

Hate speech

Comparing the US and EU approaches

SUMMARY

Differences between the United States (US) and the European Union (EU) over the regulation of online platforms have taken on a new dimension under the Trump administration. Senior members of the US administration have strongly criticised the EU for 'limiting free speech' and have called the EU's content moderation law 'incompatible with America's free speech tradition'. Much of the debate is informed by misconceptions and misunderstandings.

The differences between the US and EU hate speech regimes are striking, largely for historical reasons. The First Amendment to the US Constitution provides almost absolute protection to freedom of expression. By contrast, European and EU law curtails the right to freedom of expression. Article 10 of the European Convention of Human Rights, which applies to all EU Member States, states that freedom of expressions 'carries with it duties and responsibilities'. In a democratic society, restrictions may be imposed in the interest, among others, 'of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others'. EU legislation criminalises hate speech that publicly incites to violence or hatred and targets a set of protected characteristics: race, colour, religion, descent or national or ethnic origin. Even though legislation in EU Member States varies widely, many have extended protection from hate speech to additional characteristics.

In light of the exponential growth of the internet and the use of social media, the debate about hate speech has essentially become about regulating social media companies. The focus has been on the question of whether and to what extent service providers are responsible for removing hate speech published on social media platforms. The US has opted not to impose any obligation on social media companies to remove content created by third parties, merely granting them the right to restrict access to certain material deemed to be 'obscene' or 'otherwise objectionable'. By contrast, the EU has adopted regulation that obliges companies to remove offensive content created by third parties, including hate speech, once it is brought to their attention. Social media companies also self-regulate, by adopting community guidelines that allow users to flag hate speech and ask for its removal.



IN THIS BRIEFING

- Hate speech – attempts at definition
- US approach to hate speech
- European approach to hate speech
- EU approach to hate speech
- EU Member States' laws on hate speech
- Major US social media companies' hate speech policies



Hate speech – attempts at definition

There is no universally agreed definition of hate speech. Article 20, paragraph 2 of the **International Covenant on Civil and Political Rights** (ICCPR) of 1966 – a key international human rights treaty – establishes that 'any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law'. The 1965 **International Convention on the Elimination of All Forms of Racial Discrimination** (ICERD) prohibits discriminatory speech and action. It obliges states to criminalise certain forms of hate speech and the commission of, or incitement to, acts of violence against any race, group of persons of another colour or ethnic group. It also obliges states to prohibit organisations and propaganda activities that promote and incite racial discrimination. The 2019 **United Nations Strategy and Plan of Action on Hate Speech** defines hate speech as 'any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor'.

US approach to hate speech

Absolute freedom of expression?

The American Revolution pitched 13 British colonies in North America against their colonial masters. It ended with the defeat of the latter in 1783, when the colonies gained independence and established the United States of America. In reaction to the oppressive rule of the British colonial administration, the new country laid down the right to freedom of expression in its founding documents. The [US Constitution](#) was adopted in 1789. It was supplemented in 1791 by the [Bill of Rights](#), which spells out Americans' rights in relation to their government, including the First Amendment to the Constitution, which provided that 'Congress shall make no law ... abridging the freedom of speech'. The First Amendment also protects the free exercise of religion, freedom of the press, and the right to assemble. It applies only to government action, not to private actors, and remains in force to this day. Individual states are subject to the same restrictions on limiting the freedom of speech, by virtue of the Due Process Clause of the Fourteenth Amendment

Extensive [case law](#) has developed to interpret the [meaning](#) of the First Amendment. The US Supreme Court ultimately adopted a 'marketplace of ideas' theory approach to most freedom of expression issues. The theory was first formulated in the 1919 Supreme Court Case [Abrams v United States](#), 250 U.S. 616 (1919):

'But when men have realized that time has upset many fighting faiths, they may come to believe ... that the ultimate good desired is better reached by free trade in ideas – that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out.'

