



A CURE WORSE THAN THE DISEASE: WHY THE EU’S PROPOSAL TO
CRIMINALIZE HATE SPEECH VIOLATES FREEDOM OF EXPRESSION—AND
WON’T REDUCE INTOLERANCE OR EXTREMISM

Jacob Mchangama & Natalie Alkiviadou***

Introduction.....	1185
I. The Proposal and its Trajectory.....	1190
A. Existing EU Framework and the Commission’s Enforcement Posture	1191
1. From “procedural steps” to likely scope	1192
2. Legislative pathway and Parliament’s push for open-ended expansion.....	1194
3. Current status and institutional momentum.....	1194
II. Legal Analysis of the Proposal to Harmonize Online Hate Offences at EU Level.....	1195
A. The European Court of Human Rights.....	1196
B. The ICCPR Framework	1200
1. Legality: The lack of clear and foreseeable conceptual framework.....	1204
2. Necessity: An unproven need for EU-level criminalization ...	1206
3. Proportionality: Criminal law as an excessive response.....	1209
4. Summary	1210

* Jacob Mchangama, Executive Director, The Future of Free Speech at Vanderbilt University, and a Senior Fellow at the Foundation for Individual Rights and Expression.

** Natalie Alkiviadou, Senior Research Fellow, The Future of Free Speech at Vanderbilt University.

III.	Europe's Existing Hate Speech Framework: Scope, Enforcement, and Escalatory Dynamics	1211
A.	Historical Context: Europe's Post-War Resistance to the Criminalization of Expression	1212
B.	National Criminal Law: From Incitement to Open-Ended Prohibitions	1214
C.	Enforcement Practices: From Exceptional Sanction to Routine Intrusion	1216
D.	Scope Creep and the Targeting of Minorities and Internal Dissent.....	1220
E.	EU-Level Pressure and the Escalation of National Criminalization	1222
F.	Interaction with the Digital Services Act: Platform Enforcement as an Accelerant of Scope Creep	1224
G.	Synthesis: From National Scope Creep to EU-Level Entrenchment	1225
IV.	Speech-Intrusive Approaches to Hate Speech: Empirical Challenges and Unintended Effects	1226
A.	The Missing Causal Link between Criminalization and Reduced Harm	1227
B.	Repression, Radicalization, and the "Martyr Effect"	1230
C.	Free Expression as a Safety Valve and Predictor of Tolerance	1232
D.	Trust, Legitimacy, and Reporting Gaps	1233
E.	Implications for Necessity and Proportionality	1234
V.	The External Dimension of European Hate Speech Regulation	1234
VI.	Alternative Approaches to the European Criminalization of Hate Speech	1238
A.	The Case for Counterspeech	1238
B.	The EU Non-Discrimination Framework	1241
	Conclusion	1242

INTRODUCTION

In Germany, an artist and activist was arrested multiple times, once for wearing a Palestinian flag and again for holding a sign that read “from the river to the sea, we demand equality.”¹ In France, a mayor was held criminally liable, not for what he said, but for failing to remove Islamophobic comments posted by others beneath one of his Facebook posts.² In Finland, a member of Parliament faced criminal charges for criticizing her church’s support for gay pride events, citing a Bible verse.³

Such cases are not exceptional among European Union member states. In many places, they are standard practice.

Yet despite already possessing some of the democratic world’s most expansive and actively enforced hate-speech laws, the European Commission has proposed adding hate speech to the EU’s list of “EU crimes” under Article 83(1) of the Treaty on the Functioning of the European Union.⁴ In 2020, the European Commission (Commission) President Ursula von der Leyen proposed making hate speech and hate crime European Union (EU)-level offenses. The proposal is framed as part of the Union’s broader efforts to promote equality, inclusion, and non-discrimination.⁵ Yet it also forms part of a wider trend within EU institutions and Member States, namely a growing belief that expansive restrictions on expression are necessary to protect democratic values, counter extremism, and address social harms.

Since 2021, the Commission has reaffirmed the initiative in formal communications and public statements, with strong support from the European Parliament and other EU bodies, although unanimity in the Council of the European Union

¹ James Jackson, ‘We Jews Are Just Arrested; Palestinians Are Beaten’: Protestors in Germany, AL JAZEERA (Apr. 1, 2024).

² Sanchez v. France, App. No. 45581/15 (Eur. Ct. H.R. May 15, 2023), <https://perma.cc/C8NB-Y53A>.

³ Paul Coleman, *Päivi Räsänen Is the Face of Europe’s Free Speech Crisis*, CRITIC (Nov. 4, 2025), <https://perma.cc/EH9P-K4G6>.

⁴ *Communication from the Commission to the European Parliament and the Council: A More Inclusive and Protective Europe: Extending the List of EU Crimes to Hate Speech and Hate Crime*, COM (2021) 777 final (Dec. 9, 2021), <https://perma.cc/T92A-C8B9>.

⁵ *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: On the European Democracy Action Plan*, COM (2020) 790 final (Dec. 3, 2020), <https://perma.cc/23M5-TT3H>.

has not yet been secured. In its 2026–2030 Anti-Racism Strategy, the Commission signaled that, given the lack of progress, it is considering a legislative initiative to harmonize definitions of online hate offences by drawing on existing areas of crime covered by Article 83(1) of the Treaty on the Functioning of the European Union (TFEU).⁶ The shift from treaty expansion to reliance on the existing legal authority of Article 83 reflects not a retreat from harmonization, but a change in legal technique. Where unanimity failed, pursuing similar objectives through digital or computer-crime competences warrants scrutiny as a potential competence-stretching bypassing Member State resistance.

At the time of this report, the proposal has stalled due to a lack of unanimity in the European Council. But the Commission has signaled a narrower, “online-only” route: a legislative initiative under existing Article 83(1) crime categories (*e.g.*, computer crime) to harmonize definitions of online hate offences. In practice, this would still embed criminal prohibitions more deeply into national criminal law and EU digital enforcement frameworks. It also raises questions about whether speech that might be considered legal when said online suddenly becomes subject to criminal penalties if posted online.

The Commission seeks to justify the new initiative by pointing to an increase in hate speech and hate crimes across the EU.⁷ It also highlights divergent national definitions and enforcement practices, arguing that fragmentation undermines victim protection, creates enforcement gaps, and sends “mixed messages” about the seriousness of hate speech.⁸ It thus frames the proposal not as a novel expansion but as the logical next step in building a coherent European response that would enable the EU to address serious forms of hate speech more effectively, particularly online.

While it may be unclear how a EU-wide criminalization of hate speech will ultimately take shape, the Commission’s proposal fails to meet the foundational re-

⁶ *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Union of Equality: Anti-Racism Strategy 2026–2030*, COM (2026) 12 final (Jan. 20, 2026), <https://perma.cc/GVL7-SGLX>.

⁷ *A More Inclusive and Protective Europe*, *supra* note 4; *Joint Communication to the European Parliament and the Council: No Place for Hate: A Europe United Against Hatred*, JOIN (2023) 51 final (Dec. 6, 2023), <https://perma.cc/4ERF-64GZ>.

⁸ *A More Inclusive and Protective Europe*, *supra* note 4, ¶ 3.3.

quirements of legality, necessity, and proportionality under both European and international human rights law. It further finds that the empirical assumptions underpinning the initiative are unsubstantiated and, in some respects, contradicted by available evidence.

Legality: The principle of legality requires that criminal offenses be formulated with sufficient clarity and precision to enable individuals to regulate their conduct and to foresee, to a reasonable degree, the legal consequences of their actions. This requirement is especially stringent in the case of speech, because uncertainty itself chills expression.

