Parliamentary Assembly
Assemblée parlementaire
https://pace.coe.int

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

Resolution 2590 (2025)[1]

Regulating content moderation on social media to safeguard freedom of expression

Parliamentary Assembly

1. Social media have become an online agora where users come to exercise their right to freedom of expression and information in many ways. These platforms enable users to post their own content and enjoy the content posted by others, get informed and inform others, and communicate with other users.

2. The right to freedom of expression is not an absolute right; social media are legally obliged to remove any illegal content when they become or are made aware of its existence on their services. Moreover, it is incumbent upon social media to combat the dissemination of harmful content.

3. Social media companies are also bearers of fundamental rights, such as the right to property and freedom of enterprise, and therefore they have a say in how users can use their services and what content they can post. The content moderation rules included in their terms and conditions allow for social media companies to demote, demonetise, restrict access to or remove a concrete content item because of its incompatibility with their terms and conditions. In extreme cases, social media companies can suspend or even terminate a user's account. Their terms and conditions have a contractual character, and users are bound by them on a take-it-or-leave-it basis.

4. The major social media companies, mainly owned by entities based in the United States, have a global reach; their content moderation policies and commercial or ideological decisions about content to promote or demote may have an immense influence on public opinion and on choices of billions of people. It is, nevertheless, incumbent upon them to respect the laws of the country in which they provide their services.

5. Given the potential impact on societal behaviours and on the proper functioning of democratic processes that the information and communication flow on social media has, it is incumbent upon the State to establish the fundamental principles and institutional framework that may correct the power imbalance resulting from the unequal contractual relationship and ensure the effective protection of the right to freedom of expression.

6. It is imperative, however, that public regulation of content moderation does not have a chilling effect on free speech and is not intended to impose the views of the political power in place and censorship on opinions or ideas which may conflict with the ruling majority's vested interests. Moreover, national regulations should not place undue burdens on social media, which could result in an overzealous approach to content removal. These regulations and their implementation must uphold freedom of expression and carefully assess the necessity of any restrictions.

7. The risk of restrictive content moderation policies is increased by the lack of transparency in their implementation. Social media have been accused of a practice called "shadow banning" whereby they delist or demote content dealing with controversial issues without notifying the user in question, making that content invisible to other users. This devious, hidden practice should be forbidden: it deprives users of the possibility to defend effectively their right to freedom of expression.

---

1. *Assembly debate* on 30 January 2025 (8th sitting) (see Doc. 16089, report of the Committee on Culture, Science, Education and Media, rapporteur: Ms Valentina Grippo). *Text adopted by the Assembly* on 30 January 2025 (8th sitting).

8. The press and the media in general use social media as a platform for disseminating information to the public. It is therefore essential that content moderation practices do not unduly affect media and journalistic content that respects professional standards and the national regulatory framework.

9. Content moderation is increasingly carried out by automated means. Artificial intelligence tools are much more efficient than human moderators in processing at a high speed the colossal amount of content circulating on the web and identifying prohibited content. They lack to date, however, the capacity to fully understand the subtleties of human interaction (humour, parody, satire, etc.) and to assess the content in its context.

10. For this reason, human moderators must remain the cornerstone of any content moderation system and be responsible for making decisions in cases where automated systems are not up to the task. However, human moderation can be biased and lead to inconsistencies among countries due to cultural differences. It is therefore imperative to establish clear and comprehensive standards and to guarantee appropriate training, to ensure that all moderators have the requisite knowledge of both the applicable legislation and the company's internal guidelines, as well as of the language and the context of the country from which the content originates. However, in the event of a military conflict between two countries, moderators from one country party to the conflict should not moderate content originating from the other.

11. Regrettably, despite their fundamental role, the working conditions of human moderators are inadequate. These moderators are overexposed to disturbing content that can cause them serious mental health problems and they suffer from restrictions on their freedom to speak out about the problems they encounter at work.

12. Generative artificial intelligence tools allow synthetic content that is virtually indistinguishable from human-generated content to be produced. Such content can be highly misleading, be a tool of disinformation and manipulation, and instigate hatred and discrimination, among other dangers. It is essential that users are made aware of content that appears to be genuine, but which is in fact not. In this regard, watermarking techniques are particularly beneficial but have several drawbacks, including their lack of interoperability among social media services.

13. An independent assessment of the terms and conditions, content moderation policies and their enforcement, with a view to identifying and promoting best practices, could help to ensure their consistency with principles which uphold a human rights-based approach to content moderation.

14. The establishment of clear and transparent rules for conflict resolution is essential to ensure the protection of users and to minimise the risk of being subjected to a potentially biased decision by the social media company, or of being forced to pursue costly legal action against a multinational corporation with enormous financial resources at its disposal.