As recently as 2017, in [Matal v Tam](#), 582 U.S. 218 (2017), the Supreme Court reaffirmed that even offensive speech is protected. The Court noted that it has

'said time and time again that "the public expression of ideas may not be prohibited merely because the ideas are themselves offensive to some of their hearers".'

The contemporary approach to hate speech in the US can be defined in a few words: freedom of speech is a fundamental constitutional right, which government may only restrict under certain prescribed circumstances. The Constitution grants the government some scope to regulate the 'time, place, and manner' of speech, but not the opinion expressed, no matter how hateful or despicable. An exception is made for 'true threats' or 'incitement to imminent lawless action'. In the seminal case [Brandenburg v Ohio](#) 395 U.S. 444 (1969), the Court found that:

'a state may not forbid speech advocating the use of force or unlawful conduct unless this advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action.'

Legislation concerning freedom of expression

The First Amendment to the US Constitution governs the rights of citizens to freely express and voice their ideas and opinions; no other statute or regulation legislates the content of expression. Congress and state legislatures may adopt statutes that deal with specific issues that may also affect freedom of expression/speech. At present, federal statutes [cover](#) fraudulently representing receipt of military awards, solicitations to commit crimes of violence, bribery of foreign officials, the political speech of federal employees, the pledge of allegiance to the flag, the means by which communications are regulated, equal access in public schools, and delivering defence information to foreign governments.

Protected speech

The Supreme Court has consistently [treated](#) political and ideological speech as worthy of particular protection under the First Amendment, including speech related to topics such as politics, nationalism, religion, and other opinions. Accordingly, the Court has subjected laws that purport to regulate political or ideological speech to heightened scrutiny, and has held that the First Amendment is mainly intended to protect discussion of public matters. According to the Court,

'speech deals with matters of public concern when it "can be fairly considered as relating to any matter of political, social, or other concern to the community," ... or when it is a subject of legitimate news interest; that is, a subject of general interest and of value and concern to the public ... [regardless of] the arguably "inappropriate or controversial" character of a statement ([Snyder v Phelps](#), 562 U.S. 443 (2011)).'

Limits to protected speech

The Supreme Court has defined certain categories of speech as not deserving First Amendment protection and therefore prohibited. These include obscenity, [fighting words](#) (i.e. words meant to incite violence), [defamation](#), child sexual abuse material, pro-drug speech at schools, incitement to imminent lawless action, [true threats](#), solicitation to commit crimes, treason, speech integral to criminal conduct, fraud and perjury.

'Fighting words' are 'words which by their very utterance inflict injury or tend to incite an immediate breach of the peace'. In deciding whether certain words are fighting words, the Supreme Court ruled that '[t]he test is what men of common intelligence would understand would be words likely to cause an average addressee to fight.' ([Chaplinsky v New Hampshire](#), 315 U.S. 568 (1942)).

Regulating social media platforms

In 1996, the US Congress passed the Communications Decency Act (CDA). The aim was to encourage online services to moderate the content on their platforms to make it safe for users to use, without limiting their freedom of expression. [Section 230](#) of the Communications Act of 1934 (47 U.S.C. §230), enacted as part of the CDA, provides protection for private blocking and screening of offensive material. [Section 230](#) begins by 'lauding the extraordinary advance in the availability of educational and informational resources to US citizens the internet represents'. The section praises the internet as a forum for a true diversity of political discourse, and stresses that it has flourished 'with a minimum of government regulation'. Crucially, Section 230 stipulates that internet platforms are not responsible for the content they host, provided it is created by a third party, with some [exceptions](#). Given this lack of responsibility for third-party content, Section 230 has been criticised as encouraging the spread of hate speech on the internet. At the same time, Section 230 protects providers from liability when they take action 'in good faith' to restrict access to content that is

'obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable'. This provision has enabled internet platform providers with headquarters in the US to adopt community guidelines (see below), opting for self-regulation in an area in which government regulation is essentially absent.