The European Commission's proposal offers no conceptualization of "hate speech." Instead, it relies on broad, vague, and varying moral, identity-based, and sociological categories. The European Parliament has further encouraged the adoption of an open-ended list of protected characteristics, creating a system in which virtually any political, social, or ideological group could invoke criminal law to shield itself from alleged hate speech. If such definitional vagueness is adopted in any subsequent directive, then it will be incompatible with legal certainty, invite arbitrary enforcement, and risk chilling democratic debate on topics of public interest across all 27 Member States.

Necessity: Necessity requires showing that criminalization at the EU level is indispensable, and that less intrusive measures are insufficient to address the identified harm.

The Commission has not demonstrated that EU-level criminalization or harmonization of online hate offences is necessary. Member States already rely on extensive criminal and regulatory tools, including national speech offences, the 2008 Framework Decision on combating racism and xenophobia through criminal law, and the Digital Services Act.

Moreover, the political rationales underpinning the initiative have shifted over time, from the COVID-19 pandemic⁹ to post-October 7 tensions,¹⁰ suggesting a sense of urgency driven by the prevailing political climate rather than by a demonstrable legislative need. And empirical research provides little evidence that the criminalization of hate speech reduces hatred or violence. The empirical literature

⁹ *Id.* ¶¶ 1, 3.4.

¹⁰ BEATRIX IMMENKAMP ET AL., EUR. PARLIAMENTARY RSCH. SERV., PE 762.389, HATE SPEECH AND HATE CRIME: TIME TO ACT? 1 (2024), <https://perma.cc/VQ6K-KBUQ>.

does not establish that broader criminalization reduces hatred or violence in stable democracies, and it identifies plausible counterproductive dynamics (polarization, grievance narratives, displacement). Conversely, stronger expressive freedoms are associated with higher tolerance and lower social conflict in democratic settings.

Proportionality: Proportionality requires that restrictions on expression pursue a legitimate aim using the least intrusive means available. Criminal sanctions represent the most severe form of state interference with speech and therefore demand especially compelling justification. Even under the European Court of Human Rights' comparatively deferential approach to national hate speech laws, criminal sanctions must remain proportionate and narrowly tailored.

While the Commission's Communications briefly acknowledge the importance of freedom of expression, they omit core safeguards required under International Human Rights Law, including intent, imminence, and likelihood of harm. The proposal thus falls short of the strict incitement standard set out in Article 20(2) of the International Covenant on Civil and Political Rights, which must conform with the strict requirements for restricting freedom of expression under Article 19. Article 20(2) mandates only the prohibition of any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, as interpreted by the Rabat Plan of Action.

As a result, the proposal risks criminalizing even "shocking, offensive, or disturbing" speech that is protected under Article 10 of the European Convention on Human Rights (the right to freedom of expression). In practice, this could expose journalists, academics, artists, politicians, activists, and others to criminal liability for expression that is currently lawful in many Member States. And because the Digital Services Act defines "illegal content" by reference to EU and national law, an EU-wide regime harmonizing hate speech or online hate offences would expand platform removal and risk mitigation duties at scale, strengthening incentives to over remove lawful but controversial speech—especially where automated systems struggle with context (satire, minority speech, political critique).

An online-only offence would also risk creating a two-tier criminal regime: Speech that is lawful when delivered offline could trigger criminal liability once posted online. Because many contemporary prosecutions concern online comments, Member States would face strong legal and political pressure to align their general criminal codes with the online definition, effectively achieving the criminal

law harmonization that some States have resisted. If they do not, the result is unequal treatment and “digital disappearance.” For instance, a lecture could be lawful in a hall and reportable in the press yet be removed from social media (and even expose the speaker or person posting the lecture on social media to prosecution) when uploaded.

International Effects: The proposal also risks reshaping global norms. European approaches to online speech and hate-speech regulation already exert significant influence on countries with weak rule-of-law protections. Laws in Russia, Turkey, Venezuela, and other jurisdictions have explicitly cited European models when criminalizing or suppressing dissent—even when the effect of implementing the European models in the political systems of those countries is much more speech-restrictive than it would be in Europe. Through the so-called “Brussels Effect,” an EU-level regime harmonizing online hate offences would further entrench this dynamic, irrespective of the European Union’s pro-democratic intentions.

A More Effective and Rights-Compatible Alternative: Considering the broad range of legal tools already available to Member States to address harmful speech, the European Union should prioritize non-punitive interventions such as counter-speech. This method can foster social resilience without compromising the right to freedom of expression, which underpins democratic life and serves as a safeguard against government overreach, even when pursued with well-intentioned aims.

At a time of heightened political polarization and democratic backsliding, the European Union must reaffirm its commitment to open debate, pluralism, and restraint as a cornerstone of democratic society. The Commission’s proposal risks normalizing a model of speech governance that suppresses democratic debate rather than strengthening it. The European Union should regard freedom of expression as an essential safeguard of freedom and democracy—not a “risk” to be managed.

* * *

Rather than strengthening democratic resilience, EU-level criminalization of hate speech or harmonization of online hate offences risks entrenching overbroad enforcement, chilling lawful expression, and amplifying the very social tensions it purports to address. Accordingly, the EU should focus on strengthening rather than weakening freedom of expression and ensure that the criminalization of hate speech is used only as a last resort and in ways that complies strictly with the requirements of legality, necessity and proportionality. The EU should also refocus

its crucial efforts to combat hatred and discrimination with an increased emphasis on using non-restrictive means such as counterspeech and education.

Our analysis proceeds from the basic premise that all member states of the EU criminalize certain forms of hate speech and that such criminalization, subject to balancing against the human right to freedom of expression, is both required and compatible with EU law and International Human Rights Law. Accordingly, we assume for our purposes that the repeal of European hate speech laws is both legally and politically unfeasible, even if there are compelling normative and empirical arguments in favor of such a maximalist position. Instead, we examine the proposal through the foundational principles governing restrictions on freedom of expression in European and International Human Rights Law (IHRL): legality, necessity, and proportionality. We assess whether EU-level criminalization is compatible with these standards, whether the need for such a proposal is supported by evidence, and the proposal's implications for democratic debate, online expression, and Europe's broader global influence.

In the following Parts, we situate the initiative's institutional trajectory (Part I); assess its legality, necessity, and proportionality under European and international human rights standards (Part II); analyze Europe's existing hate-speech landscape and escalation dynamics, including the DSA multiplier (Part III); test the proposal's empirical premises (Part IV); examine external diffusion through the "Brussels Effect" (Part V); and develop rights-compatible alternatives (Part VI)

I. THE PROPOSAL AND ITS TRAJECTORY

The Commission's initiative took shape in September 2020, when President Ursula von der Leyen proposed extending the list of "EU crimes" under the TFEU. Article 83(1) allows the EU to adopt directives establishing minimum rules on criminal offences in areas of particularly serious crime with a cross-border dimension. In her 2020 speech, von der Leyen framed the initiative as part of the EU's commitment to build "a truly anti-racist Union," noting that the Commission "will propose to extend the list of EU crimes to all forms of hate crime and hate speech."¹¹ With that Council-decision track currently stalled, the Commission has increasingly pointed to an alternative path: harmonizing definitions of online hate offences under existing Article 83(1) crime areas.

¹¹ Ursula von der Leyen, President, Eur. Comm'n, 2021 State of the Union Address at the European Parliament Plenary (Sep. 15, 2021), <https://perma.cc/2USN-NRQM>.