15. The establishment of independent out-of-court dispute settlement bodies to assess content moderation decisions may prove beneficial in enhancing compliance with fundamental rights. Collaboration between social media companies in establishing such bodies could also facilitate dispute resolution.

16. As stated by the Parliamentary Assembly in its Resolution 2281 (2019) "Social media: social threads or threats to human rights?", social media companies should employ algorithms that promote the diversity of sources, topics and views, guarantee the quality of information available and thereby reduce the risk of "filter bubbles" and "echo chambers".

17. In the light of these considerations, the Assembly calls on the Council of Europe member States to review their legislation to better safeguard the right to freedom of expression on social media. In this respect, States should in particular:

17.1. require that social media uphold users' fundamental rights, including freedom of expression, in their content moderation policy and implementation practices;

17.2. require that social media platforms provide justification for any measure taken to moderate content provided by the press or media service providers prior to its implementation and allow them an opportunity to reply within an appropriate time frame;

17.3. in co-operation with the press or media organisations, implement a system of verification of media and journalist accounts, together with robust mechanisms to protect them from online harassment, hacks and fraud, and develop social media guidelines for press or media organisations on the publication of information on sensitive issues, with a view to avoiding unnecessary moderation restrictions on this type of content;

17.4. provide for minimum standards for the working conditions of human moderators, including a requirement of adequate training to carry out their often stressful tasks and of access to proper psychological support and mental healthcare when needed;

17.5. sign and ratify the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS No. 225) and adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place to meet the challenges of the identification of content generated by artificial intelligence systems;

17.6. require that content generated by artificial intelligence is disclosed as such by those initially posting it and that social media implement technical solutions allowing for such content to be easily identified by users, and encourage collaboration between social media companies to ensure the interoperability of watermarking techniques for content generated by artificial intelligence;

17.7. require that out-of-court dispute settlement bodies, when established, are independent and impartial, have the necessary expertise, are easily accessible and operate according to clear and fair rules, with certification of these requirements by the competent national regulatory authority;

17.8. promote, within the Internet Governance Forum and the European Dialogue on Internet Governance, reflection on the possibility for the internet community to develop, through a collaborative and, where appropriate, multistakeholder process, an external evaluation and auditing system aimed at determining whether algorithms are unbiased and respect the right to freedom of expression, and a "seal of good practices" which could be awarded to social media whose algorithms are designed to reduce the risk of "filter bubbles" and "echo chambers" and to foster an ideologically cross-cutting, yet safe, user experience.

18. The Assembly calls on social media companies to avoid measures that unnecessarily restrict the freedom of expression of users. These companies should, in particular:

18.1. directly incorporate principles of fundamental rights law, and in particular freedom of expression, into their terms and conditions;

18.2. use caution when moderating content that is not obviously illegal;

18.3. provide users with terms and conditions that are readily accessible, clear and informative on the types of content that are permissible on their services and the consequences for non-compliance, and which are understandable to the wide span of users notwithstanding differing levels of digital literacy and reading proficiency;

18.4. notify users without undue delay of any moderation action taken on their content, providing a comprehensive account of the rationale behind the decision, accompanied by a reference to the internal rules which have been applied;

18.5. refrain from shadow banning users' content and notify users of every instance of demotion or delisting;

18.6. ensure that automated content moderation processes are subject to human oversight and to rigorous and continuous evaluation to assess their performance;

18.7. make available a system for handling complaints that is easily accessible, user friendly and allows users to make a precise complaint;

18.8. give human moderators appropriate training and working conditions which pay attention to the heavy psychological stress they are submitted to, and ensure adequate protection of their health;

18.9. refrain from permanent deletion of content (including its metadata) that has been removed in accordance with legal obligations or with terms and conditions, in particular when the content in question may serve as evidence of war or other crimes;

18.10. ensure that the artificial intelligence systems they develop or use uphold Council of Europe standards, including the new Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law; algorithms should be designed to respect the right to freedom of expression, and to encourage plurality and diversity of views and opinions while ensuring a safe user experience, their operation modalities should be disclosed and users duly informed on how these algorithms filter and promote content;

18.11. collaborate with other online services with the aim of ensuring the interoperability of watermarking techniques for content generated by artificial intelligence;

18.12. collaborate with journalists and fact-checking organisations to effectively combat disinformation with information that adheres to the ethical and professional standards of journalism;

18.13. promote and support the creation of independent out-of-court dispute settlement bodies, and abide by their decisions and recommendations;

18.14. support the work of independent third-party oversight bodies and abide by their decisions and recommendations;

18.15. ensure that decisions related to content moderation are duly motivated and that researchers have access to full information on the legal base and reasoning behind each decision.