European approach to hate speech

Memories of the Holocaust – where hate speech targeting minorities, especially Jews, led to the demonisation, dehumanisation, and ultimately the slaughter of millions of innocent civilians – have shaped European views and legislation on hate speech. The [Nazi regime](#) replaced Germany's independent media with state-controlled radio and print media that disseminated hate speech, antisemitic, racist stereotypes, disinformation and lies. These media campaigns led to the normalisation of atrocity crimes, which facilitated the Holocaust, the state-sponsored, planned and systematic persecution and annihilation of some six million Jews, and at least half a million Roma and Sinti, by Nazi Germany and other racist states.

European Convention on Human Rights

The European Convention on Human Rights (ECHR), adopted in 1950 and ratified by 46 countries in Europe, guarantees [freedom of expression](#), but acknowledges that the exercising of this freedom may be 'subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interest of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary' (Article 10). Moreover, the Convention imposes limits on freedom of expression by prohibiting 'discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status' (Article 14).

At the same time, Article 17 of the ECHR states that 'nothing in this Convention may be interpreted as implying for any State, group or person any right to engage in any activity or perform any act aimed at the destruction of any of the rights and freedoms set forth herein or at their limitation to a greater extent than is provided for in the Convention'.

European Court of Human Rights

Interpreting the ECHR, the European Court of Human Rights (ECtHR's) approach to hate speech is [laid down](#) in two seminal judgments.

In *Handyside v the United Kingdom* (1976), the Court found that

'freedom of expression constitutes one of the essential foundations of [a democratic] society, one of the basic conditions for its progress and for the development of every man. Subject to paragraph 2 of Article 10 [of the European Convention on Human Rights], it is applicable not only to "information" or "ideas" that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no "democratic society". This means, amongst other things, that every "formality", "condition", "restriction" or "penalty" imposed in this sphere must be proportionate to the legitimate aim pursued.'

In *Erbakan v Turkey* (2006), the Court ruled that:

'tolerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance ...'

provided that any "formalities", "conditions", "restrictions" or "penalties" imposed are proportionate to the legitimate aim pursued.'

In consistent case law, the Court has declared inadmissible, on grounds of incompatibility with the values of the Convention, applications which are inspired by totalitarian doctrine or which express ideas that represent a threat to the democratic order and are liable to lead to the restoration of a totalitarian regime.

Council of Europe recommendation

The **Council of Europe**, whose members adopted the ECHR in 1950, [defined](#) hate speech in [2022](#) as 'all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as "race", colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation'.

EU approach to hate speech

Charter of Fundamental Rights

The [EU Charter of Fundamental Rights](#), declared in 2000 and in force since 2009, is a legally binding instrument that was drawn up to expressly recognise, and give visibility to, the role of fundamental rights in the legal order of the Union. The Charter provides for freedom of expression and information, but also prohibits 'any discrimination based on any ground such as sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation' (Article 21).

Legislation on hate speech

In 2008, following seven years of negotiations involving the European Commission, the European Parliament and the Council, the Council adopted a decision criminalising hate speech. [Council Framework Decision 2008/913/JHA](#) of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law (the 2008 Council Framework Decision), prohibits 'publicly inciting to violence or hatred against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin'.

The framework decision also includes two provisions on the prohibition of publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity and war crimes when the conduct is carried out in a manner likely to incite to violence or hatred against such a group or a member of such a group.

Extending the list of EU crimes to hate speech

At [present](#), the EU has no competence to criminalise hate speech based on grounds not covered by the 2008 Council Framework Decision. Article 83(1) of the Treaty on the Functioning of the European Union (TFEU) could serve as a legal basis for criminalising hate speech on other grounds, provided the offence becomes part of an exhaustive list of 'areas of particularly serious crime with a cross-border dimension resulting from the nature or impact of such offences or from a special need to combat them on a common basis' (EU crimes). Article 83(1) TFEU allows the Parliament and the Council to establish minimum rules regarding the definition of EU crimes and related sanctions.