The proposal is embedded in the Commission's broader equality agenda. In this context, the EU Gender Equality Strategy, the LGBTIQ+ Equality Strategy, and Anti-Racism Action Plan identify sexist, homophobic, transphobic, and racist expressions as obstacles to inclusion. From this perspective, the proposal presents this new EU crime as a complement to existing policies.¹²

A. Existing EU Framework and the Commission's Enforcement Posture

Before turning to the substance of existing legislation, it is helpful to clarify how authority is divided within the European Union. As a rule, EU law takes precedence over national law in areas where the EU has been given power, and, in some cases, individuals can rely on EU law directly before national courts.¹³ However, this does not mean that the EU overrides national law across the board. The EU can act only within the powers that its Member States have conferred upon it. Primary authority over criminal law remains with the Member States, which define and enforce hate speech offences within their own legal systems.

At the same time, the European Union may, under Article 83(1) TFEU, adopt directives establishing minimum rules on criminal offences, thereby harmonizing aspects of national law without fully displacing it. Alongside these layers, the Digital Services Act introduces a distinct but closely connected regulatory framework by requiring online platforms to remove "illegal content" as defined under EU or national law, effectively extending enforcement through private actors. Together, these elements form a cumulative system in which EU-level developments can significantly shape both the scope and the practical enforcement of hate speech regulation across the Union.

¹² *Gender Equality Strategy 2020–2025*, EUR. COMM'N, <https://perma.cc/NYZ3-AWTZ>; *LGBTIQ+ Equality Strategy 2026–2030*, EUR. COMM'N, <https://perma.cc/MBE8-SWQD>; *Anti-Racism Strategy*, EUR. COMM'N, <https://perma.cc/WVP9-AW9Q>; *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Union of Equality: EU Anti-Racism Action Plan 2020–2025*, COM (2020) 565 final (Sep. 18, 2020), <https://perma.cc/7Z6W-U27U>; *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Union of Equality: LGBTIQ Equality Strategy 2020–2025*, COM (2020) 698 final (Nov. 12, 2020), <https://perma.cc/SH5J-P8J5>.

¹³ Case 6/64, *Costa v. ENEL*, 1964 E.C.R. 585; Case 26/62, *Van Gend en Loos v. Nederlandse Administratie der Belastingen*, 1963 E.C.R. 1.

The Commission's proposal also sits within a growing patchwork of existing EU and national instruments that already (directly or indirectly) regulate hate and extremist speech. At EU level, these include:

- 2008 Framework Decision on Combating Certain Forms and Expressions of Racism and Xenophobia (Framework Decision),
- EU Code of Conduct on countering illegal hate speech online (revised in 2025),
- Digital Services Act (DSA),
- 2018 Audiovisual Media Services Directive (AVMSD),
- 2018 Regulation on Preventing the Dissemination of Terrorist Content Online (Regulation on Terrorist Content Online), and
- Directive on Combating Violence against Women and Domestic Violence.¹⁴

The Commission has also pursued infringement proceedings against several member states aimed at strengthening national implementation of the Framework Decision and has argued that divergent definitions and enforcement practices create gaps and “mixed messages.”¹⁵

1. From “procedural steps” to likely scope

The initiative was formalized in the Commission's 2021 Communication “A more inclusive and protective Europe” and reiterated in a 2023 Communication entitled “No place for hate: a Europe united against hatred.” The proposal and its

¹⁴ Council Framework Decision 2008/913/JHA of 28 Nov. 2008 on Combating Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law, 2008 O.J. (L 328) 55; *Code of Conduct on Countering Illegal Hate Speech Online*, EUR. COMM'N (Jan. 20, 2025), <https://perma.cc/27ZN-LSB6>; Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 Nov. 2018 Amending Directive 2010/13/EU, 2018 O.J. (L 303) 69; Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 Apr. 2021 on Addressing the Dissemination of Terrorism Content Online, 2021 O.J. (L 172) 79; Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 Oct. 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1 [hereinafter Digital Services Act]; Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on Combating Violence Against Women and Domestic Violence, 2024 O.J. (L 1385).

¹⁵ *A More Inclusive and Protective Europe*, *supra* note 4, ¶ 3.3.

importance have since been reaffirmed by the Commission in less formal frameworks such as public speeches. Other EU bodies, including the European Parliament, the Committee of the Regions, and the European Economic and Social Committee, have publicly supported the proposal. The Parliament's 2024 Resolution urged the Council to proceed with adding hate speech and hate crimes to the list of EU crimes. However, with that track stalled, in January 2026, the Commission indicated that it may instead pursue harmonization of online hate offences under existing Article 83(1) crime areas.

Although formally presented as a procedural first step, the 2021 Communication from the Commission to the Parliament and Council provides clear indications of the likely scope and underlying logic of any future directive. The Commission explicitly envisions expanding protected characteristics beyond those covered by the 2008 Framework Decision, extending protection to a broad, potentially open-ended range of "fundamental characteristic[s]," including sex, sexual orientation, gender identity, age, and disability.¹⁶

Further, the Communication frames hate speech primarily through sociological and identity-based descriptions rather than legally precise criteria capable of delimiting criminal liability. It describes hate as targeting individuals and groups sharing or "perceived as sharing a common characteristic," defined as an aspect of identity that is "unchangeable or fundamental to a person's sense of self," "noticeable to others," and linked to "group identity."¹⁷ Hate speech and hate crimes are further characterized as "'identity' or 'message' acts," whose gravity lies in the symbolic message that targeted persons "do not belong to that society."¹⁸

The Communication asserts that hatred "leads to the devaluation of and threat to the human dignity of a person or a group," "negates their equal footing as members of society," and "poisons the political dialogue."¹⁹ While these descriptions reflect real social concerns, they are overly abstract and prone to subjective interpretation and value judgements. The proposed criteria fail to identify elements necessary for the operation of criminal law, such as intent, imminence, and likelihood of harm.

¹⁶ *Id.* ¶ 1.

¹⁷ *Id.* (citation modified).

¹⁸ *Id.* ¶ 3.1 (citation modified).

¹⁹ *Id.* ¶¶ 3.1, 3.23.

As a result, core assumptions about identity, vulnerability, and group harm are relied upon to justify criminalization, without being translated into the precise and foreseeable criteria required by criminal law and freedom-of-expression standards, with direct implications for legal certainty. Also, the communications do not meaningfully assess the chilling effects of criminal sanctions on lawful political, journalistic, artistic, or minority expression, and they tend to treat criminal law as a default response rather than a measure of last resort.

2. Legislative pathway and Parliament's push for open-ended expansion

The 2021 Communication sets out a two-step legislative process: A unanimous Council decision (not yet achieved), with Parliament's consent, to add hate speech and hate crimes to the list of EU crimes, and a subsequent Directive, establishing minimum rules on definitions and sanctions (which would follow only if the Council Decision were adopted).²⁰

The Council discussed the proposal in March 2022. A large majority of Member States expressed support in principle, but unanimity has not yet been achieved.²¹ In January 2024, the European Parliament adopted a Resolution that went beyond the Commission's framing, explicitly encouraging an "open-ended" approach to protected characteristics—allowing future expansion in response to "new and changing social dynamics."²² Such an approach risks making the boundaries of speech-related criminal offences politically elastic and less foreseeable.

3. Current status and institutional momentum

Despite the absence of unanimity, the initiative remains institutionally active. The Commission continues to affirm its political commitment, while Member State and stakeholder discussions have periodically called for renewed deliberation. At the same time, in its 2026–2030 Anti-Racism Strategy, the Commission has signaled that, given the absence of progress, it is considering a legislative initiative

²⁰ *Id.* ¶ 2.1.

²¹ *Justice and Home Affairs Council*, 3–4 March 2022, COUNCIL EUR. UNION, <https://perma.cc/TF9E-7P5U>.

