example the confirmation that their report has been received, information on whether any action was taken and, where no action was deemed necessary, a clear explanation of the rationale.

221. Platforms should be required to enable users to submit notices concerning content they believe to be legally restricted. This constitutes an important accountability mechanism, as user notification is often the trigger for platforms to become aware of the potentially legally restricted nature of the content, necessitating their action as highlighted in paragraph 55 of the Appendix. To facilitate the submission of sufficiently precise and substantiated notices, intermediaries should be required to design their notice-and-action mechanisms in a way allowing users to provide them with all the elements they need to act quickly and efficiently on the notices. By way of example, the EU Digital Services Act requires that notice-and-action mechanisms allow for the submission of "a sufficiently substantiated explanation of the reasons why the individual or entity alleges the information in question to be illegal content", as well as the exact electronic location of the content in question (Article 16). A similar standard applies in the United Kingdom under the e-commerce Regulation.^[56] Article 28.b of the EU Audiovisual Media Services Directive requires video-sharing platforms to provide users with content rating systems and to establish and operate transparent and user-friendly mechanisms for users to report content inciting hatred or violence, content harmful to minors or certain categories of illegal content. In order to facilitate user participation, notice mechanisms should be

prominently displayed, easily accessible and user-friendly, mirroring the design and usability of flagging tools used for reporting breaches of a platform's own rules.

On paragraph 95

222. Paragraph 95 encourages States to promote the identification and establishment of professionalised actors who can serve as trusted experts in notifying legally restricted content or flagging breaches of contractual policies and rules on platforms. By virtue of their expertise, such professionals, often referred to as "trusted flaggers" or "trusted notifiers", can provide higher-quality, better-substantiated notifications than ordinary users, improving the accuracy and timeliness of platforms' responses. For example, a civil society organisation specialising in the monitoring of hate speech can identify legally restricted content in this area more reliably than the general public. Similarly, flags by journalist associations or media councils could help platforms respond to systemic risks related to freedom of expression without unnecessarily restricting it (see also Guidance note on countering online mis- and disinformation, para. 37).

223. States should incentivise platforms to recognise and cooperate with such expert actors by granting them specific privileges, such as prioritised handling of their notifications and flags, accelerated review in appeal procedures and access to improved technical reporting interfaces such as bulk reporting tools or APIs. States should develop financial or institutional support schemes to ensure that these experts remain sustainable. At the same time, safeguards are necessary to ensure their independence and accountability. Recognition of professional notifiers should be based on transparent criteria, such as

demonstrated expertise, independence from political or commercial interests, and compliance with ethical and data protection standards. Oversight mechanisms should be in place to monitor performance and, where necessary, revoke their status in cases of misuse.

On paragraph 96

224. States should also foster an enabling environment for the establishment and professionalisation of independent user groups or associations that can act collectively on behalf of users and content creators. Such organisations could provide legal support, advocacy and assistance in navigating appeals processes, thereby helping to rebalance the power asymmetry between individual users and large platforms. To ensure their effectiveness, States may grant these groups specific privileges, including priority consideration of their collective appeals before platforms, financial or logistical support to guarantee their long-term sustainability, access to enhanced reporting tools or data to identify systemic problems and formal recognition of their right to initiate collective action in cases of rights violations. Particular attention should be paid to promoting diversity and inclusivity within these organisations, so that they adequately represent the interests of not only professional content creators but also vulnerable groups, children and marginalised communities who may face disproportionate risks online.

[1] The acronym LGBTI stands for lesbian, gay, bisexual, transgender and intersex, in the ordinary meaning given to these terms in relevant Council of Europe standards and documents, see, in particular, Recommendation CM/Rec(2010)5 on measures to combat discrimination on grounds of sexual orientation or gender identity and Recommendation CM/Rec(2025)7 on equal rights for intersex persons and European Commission against Racism and Intolerance (ECRI) (2023), *General Policy Recommendation N°17 on preventing and combating intolerance and discrimination against LGBTI persons*, available at <https://go.coe.int/ccZ8k>.