In 2021, the Commission [invited](#) the Council to adopt a decision identifying hate speech (and hate crime) as an area of crime under Article 83(1). In 2022, the Council examined the proposal, with a broad majority in favour of the initiative. However, even though the Parliament also expressed its support for the initiative in 2024, the file has stalled in the Council, which has not reached the unanimity required. The Commission, in collaboration with the Polish Presidency, put the issue back on the agenda, with a discussion during the [High Level Forum on the future of EU Criminal Justice](#) on 4 March 2025.

The framework decision requires Member States to ensure that public incitement to violence or violence on the above grounds is punishable under their national legislation. The framework decision also applies to public incitement to violence or hatred manifesting itself online. The scope of the 2016 Code of Conduct (see below) makes explicit reference to content deemed illegal as per national laws transposing the framework decision.

Whereas all EU Member States have criminalised hate speech based on the grounds covered by the 2008 framework decision, national laws differ with regard to other protected characteristics, such as gender, disability, sexual orientation and age (see below).

EU Member States' laws on hate speech

At EU level, the 2008 Council Framework Decision criminalises hate speech based on a range of grounds (see above). Member States transpose the Framework Decision into national law, and all Member States have criminalised hate speech on the grounds covered by the Framework Decision: race, colour, religion, descent or national or ethnic origin.

However, EU Member States nevertheless treat [hate speech](#) in very different ways. Some criminalise hate speech 'generally', without reference to any specific grounds. These Member States have criminalised hate speech without reference to any specific protected characteristic, essentially barring any form of intolerance. In other Member States, it is illegal to denigrate a person or persons on the specific grounds of race, colour, religion, descent, or national or ethnic origin, as provided under the 2008 Council Framework Decision. Others extend the list of grounds provided by the 2008 Framework Decision to include sex (and/or gender, including, where relevant, gender identity and sex characteristics), sexual orientation, age, disability, language, political status and political convictions.

Regulating social media platforms

Digital Services Act

Social media platforms have been subject to some form of regulation in the EU since 2000. The need to tackle the increasing spread of illegal and harmful content, including hate speech, led to the adoption of the [Digital Services Act](#) (DSA), which entered into force in November 2022. The DSA put a framework of layered responsibilities in place, targeted at different types of online intermediary services, depending on their 'role, size and impact in the [online ecosystem](#)'. These include network infrastructure services (e.g. cloud and webhosting), online platform services (e.g. app stores and social media platforms), and services provided by [very large online platforms](#) (VLOPs) and very large online search engines (VLOSEs) that pose particular risks in the dissemination of illegal content and societal harms. By February 2025, the European Commission had designated [25 services](#) as VLOPs and VLOSEs, including social media platforms such as Instagram, X, Google and YouTube.

The DSA requires all online platforms to have notice and action systems in place to allow for user reports on illegal content, including hate speech. [Online platforms](#) have to follow up on users' reports and block access to illegal content. The DSA also requires online platforms to inform law enforcement agencies if they become aware of a serious criminal offence involving a threat to life or safety. Online platforms can also voluntarily remove content that violates their community standards. However, they cannot be forced to proactively search their users' posts for potentially illegal content.

Moreover, the DSA introduced strict rules for VLOPs with more than 45 million users per month in the EU. These platforms and search engines must assess and mitigate systemic risks, including concerning the spread of illegal hate speech. Large online platforms also need to increase transparency, including by providing information on the functioning of their algorithms and recommending systems.

Code of Conduct on countering illegal hate speech online

To tackle the spread of illegal hate speech online, the Commission drew up a '[Code of conduct on countering illegal hate speech online](#)' in 2016. Major information technology companies signed up to the voluntary Code of Conduct. By signing up, these companies committed to ensuring that online platforms do not offer opportunities for illegal online hate speech to spread virally. More specifically, they agreed to put clear and effective processes in place to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content. Upon receipt of a valid removal notification, companies committed to review such requests against their rules and community guidelines and, where necessary, national laws transposing Framework Decision 2008/913/JHA, with dedicated teams reviewing requests. Moreover, they agreed to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary.