²² European Parliament Resolution of 18 January 2024 on Extending the List of EU Crimes to Hate Speech and Hate Crime, EUR. PARL. DOC. P9_TA(2024)(0044) (2024), <https://perma.cc/UAQ2-EKNT>.

based on existing areas of crime covered by Article 83(1) TFEU to harmonize definitions of online hate offences.²³ Given the likely interaction with EU digital enforcement architectures (including the DSA), both tracks warrant careful legal and empirical scrutiny.

The Commissioner for Democracy, Justice, the Rule of Law and Consumer Protection, Michael McGrath, has consistently championed the initiative, emphasizing (as late as June 2025) the Commission's intention to "explor[e] all possibilities under the Treaties" in order to revive the measure once unanimity is secured.²⁴ Indeed, during discussions about the proposal at the High-Level Forum on Future of EU Criminal Justice, which took place in December 2025, a majority of Member State representatives and external stakeholders stressed the need to "resume discussions on the criminalization of hate speech and hate crime at EU level."²⁵

The proposal remains pending, but its trajectory demonstrates a clear institutional desire to reshape how hate speech is understood and addressed across the Union. This makes a careful legal and empirical assessment essential before the EU proceeds with what would constitute a major shift in the status and protection of free speech in Europe.

II. LEGAL ANALYSIS OF THE PROPOSAL TO HARMONIZE ONLINE HATE OFFENCES AT EU LEVEL

Any EU initiative that limits freedom of expression—whether framed as a new "EU crime" under Article 83(1) TFEU or as an online-only instrument under an existing Article 83(1) crime area must comply with the foundational principles governing restrictions under European and international human rights law (IHRL): legality, necessity, and proportionality. Although the EU is not a party to the International Covenant on Civil and Political Rights (ICCPR), all EU Member States have ratified it.²⁶ As such, the standards articulated by the ICCPR and the Human Rights Committee (HRC) are directly relevant.

²³ *Union of Equality: Anti-Racism Strategy 2026–2030*, *supra* note 6.

²⁴ Michael McGrath, Eur. Comm'r, *Speech to Mark No Hate Speech Week 2025* (June 18, 2025), <https://perma.cc/6MQ3-R3AN>.

²⁵ EUROPEAN COMMISSION, REPORT OF THE HIGH-LEVEL FORUM ON THE FUTURE OF EU CRIMINAL JUSTICE 8 (2025), <https://perma.cc/QX79-ZPCA>.

²⁶ Status of Treaties: International Covenant on Civil and Political Rights, U.N. TREATY COLLECTION, <https://perma.cc/KTS7-33PR>.

The preamble to the Charter of Fundamental Rights of the European Union (which we'll just call the Charter) explicitly recognizes the rights arising from the constitutional traditions of the Member States and from international obligations.²⁷ The Court of Justice of the European Union (CJEU) has treated major United Nations human rights treaties such as the ICCPR as relevant sources for identifying general principles of EU law and for interpreting rights under the Charter.²⁸ As a result, any effort to harmonize criminal law in the EU must take into account the existing international obligations of Member States.

A. *The European Court of Human Rights*

Article 52 of the Charter provides that the rights it contains have the same meaning and scope as the corresponding rights guaranteed by the European Convention on Human Rights (ECHR), while permitting the EU to afford a higher level of protection. Its preamble reaffirms the rights as they emanate from the case law of the European Court of Human Rights (ECtHR).²⁹ As such, ECtHR case law establishes minimum standards of protection for freedom of expression rather than an exhaustive or maximal framework for EU regulation.

The ECtHR has failed to properly define hate speech, and its approach to relevant cases has been criticized for being unclear, inconsistent, and paternalistic.³⁰ Its case law sets a notably low threshold for permissible restrictions on hate speech, which allows restrictions not only on speech that explicitly calls for violence or other criminal acts, but also on “attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population.”³¹

²⁷ Charter of Fundamental Rights of the European Union, 2012 O.J. (C 326) 391.

²⁸ See, e.g., Case C-274/87, *Orkem v. Comm'n*, 1989 E.C.R. 3283; *Parliament v. Council*, 2006 E.C.R. I-5769.

²⁹ Charter of Fundamental Rights of the European Union, *supra* note 27.

³⁰ ADAM TOMKINS, ON THE LAW OF SPEAKING FREELY 115, 117–18, 122 (2025); Stefan Sottiaux, *Conflicting Conceptions of Hate Speech in the ECtHR's Case Law*, 23 GER. L.J. 1193 (2022); Alessio Sardo, *Hate Speech: A Pragmatic Assessment of the European Court of Human Rights' Jurisprudence*, 2022 EUR. CONV. HUM. RTS. L. REV. 1; Jacob Mchangama & Natalie Alkiviadou, *Hate Speech and the European Court of Human Rights: Whatever Happened to the Right to Offend, Shock or Disturb?*, 21 HUM. RTS. L. REV. 1008 (2021); Hannes Cannie & Dirk Voorhoof, *The Abuse Clause and Freedom of Expression in the European Human Rights Convention*, 29 NETH. Q. HUM. RTS. 54 (2011).

³¹ *Vejdeland v. Sweden*, App. No. 1813/07, ¶ 55 (Eur. Ct. H.R. Feb. 9, 2012), <https://perma.cc/BU7A-UKCX>.

In *Féret v. Belgium* (2009), a politician was convicted for distributing Islamophobic materials, including slogans such as “Stop the Islamization of Belgium.”³² The Court held that incitement to hatred does not require calls for violence: insults, ridicule, or defamation targeting specific groups, or incitement to discrimination, may suffice where expression undermines dignity or safety.³³ The Court also found that Féret’s statements were “inevitably of such a nature as to arouse, particularly among the less informed members of the public, feelings of distrust, rejection or hatred towards foreigners.”³⁴ Such assumptions risk justifying extensive restrictions on political speech by portraying democratic audiences as inherently vulnerable and manipulable.

The Court has also invoked Article 17 (the prohibition of abuse of rights) to place certain expressions outside Convention protection, effectively bypassing Article 10 safeguards—especially proportionality review. It has done so consistently in Holocaust-denial cases, including ones involving private communications. Yet in *Perinçek v. Switzerland* (2015), concerning the denial of the Armenian genocide, it declined to apply Article 17.³⁵ Instead, it found that the restriction on speech violated Article 10, an inconsistency that has fueled criticism that Article 17 lacks clear limiting principles in the hate-speech context.³⁶ In *Norwood v. United Kingdom* (2004), the Court applied Article 17 to exclude from Article 10 protection a poster depicting the Twin Towers in flames with the slogan “Islam out of Britain—Protect the British People.”³⁷ And in *Lenis v. Greece* (2023), it treated a church official’s inflammatory remarks as literal calls to violence, again removing them from Article 10’s protection.³⁸

³² *Féret v. Belgium*, App. No. 15615/07 (Eur. Ct. H.R. July 16, 2009), <https://perma.cc/X686-QE5W>.

³³ *Id.* ¶ 73.

³⁴ *Id.* ¶ 69.

³⁵ *Perinçek v. Switzerland*, App. No. 27510/08, ¶ 115 (Eur. Ct. H.R. Oct. 15, 2015), <https://perma.cc/2WF3-YV28>.

³⁶ *Id.* ¶ 237.

³⁷ *Norwood v. United Kingdom*, App. No. 23131/03 (Eur. Ct. H.R. Nov. 16, 2004), <https://perma.cc/SBF2-NWEP>.

³⁸ *Lenis v. Greece*, App. No. 47833/20 (Eur. Ct. H.R. June 27, 2023), <https://perma.cc/WS6N-YEKZ>.