[2] United Nations (2011), *Guiding principles on business and human rights: Implementing the United Nations "Protect, Respect and Remedy" framework*, UN doc. HR/PUB/11/04, United Nations, New York and Geneva, hereinafter "UN Guiding Principles on business and human rights", available at <https://digitallibrary.un.org/record/720245?ln=en&v=pdf>, endorsed by the UN Human Rights Council, Resolution 17/4, 16 June 2011, UN doc. A/HRC/RES/17/4, available at https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/RES/17/4.

[3] See, European Union, *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act)*, available at <http://data.europa.eu/eli/reg/2022/2065/oj>, hereinafter "EU Digital Services Act", and United Kingdom, *Online Safety Act 2023*, c. 50, available at www.legislation.gov.uk/ukpga/2023/50, hereinafter "United Kingdom Online Safety Act".

[4] See, for an overview of emerging definitions in domestic law, European Audiovisual Observatory (2024), *National rules applicable to influencers*, Strasbourg, p. 22 ff., available at <https://go.coe.int/WaGoW>. The study covers EU member States, Norway and Switzerland.

[5] See, for example, European Parliament: Directorate-General for Parliamentary Research Services (2025), *Preventing radicalisation in the European Union – How EU policy has evolved – In-depth analysis*, section 4.2 and references therein, available at <https://data.europa.eu/doi/10.2861/3681328>.

- [6] Council of Europe, Committee on Counter-Terrorism (CDCT) (2025), *Report on the emerging patterns of misuse of technology by terrorist actors*, available at <https://go.coe.int/D69xT>.
- [7] Council of Europe (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*, adopted by the Steering Committee for Media and Information Society (CDMSI) at its 24th meeting, 29 November-1 December 2023, CM(2024)9-add1.
- [8] Council of Europe (2025), *Guidelines on the implications of generative artificial intelligence for freedom of expression*, adopted by the CDMSI at its 28th Meeting, 3-5 December 2025, CDMSI(2025)15-rev, available at <https://go.coe.int/CrPF>.
- [9] Zamfir I. and Murphy C. (2024), *Cyberviolence against women in the EU*, European Parliamentary Research Service, Policy Briefing PE 767.146, Brussels, available at [www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2024\)767146](http://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2024)767146).
- [10] Posetti J. et al. (2020), *Online violence against women journalists: A global snapshot of incidence and impacts*, UNESCO, Paris, available at <https://unesdoc.unesco.org/permalink/P-bd67e208-065d-41f5-ba4e-ca4aa6098871>.
- [11] Commissioner for Human Rights (2022), *Human rights comment: No space for violence against women and girls in the online world*, available at <https://go.coe.int/aJQth>.
- [12] European Union Agency for Fundamental Rights (FRA) (2024), *Fundamental rights report 2024*, available at <https://fra.europa.eu/en/publication/2024/fundamental-rights-report-2024>, p. 12. See also Recommendation CM/Rec(2026)2 on accountability for technology-facilitated violence against women and girls.
- [13] Clark M. and Grech A. (2017), *Journalists under pressure: Unwarranted interference, fear and self-censorship among journalists in Europe*, Council of Europe, Strasbourg, available at <https://go.coe.int/RoHyP>. See, also, Clark M. and Horsley W. (2020), *A mission to inform: Journalists at risk speak out*, Council of Europe, Strasbourg, available at <https://go.coe.int/s1kfx>.
- [14] Posetti J. et al. (2021), *The chilling: Global trends in online violence against women journalists*, UNESCO, Paris, available at <https://unesdoc.unesco.org/ark:/48223/pf0000377223>.
- [15] See: Congress of local and Regional Authorities, Resolution 459 (2020) and Recommendation 449 (2020) on Fighting sexist violence against women in politics at local and regional levels; Council of Europe Commissioner for Human Rights (2023), *Report: Human rights defenders in the Council of Europe area in times of crises*, available at <https://go.coe.int/yuHwm>; GREVIO (2025), *6th general report on GREVIO's activities, covering the period from January to December 2024*, Council of Europe, p. 33-38, available at <https://go.coe.int/hGP8f>.
- [16] See, Council of Europe, *Gender Equality Strategy 2024-2029*, CM(2024)17-final, adopted by the Committee of Ministers on 6 March 2024, CM/Del/Dec(2024)1491/4.3.
- [17] Commissioner for Human Rights (2022), *Human rights comment: No space for violence against women and girls in the online world*, available at <https://go.coe.int/aJQth>.
- [18] European Union Agency for Fundamental Rights (FRA) (2022), *Bias in algorithms - Artificial intelligence and discrimination*, available at <https://data.europa.eu/doi/10.2811/25847>, p. 12.