In January 2025, the Commission published a revised Code of Conduct, the '[Code of Conduct+](#)', which it developed together with the signatories of the original Code of Conduct, national authorities and civil society organisations. Dailymotion, Facebook, Instagram, Jeuxvideo.com, LinkedIn, Microsoft hosted consumer services, Snapchat, Rakuten Viber, TikTok, Twitch, X and YouTube all signed the Code of Conduct+.

The Code of Conduct+ has been integrated into the framework of the DSA, as stipulated by Article 45 DSA, and will strengthen the way online platforms deal with content that EU and national laws define as illegal hate speech. For signatories who are designated as VLOPs, this may help to ensure that appropriate risk mitigation measures are put in place. In addition, VLOPs are subject to an obligatory annual audit under the DSA to verify their compliance with commitments under the Code of Conduct+.

The signatories of the Code of Conduct+ commit to:

- 'Allow a network of 'monitoring reporters', which are not-for-profit or public entities with expertise on illegal hate speech, to regularly monitor how the signatories are reviewing hate speech notices (monitoring reporters may include entities designated as 'trusted flaggers' under the DSA).
- Undertake best efforts to review at least two thirds of hate speech notices received from monitoring reporters within 24 hours.
- Engage with well-defined and specific transparency commitments as regards measures to reduce the prevalence of hate speech on their services, including through automatic detection tools.
- Participate in structured multi-stakeholder cooperation with experts and civil society organisations that can flag the trends and developments of hate speech they observe, helping to prevent waves of hate speech from going viral.
- Raise, in cooperation with civil society organisations, users' awareness about illegal hate speech and the procedures to flag illegal content online'.

Major US social media companies' hate speech policies

[Social media companies](#) have put content moderation policies in place defining what is and is not allowed on their online platforms. These policies are intended to ensure that content is in line with specific platforms' guidelines and does not harm users. Violations of these policies can result in offensive content being removed or fake accounts being closed. Internet platforms proactively monitor content, to deal with topics including hate speech. Facebook, for example, uses [technology](#) and human reviewers to find, review, and take action on content that may go against the platform's community standards (see below). Similarly, Google (owner of YouTube) takes a [range of actions](#) to enforce their policy.

In the EU, the DSA laid down rules for content moderation. Online platforms must implement measures that prevent the spread of illegal content, and protect their users from harm. However, platforms are not required to monitor proactively the content they display.

FRA 2023 report on online hate

A European Union Agency for Fundamental Rights (FRA) [2023 report](#) on online hate revealed that more than half the posts included in the study, which were already assessed by content moderation tools, are still considered hateful by human coders. As misogyny was the most prevalent form of online hate across platforms (in all platforms included in the study, including X and YouTube), the FRA proposed to include it in the systemic risks to be addressed by VLOPs.

Meta's hateful content policy

According to Meta's [community standards](#) (which outline what is and is not allowed on Facebook, Instagram, Messenger and Threads), the company defines [hateful conduct](#) as direct attacks against people – rather than concepts or institutions – on the basis of what it calls protected characteristics (PCs): race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity, and serious disease. Additionally, Meta considers age a protected characteristic when referenced along with another protected characteristic. Meta also protects 'refugees, migrants, immigrants, and asylum seekers from the most severe attacks', though the company does allow commentary on and criticism of immigration policies. Meta removes 'dehumanising speech, allegations of serious immorality or criminality, and slurs'. The company also removes 'harmful stereotypes', defined as 'dehumanising comparisons that have historically been used to attack, intimidate, or exclude specific groups'. Moreover, the company removes 'serious insults, expressions of contempt or disgust, cursing, and calls for exclusion or segregation when targeting people based on protected characteristics'.