All in all, ECtHR case law on hate speech is systematically less protective of freedom of expression than the approach developed under Articles 19 and 20(2) of the ICCPR. In the words of former Special Rapporteur David Kaye and Evelyn Aswad, “the European Court of Human Rights frequently lessens or removes the burden on governments to show hate speech restrictions are properly imposed, allows for the imposition of hate speech restrictions for reasons not accepted at the global level, and does not assess whether restrictions on speech are the least intrusive means to achieving legitimate public interest objectives.”³⁹

ECtHR jurisprudence has, at times, blurred the line between hate speech and blasphemy by treating the protection of religious believers from offense as a legitimate basis for restricting expression. In *E.S. v. Austria* (2018), the Court upheld criminal sanctions for calling the Prophet Muhammad a pedophile, emphasizing “that presenting objects of religious worship in a provocative way capable of hurting the feelings of the followers of that religion could be conceived as a malicious violation of the spirit of tolerance” and that the comments were “capable of stirring up prejudice and putting religious peace at risk” and “contained elements of incitement to religious intolerance.”⁴⁰ *Norwood* (the “Islam out of Britain” case discussed above) is another case that blurs the distinction between attacks on a religion and its followers.

Despite its intrusive approach to hate speech, the ECtHR has articulated safeguards regarding the criminalization of speech, which raise issues for the Commission’s proposal. The ECtHR has long recognized that freedom of expression under Article 10 ECHR extends

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.’⁴¹

³⁹ Evelyn Aswad & David Kaye, *Convergence & Conflict: Reflections on Global and Regional Human Rights Standards on Hate Speech*, 20 NW. J. HUM. RTS. 165, 165 (2022).

⁴⁰ *E.S. v. Austria*, App. No. 38450/12, ¶ 53, ¶ 57 (Eur. Ct. H.R. Oct. 25, 2018), <https://perma.cc/WS9R-X527>.

⁴¹ *Handyside v. United Kingdom*, 24 Eur. Ct. H.R. (ser. A) 49 (1976).

The right to hyperbolic and provocative language and speech is a central part of political speech, which protects polemical, sarcastic, and satirical language.⁴² A certain degree of exaggeration is broadly understood to be accepted, particularly in discussions on political issues.⁴³

Legal certainty and foreseeability are central requirements for any restriction on freedom of expression, particularly where criminal sanctions are involved. In *Sunday Times v. United Kingdom* (1979), the Court highlighted that laws must be “formulated with sufficient precision to enable the citizen to regulate his conduct . . . to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail.”⁴⁴ In *Savva Terentyev v. Russia* (2018), the Court stressed that:

it is vitally important that criminal law provisions directed against expressions that stir up, promote or justify violence, hatred or intolerance clearly and precisely define the scope of relevant offences, and that those provisions be strictly construed in order 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.⁴⁵

Proportionality requires that restrictions on expression constitute the least intrusive means of achieving a legitimate aim. The Court has consistently emphasized that criminal sanctions have a pronounced chilling effect on freedom of expression. In *Mariya Alekhina v. Russia* (2018), it held that “interference with freedom of expression in the form of criminal sanctions may have a chilling effect on the exercise of that freedom, which is an element to be taken into account when assessing the

⁴² *Lopes Gomes da Silva v. Portugal*, App. No. 37698/97, ¶ 35 (Eur. Ct. H.R. Sept. 28, 2000), <https://perma.cc/E7D6-6L27>; *Katrami v. Greece*, App. No. 19331/05 (Eur. Ct. H.R. Dec. 6, 2007), <https://perma.cc/SS9J-QSZ7>; *Eon v. France*, App. No. 26118/10, ¶ 61 (Eur. Ct. H.R. Mar. 14, 2013), <https://perma.cc/697W-JN2P> (“The Court considers that criminal penalties for conduct such as that of the applicant in the present case are likely to have a chilling effect on satirical forms of expression relating to topical issues. Such forms of expression can themselves play a very important role in open discussion of matters of public concern, an indispensable feature of a democratic society.”).

⁴³ *Dabrowski v. Poland*, App. No. 18235/02, ¶ 35 (Eur. Ct. H.R. Dec. 19, 2006), <https://perma.cc/6MST-VDR4>.

⁴⁴ *Sunday Times v. United Kingdom*, App. No. 6538/74, ¶ 49 (Eur. Ct. H.R. Apr. 26, 1979), <https://perma.cc/7M8J-5W26>.

⁴⁵ *Savva Terentyev v. Russia*, App. No. 10692/09, ¶ 85 (Eur. Ct. H.R. Aug. 28, 2018), <https://perma.cc/3RG8-8H3C>.

proportionality of the interference in question.”⁴⁶ The Court underlined that “alternative preventive measures” should have been considered.⁴⁷ In *Lehideux & Isorni v. France* (1998), the ECtHR noted “the seriousness of a criminal conviction for publicly defending the crimes of collaboration, having regard to the existence of other means of intervention and rebuttal, particularly through civil remedies.”⁴⁸ This jurisprudence underscores that criminal law should remain a measure of last resort when regulating speech.

Despite the ECtHR’s general acceptance of hate speech laws, the Commission’s proposal to expand and harmonize criminalization of hate speech may therefore run afoul of core ECtHR doctrines on legality and proportionality. Even so, the ECtHR’s deferential approach to national hate speech prohibitions cannot be the sole measure of the proposal’s compatibility with the fundamental human right to freedom of expression. It must also consider IHRL, which provides a significantly more speech-protective framework.

B. *The ICCPR Framework*

Article 19(2) of the ICCPR guarantees “[e]veryone . . . the right to freedom of expression,” which includes “the freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media.”⁴⁹ Article 19(3) allows restrictions but only if “provided by law” and “necessary” for respect of the rights or reputations of others, national security, public order, or public health or morals.⁵⁰

While Article 19(3) *permits* restrictions of free expression, ICCPR Article 20 (2) *obliges* ratifying States to prohibit by law “[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.”⁵¹ The ICCPR is interpreted authoritatively by the HRC, an independent

⁴⁶ *Mariya Alekhina v. Russia*, App. No. 38004/12, ¶ 227 (Eur. Ct. H.R. July 17, 2018), <https://perma.cc/98XB-HTLR>.

⁴⁷ *Id.* ¶ 156.

⁴⁸ *Lehideux & Isorni v. France*, App. No. 55/1997/839/1045, ¶ 57 (Eur. Ct. H.R. Sept. 23, 1999), <https://perma.cc/ZTB8-YMMC>.

⁴⁹ International Covenant on Civil and Political Rights, art. 19(2), *opened for signature* Dec. 19, 1966, T.I.A.S. No. 92–908, 999 U.N.T.S. 171 [hereinafter ICCPR].

⁵⁰ *Id.* art. 19(3).

⁵¹ *Id.* art. 20(2).

body of human rights experts.⁵² The UN Special Rapporteur for Freedom of Opinion and Expression (The Special Rapporteur) is an independent expert who contributes reports and statements that clarify the scope of Articles 19 and 20.⁵³

A 2012 report on hate speech by the Special Rapporteur clarifies that for expression to reach the threshold of Article 20(2), it must satisfy a cumulative test: The expression must amount to advocacy, that advocacy must constitute incitement, and the incitement must be directed toward one of the specified harmful outcomes.⁵⁴ Mere advocacy of hatred, without the additional elements of incitement and likely resultant harm, does not fall within the prohibition.⁵⁵

The interplay of Articles 19(3) and 20(2) thus provides a comparatively clear and operational benchmark for assessing speech restrictions, particularly when criminal sanctions are in play.⁵⁶ Article 20(2) imposes a duty to prohibit a narrow category of intentional advocacy of hatred that reaches the threshold of incitement; but the HRC stresses that implementing that duty remains constrained by Article 19(3)'s requirements of legality, legitimate aim, necessity, and proportionality. This prevents Article 20 from being treated as a free-standing license for broad or indeterminate criminalization of offensive or controversial speech.