[19] Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS No. 201) (Lanzarote Convention); Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence (CETS No. 210) (Istanbul Convention); Recommendation CM/Rec(2019)10 on developing and promoting digital citizenship education; Recommendation CM/Rec(2019)1 on preventing and combating sexism. The Steering Committee for the Rights of the Child (CDENF) is preparing, a draft recommendation on the protection of children against violence through age-appropriate comprehensive sexuality education, to be submitted to the Committee of Ministers in 2026, see for more information <https://go.coe.int/MajAW>.

[20] Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership. See also Council of Europe (2025), *National media and information literacy (MIL) strategies – practical steps and indicators*, adopted by the Steering Committee on Media and Information Society (CDMSI) at its 28th Meeting, 3–5 December 2025, CDMSI(2025)09, available at <https://go.coe.int/bdThG>.

[21] As per the resources suggested on the Council of Europe's Cyberviolence portal: www.coe.int/en/web/cyberviolence/home. See also Recommendation CM/Rec(2018)7 on Guidelines to respect, protect and fulfil the rights of the child in the online environment.

[22] Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age; Recommendation CM/Rec(2024)2 on countering the use of strategic lawsuits against public participation (SLAPPs).

[23] Explanatory Report to the Convention on Cybercrime (ETS No. 185), para. 125.

[24] Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors.

[25] See, for example, All European Academy (ALLEA) (2023), *The European Code of Conduct for research integrity – Revised edition 2023*, Berlin, available at www.doi.org/10.26356/ECOC, para. 2.7.

[26] See, for example, European Union (2024), *Council conclusions on support for influencers as online content creators*, 23 July 2024, OJ C/2024/3807, available at <https://op.europa.eu/s/Aapy>.

[27] France, *Law No. 2023-451 of 9 June 2023 aimed at regulating commercial influence and combating the excesses of influencers on social networks*, available at www.legifrance.gouv.fr/eli/loi/2023/6/9/ECOX2308125L/jo/texte.

[28] Spain, *Law No. 13/2022 of 7 July 2022, General Law on Audiovisual Communication*, available at www.boe.es/eli/es/l/2022/07/07/13/con, and *Royal Decree 444/2024 of 30 April 2024*, available at www.boe.es/eli/es/rd/2024/04/30/444/con.

[29] European Union, *Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive)*, as amended by *Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media*

Services Directive). The consolidated text is available at <http://data.europa.eu/eli/dir/2010/13/2025-02-08>. For an overview, see European Audiovisual Observatory (2024), *National rules applicable to influencers*, cited above.

[30] For references and further examples, see European Audiovisual Observatory (2024), *National rules applicable to influencers*, cited above.

[31] France, *Law No. 2020-1266 of 19 October 2020 aimed at regulating the commercial exploitation of the image of children under the age of 16 on online platforms*, consolidated text available at www.legifrance.gouv.fr/loda/id/JORFTEXT000042439054/.

[32] See, for example, Court of Justice of the European Union (Grand Chamber), case C-507/17, *Google LLC v. CNIL*, 24 September 2019, on the lack of extraterritorial effects of the right to be forgotten.

[33] Council of Europe (2021), *Guidance note on content moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation*, adopted by the Steering Committee on Media and Information Society (CDMSI) at its 19th meeting, 19-21 May 2021, pp. 36-39, available at <https://go.coe.int/kiRAW>, hereinafter "Guidance Note on content moderation".

[34] UNESCO (2023), *Guidelines for the governance of digital platforms: Safeguarding freedom of expression and access to information through a multistakeholder approach*, available at <https://unesdoc.unesco.org/ark:/48223/pf0000387339.locale=en>, para. 90.

[35] See, also: EU Digital Services Act, Articles 4-7; United Kingdom, The Electronic Commerce (EC Directive) Regulations 2002, 2002 No. 2013, available at www.legislation.gov.uk/uksi/2002/2013/contents, Regulations 17-19.