On 7 January 2025, Marc Zuckerberg, CEO of Meta, [announced](#) that, starting in the US, the company would end its third-party fact-checking programme and replace it with community notes. Referring explicitly to the 'cultural tipping point' of the election of Donald Trump to the US presidency 'towards once again prioritising speech', Zuckerberg announced that the company would 'allow more speech' and 'restore free expression' by lifting restrictions on some topics that are 'part of mainstream discourse' (such as racism, immigration and gender identity) and focus enforcement on 'illegal and high-severity violations', such as terrorism, child sexual exploitation, drugs, fraud and scams. Meta accordingly revised its content moderation policy by updating the company's hateful conduct policy (the [7 January 2025](#) version of which shows the revisions introduced, including changing the reference from hate 'speech' to 'hateful conduct').

Critics have noted that, 'under Meta's newly [relaxed moderation policies](#), women can be compared to [household objects](#), ethnic groups can be called "filth", users can call for the exclusion of [LGBTQ people](#) from certain professions and [people can refer](#) to a transgender or non-binary person as an "it"'. Announcing the changes, Mark Zuckerberg praised the US for what he [described](#) as the 'strongest constitutional protections for free expression in the world'. He contrasted that with the EU, where he claimed there was 'an ever-increasing number of laws institutionalising censorship and making it difficult to build anything innovative there'.

Proceedings against Facebook and Instagram under the DSA related to hate speech

In April 2024, the European Commission opened [formal proceedings](#) against Facebook and Instagram under the DSA, for non-compliance with Meta's notice-and-action mechanism to flag illegal content, among other issues. Proceedings, which are ongoing, were preceded by a [request for information](#) under the DSA in October 2023, which included investigating the measures the company put in place to prevent dissemination of hate speech.

X's hateful content policy

The X platform (formerly Twitter) [prohibits](#) the promotion of hateful content globally. Examples of hateful content cited in the company's hateful content policy include: hate speech or advocacy against a protected group, individual, or organisation based on, but not limited to, the following: race, ethnicity, colour, national origin, sexual orientation, sex, gender identity, religious affiliation, age, disability, medical or genetic condition, status as a veteran, refugee or immigrant. Hateful content also includes organisations, groups, or individuals associated with promoting hate, criminal, or terrorist-related content; promotion of, support for, or affiliation with, organisations, groups or individuals associated with promoting hate, criminal and/or terrorist-related content; and degrading, mocking, or harassing references to events or practices that negatively affected a protected group.

According to Elon Musk, who acquired X in 2022, hate speech on the platform [decreased](#) after his acquisition (finalised on [28 October 2022](#)). However, a November 2022 [study](#) found that hate speech 'spiked' after Musk acquired the platform. A February 2025 [study](#) comparing posts on X before and after Musk's takeover found an [increase](#) in hate speech 'in the months immediately following Musk's acquisition'. The latter study also found that the number of 'likes' on hate posts doubled, indicating a worrying increase in user interaction with hateful material.

However, [research](#) published in October 2024 found evidence of a more nuanced approach by X to hate and violent speech after Musk's takeover. While X removed a passage which specifically prohibited the 'misgendering or deadnaming of transgender individuals' from its hateful conduct policy, other policies, such as on violent threats or violent content have been strengthened.

Proceedings against X under the DSA related to hate speech

In October 2023, the Commission sent a [request for information](#) under the DSA to X, on spreading of illegal content, such as terrorist and violent content and hate speech. In December 2023, the Commission opened [official proceedings](#) against X under the DSA, including on non-compliance with the DSA obligations related to countering the dissemination of illegal content. The proceedings are ongoing.

Google's hate speech policy

Hate speech is not allowed on [YouTube](#). Google, YouTube's parent company, does not allow content that promotes violence or hatred against individuals or groups based on any of the following attributes, which indicate a protected group status under YouTube's policy: age, caste, ethnicity, or race, disability, immigration status, nationality, religion, sex, gender, or sexual orientation, veteran status, and victims of a major violent event and their kin. However, [research](#) shows that the platform plays a key role in enabling audiences to access potentially harmful content from radical channels. The company has never [included](#) fact-checking as part of its content moderation practices.