In 2012, a group of experts under the auspices of the UN's Office for the High Commissioner for Human Rights adopted the Rabat Plan of Action (RPA).⁵⁷ The RPA is a key framework for interpreting how Article 19 ICCPR's protection of free expression relates to Article 20(2)'s obligation to prohibit any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility, or

⁵² See *Human Rights Committee*, UNITED NATIONS, <https://perma.cc/MBP3-JZ26>.

⁵³ See *Special Rapporteur on Freedom of Opinion and Expression*, UNITED NATIONS, <https://perma.cc/67DA-R63E>.

⁵⁴ Frank La Rue (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 43, U.N. Doc. A/67/357 (Sep. 7, 2012).

⁵⁵ *Id.* ¶ 46.

⁵⁶ ICCPR, *supra* note 49, arts. 19(3), 20(2).

⁵⁷ U.N. High Comm'r for Human Rights, *Rabat Plan of Action on the Prohibition of Advocacy of National, Racial, or Religious Hatred That Constitutes Incitement to Discrimination, Hostility or Violence*, at 6, U.N. Doc. A/HRC/22/17/Add.4 (Jan. 11, 2013) [hereinafter *Rabat Plan of Action*].

violence. It sets “a high threshold . . . for defining restrictions on freedom of expression,” prohibiting overly broad categories of offensive or hateful speech.⁵⁸ It operationalizes this through a strict six-part “threshold test” (context, speaker, intent, content/form, extent, and likelihood/imminence of harm) to determine when restriction or prosecution can be justified.⁵⁹

The RPA six-part test has been referenced by the ECtHR and endorsed by the Committee on the Elimination of Racial Discrimination, the treaty body under the International Convention for the Elimination of All Forms of Racial Discrimination, which includes an obligation to criminalize racist hate speech.⁶⁰ It is also frequently invoked by Meta’s Oversight Board, which makes binding content moderation decisions for Meta platforms.⁶¹ The RPA is widely regarded as a norm-setting benchmark for human-rights-compliant hate speech regulation, anchored in the strict application of legality, necessity, and proportionality.

The HRC places the burden on States to justify restrictions.⁶² Criminal sanctions represent the most intrusive form of interference, carrying a significant risk of chilling lawful expression, and must therefore satisfy the strictest of reviews. The RPA clarifies that criminal law should be a last resort, limited to intentional, public, and imminent incitement to discrimination, hostility, or violence.⁶³ International human rights standards also recognize a spectrum of non-coercive responses, political leadership, education, dialogue, and support for affected communities along-

⁵⁸ *Id.* ¶ 29.

⁵⁹ *Id.*

⁶⁰ See, e.g., Comm. on the Elimination of Racial Discrimination, *General Recommendation No. 35*, U.N. Doc. CERD/C/GC/35 (Sep. 26, 2013). Additionally, there are several ECtHR cases. See, e.g., *Sanchez v. France*, App. No. 45581/15, ¶ 56 (Eur. Ct. H.R. May 15, 2023), <https://perma.cc/C8NB-Y53A>; *Mariya Alekhina v. Russia*, App. No. 38004/12, ¶ 110 (Eur. Ct. H.R. July 17, 2018), <https://perma.cc/98XB-HTLR>.

⁶¹ See, e.g., *Posts Sharing Speeches in Syrian Conflict*, OVERSIGHT BD. (Oct. 2, 2025), <https://perma.cc/Q2DJ-Q6SJ>.

⁶² Hum. Rts. Com., *Views, Kungurov v. Uzbekistan*, U.N. Doc. CCPR/C/102/D/1478/2006 (Sep. 15, 2011).

⁶³ *Rabat Plan of Action*, *supra* note 57, ¶ 34.

side narrowly tailored criminalization of incitement to imminent violence, as reflected in Human Rights Council Resolution 16/18.⁶⁴ The Council of Europe has also developed initiatives around alternative methods to criminalization or regulation with Recommendation CM/Rec (2022) 16, emphasizing education and training.⁶⁵ In *Mohamed Rabbae, A.B.S. & N.A. v. Netherlands* (2011), the HRC offered its most detailed interpretation of Article 20(2) ICCPR. The complainants argued that Dutch politician Geert Wilders' acquittal for hate speech breached the Netherlands' duty to prohibit incitement to hatred. The HRC emphasized that Article 20(2) is "crafted narrowly" to safeguard free expression, including speech that is "deeply offensive" or disrespectful of religion unless the provision's strict incitement threshold is met. The HRC found no violation of Article 20(2).⁶⁶ The Special Rapporteur has likewise stated that,

[I]t is important to emphasize that expression that may be offensive or characterized by prejudice and that may raise serious concerns of intolerance may often not meet a threshold of severity to merit any kind of restriction. There is a range of expression of hatred, ugly as it is, that does not involve incitement or direct threat, such as declarations of prejudice against protected groups.⁶⁷

This marks a stark contrast to the ECtHR, which accepts that mere "insulting" and "offensive" statements can qualify and be prohibited as hate speech.⁶⁸ Unlike the

⁶⁴ Human Rights Council Res. 16/18, U.N. Doc. A/HRC/RES/16/18 (Apr. 12, 2011); *see also* KARI TELLE, UN RESOLUTION 16/18 AND THE ISTANBUL PROCESS: WHAT HAS BEEN ACHIEVED? A CHARTING OF BLASPHEMY TRENDS IN PAKISTAN AND INDONESIA (2022), <https://perma.cc/PGV7-77PG>.

⁶⁵ *Recommendation CM/Rec(2022)16 of the Committee of Ministers to Member States on Combating Hate Speech*, COMM. OF MINISTERS (May 5, 2022), <https://perma.cc/S3FB-97G9>.

⁶⁶ *Rabbae v. Netherlands*, Communication No. 2124/2011, Views, ¶ 10.4, Human Rights Committee, U.N. Doc. CCPR/C/117/D/2124/2011 (July 14, 2016), <https://perma.cc/4ZW9-T5FS>.

⁶⁷ David Kaye (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression), *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶ 10, U.N. Doc. A/74/486 (Oct. 9, 2019).

⁶⁸ *See* *Vejdeland v. Sweden*, App. No. 1813/07, ¶ 55 (Eur. Ct. H.R. Feb. 9, 2012), <https://perma.cc/BU7A-UKCX>; *Handyside v. United Kingdom*, 24 Eur. Ct. H.R. (ser. A) 49 (1976).

ECtHR's case law, General Comment 34 also makes clear that prohibitions of blasphemy and religious insult are incompatible with Article 19 unless they fall within the specific framework of Article 20(2).⁶⁹

As demonstrated below, the Commission's proposal departs from international standards in each of the three core areas: legality, necessity, and proportionality.

Another important element of this case is the HRC's observation that "hate speech and similar laws ironically are often employed to suppress the very minorities they purportedly are designed to protect."⁷⁰ In his report on online hate speech, Special Rapporteur Kaye recognized this risk and warned that,

a person advocating a minority or even offensive interpretation of a religious tenet or historical event, or a person sharing examples of hatred and incitement to report on or raise awareness of the issue, is not to be silenced under article 20 (or any other provision of human rights law).⁷¹

1. Legality: The lack of clear and foreseeable conceptual framework

One of the central concerns regarding legality arising from the Commission's proposal and accompanying documents lies in the absence of a sufficiently clear and non-elusive *conceptualization* of hate speech capable of structuring subsequent legislative action. While the term hate speech appears repeatedly across Commission documents, the Parliament's Resolution and accompanying political discourse are characterized by extremely broad, vague, and ill-defined categories, making them vulnerable to subjective interpretation, as described above in Part I(A)(1).