[36] UNESCO (2025), *Towards user empowerment: a multistakeholder action plan for integrating media and information literacy on digital platforms*, available at <https://unesdoc.unesco.org/ark:/48223/pf0000394855>, Action 27.

[37] European Commission (2025), *Guidelines on measures to ensure a high level of privacy, safety and security for minors online, pursuant to Article 28(4) of Regulation (EU) 2022/2065*, Communication C/2025/5519, 7 October 2025, available at <http://data.europa.eu/eli/C/2025/5519/oj>, hereinafter "Guidelines on the protection of minors", paras 65 and 89.

[38] European Commission (2025), *Guidelines on the protection of minors*, cited above, Section 6.5.

[39] See also paragraph 1.5. of Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression.

[40] European Commission, Directorate-General for Communications Networks, Content and Technology (2025), *Code of conduct on disinformation – As amended in October 2024*, available at <https://data.europa.eu/doi/10.2759/5029213>.

[41] France, Conseil national du numérique (National Digital Council) (2024), *Fostering the wealth of networks*, available at www.conseil-ia-numerique.fr/files/archive/en/communiqu/fostering-wealth-networks.html.

- [42] United Nations, Convention on the Rights of Persons with Disabilities, UNTS 2515, p. 3, available at https://treaties.un.org/pages/viewdetails.aspx?src=treaty&mtdsg_no=iv-15&chapter=4&clang=_en.
- [43] United Kingdom, Ofcom (2025), *Part 3 Guidance on highly effective age assurance*, available at www.ofcom.org.uk/online-safety/online-safety-regulatory-documents, pp. 10-12.
- [44] United Kingdom, Online Safety Act, Part 5; European Commission (2025), *Guidelines on the protection of minors*, cited above;
- [45] European Union, *Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)*, available at <http://data.europa.eu/eli/reg/2022/1925/oj>.
- [46] See, for example, UNESCO (2025), *Towards user empowerment: a multistakeholder action plan for integrating media and information literacy on digital platforms*, cited above.
- [47] *Santa Clara Principles 2.0 on transparency and accountability in content moderation*, 2021, available at <https://santaclaraprinciples.org>. It should be noted that the Santa Clara Principles "have been developed to support companies to comply with their responsibilities to respect human rights and enhance their accountability, and to assist human rights advocates in their work. They are not designed to provide a template for regulation".
- [48] This information should be made available to users already in terms of service, reflecting the principle provided for in paragraph 3.4. of the Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression, see also Explanatory Memorandum on para. 80.
- [49] United Kingdom, Ofcom (2025), *Online Safety Transparency Reporting. Final Transparency Guidance*, available at <https://www.ofcom.org.uk/online-safety/online-safety-regulatory-documents>, in particular Annex A.
- [50] European Commission (2021), *Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market*, Notice 2021/C 526/01, available at [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021XC1229\(05\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021XC1229(05)), Section 4.2.6.
- [51] Lane M. (2020), "Regulating platform work in the digital age", *OECD Going Digital Toolkit Notes*, No. 1, OECD Publishing, Paris, available at <https://doi.org/10.1787/181f8a7f-en>.
- [52] Ireland, Competition and Consumer Protection Commission (CCPC) and Advertising and Standards Authority (ASAI) (2023), *Guidance on influencer advertising and marketing*, available at www.ccpc.ie/business/help-for-business/guidelines-for-business/influencer-advertising-and-marketing/.
- [53] OECD (2019), *Recommendation of the Council on artificial intelligence*, available at <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>, Principles 116-118 (Data access for research purposes).

[54] Santa Clara Principles; UNESCO Guidelines for the governance of digital platforms, Principle 5; EU Digital Services Act.

[55] For example, the UN Guiding Principles on business and human rights clearly highlight the role of the States in ensuring that business-related human rights restrictions are remedied. They also highlight principles for effective complaints mechanisms to that end.

[56] United Kingdom, *The Electronic Commerce (EC Directive) Regulations 2002, 2002 No. 2013*, available at www.legislation.gov.uk/ukxi/2002/2013/contents, Regulation 19.