Under the EU Code of Conduct on Disinformation, which will become a co-regulatory [instrument](#) under the DSA on 1 July 2025, Google is required to commit to third-party fact-checking. However, the company has informed the Commission that it will not commit to adding fact checks to search results and YouTube videos, or use them in ranking or removing content. Instead, the company is introducing [community-sourced notes](#) (only available in the United States) to add context to videos. This could potentially have an impact on the content displayed and consequently also on hate speech.

Request for information on hate speech to You Tube under the DSA

The European Commission sent a [request for information](#) under the DSA to You Tube in October 2024 on details of its recommender systems, including on their potential influence on hate speech.

US criticism of the regulation of social media companies

[Republican Party](#) members and big US technology firm representatives have criticised EU efforts to regulate online content for some time. However, this criticism has taken on a new dimension under the Trump administration, which has committed itself to 'restoring freedom of speech and ending federal censorship'. In an executive order issued on 20 January 2025, the [White House](#) criticised the Biden administration for 'trampling on free speech rights by censoring Americans' speech on online platforms'. Regarding the EU's regulation of online content, in September 2024 JD Vance, then a US Senator and now Vice-President, [suggested](#) withdrawing the US from NATO if the EU tries to regulate US companies (in this case, X). He stated that '[i]t's insane that we would support a military alliance if that military alliance isn't going to be pro-free speech. ... American power comes with certain strings attached. One of those is respect for free speech'. At the Munich Security Conference in February 2025, Vice-President Vance accused the EU of [curtailing free speech](#), going as far as calling content moderation '[authoritarian censorship](#)'. In March 2025, the chair of the US Federal Communications Commission (FCC), Trump-appointed Republican Brendan Carr, [called](#) the EU's content moderation law 'incompatible with America's free speech tradition' and warned of 'a risk that it will excessively restrict freedom of expression'. Members of the European Parliament [travelled](#) to Washington in February 2025 to meet US officials and members of Congress and to explain what EU regulation of online content is trying to achieve.

MAIN REFERENCES

- Alkaviadou, N., [Case law on Hate Speech: The Enduring Question of Thresholds](#), Columbia University, June 2023.
- Casarosa, F., Moraru, M., [Freedom of expression and countering hate speech](#), Handbook on Techniques of Judicial Interaction in the Application of the EU Charter, European University Institute, May 2020.
- Cho, C., Zhu, L., [Social Media: Content Dissemination and Moderation Practices](#), Congressional Research Service, March 2025.
- European Court of Human Rights, [Fact Sheet on Hate Speech](#), November 2023.
- Immenkamp, B., [Hate speech and hate crime – time to act?](#), EPRS, September 2024.
- Immenkamp, B., [Criminalisation of hate speech and hate crime in selected EU countries](#), EPRS, November 2024.
- Killion, V., [The First Amendment: Categories of Speech](#), Congressional Research Service, March 2024.
- Madiega, T., [Digital services act](#), EPRS, November 2022.
- McKeown, M., [Hate Speech: A Comparative Analysis of the United States and Europe](#), in Regulating Cyber Technologies, 2023.
- Velenchuk, T., [Freedom of expression, a comparative law perspective – the United States](#), EPRS, October 2019.
- Windwehr, S., [Trump vs. Europe: The role of the Digital Services Act](#), Heinrich Böll Foundation, 18 February 2025.

DISCLAIMER AND COPYRIGHT

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

© European Union, 2025.

Photo credits: © Stillfx / Adobe Stock.

eprs@ep.europa.eu (contact)

<https://eprs.in.ep.europa.eu> (intranet)

www.europarl.europa.eu/thinktank (internet)

<http://epthinktank.eu> (blog)