An online only instrument would compound these legality concerns by creating parallel offences defined by medium. If the same statement is lawful in a lecture hall or in a printed letter to the editor but criminal once posted online, citizens may not reliably be able to foresee the legal consequences as speech moves across different channels of communication. Such a two-tier structure is also likely to generate pressure to align general criminal codes with the online definition, effectively expanding criminal liability, despite the lack of unanimity required to change Article 83(1) of the TFEU.

⁶⁹ Hum. Rts. Comm., *General Comment No. 34*, ¶ 48, U.N. Doc. CCPR/C/GC/34 (Sep. 12, 2011).

⁷⁰ Hum. Rts. Comm., *Rabbae v. Netherlands*, *supra* note 66, ¶ 8.

⁷¹ Kaye, *Report of the Special Rapporteur*, *supra* note 67, ¶ 10.

While a precise definition of criminal offenses is not required at the proposal stage, the principle of legality still requires sufficiently clear descriptions of the conduct proposed for future criminalization. This ensures that the laws can be properly assessed under Article 83(1) TFEU, which requires criminalization of offenses on an EU-level and defines the boundaries of any potential interference with freedom of expression.

The Commission's own public rhetoric underscores this ambiguity. Justice Commissioner Michael McGrath has described hate speech interchangeably as "incitement," "harmful speech," "toxic content," and as part of a broader category of "threats to democracy," which encompasses bots, trolls, disinformation, and foreign information manipulation.⁷² Such formulations blur the line between expression, conduct, and manipulation, making it impossible to determine whether the proposed offense would be limited to intentional incitement or extend to merely offensive, negligent, or controversial speech. Where the conceptual boundaries of the envisaged offense remain unclear, this requirement of legality is undermined. This concern is particularly acute in the field of hate speech, where the underlying concept is inherently contested. In fact, the profound dangers of this conceptual elasticity are vividly illustrated in the 2026 judgment of the CJEU in *Commission v. Hungary*.⁷³ While centered on Hungary's "Child Protection Act,"⁷⁴ the judgment provides critical guidance on how online speech governance must comply with the Charter of Fundamental Rights. The CJEU stated that the law's provisions "limit the freedom of expression and information not only with regard to minors, but also with regard to members of the general public wishing to receive such content, as well as with regard to service providers disseminating that content in the form of advertising, communications of public interest or messages promoting awareness broadcast in the public sphere."⁷⁵ When transposed to the EU's initiative to harmo-

⁷² Michael McGrath, Eur. Comm'r, *Future Proofing EU Democracy at the International Institute of International and European Affairs* (May 18, 2025), <https://perma.cc/H48U-WRA2>; Michael McGrath, Eur. Comm'r, *Shielding Europe's Democracy from Hybrid Threats at the Copenhagen Democracy Summit* (May 12, 2025), <https://perma.cc/D4JG-EVWG>.

⁷³ Case C-769/22, *Comm'n v. Hungary*, ECLI:EU:C:2026:326 (Apr. 21, 2026).

⁷⁴ 1997. évi XXXI. törvény a gyermekek védelméről és a gyámügyi igazgatásról [Act XXXI on the Protection of Children and on Guardianship Administration] § 6/A (Hung.).

⁷⁵ *Hungary*, Case C-769/22, ¶ 468.

nize online hate offenses, *Commission v. Hungary* demonstrates that when statutory definitions of speech are left broad and subjective, the resulting framework fails the strict strictures of legality and proportionality mandated by Article 11. Ultimately, this case suggests that the EU proposal is fundamentally incompatible with EU law; the Commission cannot deploy an open-ended and elastic regulatory architecture that mirrors the exact disregard for Article 11 protection which was highlighted in the Hungarian case.

Ultimately, criminal law cannot be structured around an indeterminate and vague conception of hate speech that invites arbitrary definitions and enforcement without violating freedom of expression guarantees under IHRL. As the HRC stated in General Comment 34 and in *Koktish v. Belarus* (2014), “for the purposes of article 19, paragraph 3, a norm must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly.”⁷⁶ Vague or open-ended criminal prohibitions fail this test, particularly in the speech context, where the chilling effect of uncertainty is itself a form of restriction.

Without a sufficiently defined offense, the Council cannot meaningfully assess whether the conduct in question satisfies the criteria under Article 83(1).

2. Necessity: An unproven need for EU-level criminalization

Necessity requires the State to demonstrate that restricting expression is indispensable to achieving a legitimate objective, including by showing that less intrusive measures are unavailable or insufficient, and that the selected measure is the least intrusive means capable of achieving that objective.⁷⁷ General Comment 31 of the HRC, on the nature of the general legal obligation imposed on States Parties to the Covenant, specifies that “where such restrictions are made, States must demonstrate their necessity and only take such measures as are proportionate to the legitimate aim pursued.”⁷⁸

⁷⁶ Hum. Rts. Comm., *General Comment No. 34*, *supra* note 69; *Koktish v. Belarus*, Communication No. 1985/2010, Views, ¶ 8.5, Human Rights Committee, U.N. Doc. CCPR/C/111/D/1985/2010 (July 24, 2014), <https://perma.cc/RU8N-TK3B>.

⁷⁷ *Rabat Plan of Action*, *supra* note 57; *Lehideux & Isorni v. France*, App. No. 55/1997/839/1045, ¶ 4 (Eur. Ct. H.R. Sept. 23, 1999), <https://perma.cc/ZTB8-YMMC>.

⁷⁸ Hum. Rts. Comm., *General Comment No. 31*, ¶ 6, U.N. Doc. CCPR/C/21/Rev.1/Add. 13 (May 26, 2004).

Kaye and Aswad note that for the least intrusive means test to be met, the State must determine whether non-censorial measures, such as enforcing anti-discrimination laws or engaging in counterspeech by public officials, are available and sufficient to address the identified harm. Where such measures are effective, restricting speech cannot be justified. If such measures are inadequate, the state must evaluate and rank available speech-intrusive options and select the measure that imposes the least burden on freedom of expression, with particular attention to the rights of both speakers and audiences. The state must demonstrate, through transparent and evidence-based reasoning, that the chosen restriction is capable of achieving the stated objective; measures that are ineffective or counterproductive cannot satisfy the requirement of necessity.⁷⁹

The RPA also emphasizes that criminal sanctions should be “*last resort measures*” reserved for cases of intentional and imminent incitement.⁸⁰ The Commission has not met this burden. Its proposal neither assesses the question of necessity nor explains why civil remedies, administrative measures, counterspeech, or educational interventions would be insufficient to address the identified harms.

As described above, Europe already operates within a dense regulatory environment. In particular, the Framework Decision obliges Member States to criminalize public incitement to violence or hatred on racial, national, or religious grounds.⁸¹ The Commission has initiated several infringement proceedings to compel broader national transposition, thereby reinforcing the ongoing expansion of criminalization of speech.⁸² At the same time, the DSA now imposes harmonized obligations for the removal of illegal content and systemic risk mitigation.⁸³

⁷⁹ Aswad & Kaye, *Convergence & Conflict*, *supra* note 39, at 175.

⁸⁰ *Rabat Plan of Action*, *supra* note 57, ¶ 34 (emphasis added).

⁸¹ Council Framework Decision 2008/913/JHA, *supra* note 14, art. 1(1)(a).

⁸² European Commission Infringement Decisions Package INF/25/982, May Infringements Package: Key Decisions (May 6, 2025), <https://perma.cc/43UK-5RFS>; Thomas Wahl, *Commission Launches/Continues Infringements Proceedings in Several JHA Matters*, EUCRIM (Mar. 20, 2021), <https://perma.cc/QFQ4-KEDU>.

⁸³ Digital Services Act, *supra* note 14, arts. 1(1)–2, 9, 16.

The Commission asserts that divergent national approaches “may send mixed messages to the public that such acts are not being taken seriously and can be perpetrated with impunity.”⁸⁴ Yet it provides no evidence that perpetrators escape accountability because of these divergences. On the contrary, prosecutions and convictions for hate speech have increased across the Union.⁸⁵

The political rationales advanced in support of the proposal have also shifted. Initially framed as a response to online hate during the COVID-19 pandemic, the initiative was later revived amid, amongst others, post-October 7 tensions.⁸⁶ Such shifting justifications underscore that the proposal is driven by political urgency rather than a demonstrated legislative necessity. Crucially, the proposal is put forward by the European Commission, which functions as the executive and administrative authority of the EU. As well as the body putting forth the proposal, the Commission would also be monitoring its implementation and could, to this end, institute infringement proceedings (legal action) against Member States which are found to not be complying or partially complying. As an inherently political body, its reliance on fluctuating, crisis-driven pretexts underscores that the proposal is propelled by immediate political urgency rather than a demonstrated, durable legislative necessity.

The HRC has stressed that restrictions on expression must be “directly related to the specific need on which they are predicated.”⁸⁷ In *Velichkin v. Belarus* (2005), the HRC held that “the right to freedom of expression is of paramount importance in any democratic society, and any restrictions on the exercise of this right must meet a strict test of justification.”⁸⁸ The notion of a restriction responding to a “pressing social need” has also been established in ECtHR jurisprudence.⁸⁹ As such, reliance on fluctuating political anxieties does not satisfy the requirements of what

⁸⁴ *A More Inclusive and Protective Europe*, *supra* note 4, ¶ 3.3.

⁸⁵ Natasha Dangoor et al., *Europe’s Crackdown on Speech Goes Far and Wide*, WALL ST. J. (July 7, 2025, at 21:00 ET).

⁸⁶ *No Place for Hate: A Europe United Against Hatred, A More Inclusive and Protective Europe*, *supra* note 7, nn.8 & 37.

⁸⁷ Hum. Rts. Comm., General Comment No. 34, *supra* note 69, ¶ 22.

⁸⁸ *Velichkin v. Belarus*, Communication No. 1022/2001, Views, ¶ 7.3, Human Rights Committee, U.N. Doc. CCPR/C/85/D/1022/2001 (Oct. 20, 2005), <https://perma.cc/9PK2-KMFC>.

⁸⁹ *Sunday Times v. United Kingdom*, App. No. 6538/74, ¶ 59 (Eur. Ct. H.R. Apr. 26, 1979), <https://perma.cc/7M8J-5W26>.

constitutes a pressing social need. This issue becomes particularly relevant if the proposal's justifications of increased hate crimes against specific minorities (Jews and Muslims, for instance) are used to justify the inclusion of broader or open-ended protected categories.

3. Proportionality: Criminal law as an excessive response

Even where a restriction is deemed necessary, proportionality requires the state to demonstrate that the benefit to the legitimate objective outweighs the harm caused to freedom of expression and related rights, and that the interference does not unduly intrude upon the rights of speakers or audiences.

General Comment 34 notes that the principle of proportionality must be “respected not only in the law that frames the restrictions but also by the administrative and judicial authorities in applying the law.”⁹⁰ In decisions, the HRC has underlined that any restriction must be “appropriate to achieve their protective function” and that the least intrusive measures opted for “must be proportionate to the interests to be protected.”⁹¹

Kaye and Aswad note that proportionality operates as a distinct and additional safeguard alongside the necessity analysis.⁹² It requires governments to demonstrate that speech restrictions pursue a specific public interest objective and do not unduly intrude upon the rights of affected individuals, including third parties. This assessment further demands that any interference with freedom of expression be limited and justified in light of the interest served.

While the proportionality inquiry may overlap with the least-intrusive-means analysis, the two are not interchangeable, and governments must satisfy both. Proportionality thus functions as a final balancing exercise, requiring states to show that the benefit of the restriction to the legitimate objective outweighs the harm imposed on freedom of expression and related rights, even where the least intrusive means requirement has been met.⁹³

The Commission's proposal fails to meet the proportionality requirement because its abstract and open-ended framing produces an excessive intrusion into

⁹⁰ Hum. Rts. Comm., General Comment No. 34, *supra* note 69, ¶ 34.

⁹¹ *Id.*

⁹² Aswad & Kaye, *Convergence & Conflict*, *supra* note 39, at 176.

⁹³ *Id.*

protected expression. The absence of core limiting criteria such as intent, imminence, likelihood of harm, and contextual assessment significantly broadens the scope of criminal liability, thereby imposing a disproportionate burden on freedom of expression. As a result, the proposal risks capturing political debate, journalistic inquiry, satire, and other forms of protected speech, rather than confining criminal sanctions to a narrowly defined category of expression whose restriction can plausibly be justified in light of the competing interests at stake.

The disproportionality concern is magnified by the proposal's downstream interaction with the DSA: Expanding criminal illegality expands platform compliance duties and heightens incentives for risk-averse removal of lawful but controversial speech, extending the effects of the restriction beyond the original speaker to third parties such as journalists but also minority voices such as opposition groups through interference with their right to receive and access information. This happens not only through the removal of specific content, but also through the risk of over-removal, reduced visibility, and the deterrence of publication in the first place, thereby interfering with their right to receive and access information (see Part III). Such extensive and indirect interference with third-party rights weighs heavily against proportionality, as set out in *Commission v. Hungary* which did not accept the justification of protecting children for purposes of preventing them from accessing information pertaining to LGBTQ+ rights.⁹⁴

Proportionality also requires that expected benefits outweigh the predictable costs. Where the effectiveness of broader criminalization is empirically uncertain, and the risks to democratic discourse and minority expression are substantial (see Part III D), criminal sanctions are difficult to defend as proportionate, even if necessity were otherwise established.

4. Summary

Assessed against the requirements of legality, necessity, and proportionality, the Commission's effort to harmonize criminal prohibitions on hate speech, whether through an "EU crime" under Article 83(1) TFEU or an emerging online-only instrument, raises serious concerns. It relies on an imprecise and potentially open-ended concept of hate speech, failing to define constituent elements

⁹⁴ Case C-769/22, *Comm'n v. Hungary*, ECLI:EU:C:2026:326 (Apr. 21, 2026).

with the clarity required by legality. It does not demonstrate that EU-level criminalization or harmonization is necessary given existing national laws and the regulatory framework established by the DSA. Finally, resorting to criminal law without robust safeguards risks disproportionate interference with lawful and democratically valuable expression, with chilling effects amplified by downstream digital enforcement.

And the proposal cannot be justified by appealing to either IHRL or European consensus. It exceeds the limits imposed by the UN system and cannot plausibly be defended as a natural extension of existing ECtHR jurisprudence:

(a) Under the UN framework, an EU-level regime harmonizing online hate offences would fail unless it confined criminal liability to the narrow category that Article 20(2) ICCPR obliges States to prohibit: intentional advocacy of national, racial, or religious hatred constituting incitement to discrimination, hostility, or violence. It would also need to incorporate the core safeguards required under IHRL, intent, context, and the likelihood/imminence of harm—and treat criminal law as a measure of last resort, as articulated by the Human Rights Committee, the Special Rapporteur, and the Rabat Plan of Action.

(b) The proposal fails even when assessed against the ECtHR's comparatively lower standards. Despite its deferential approach to hate speech regulation, the ECtHR has consistently emphasized the need for legal precision, foreseeability, and restraint in the use of criminal sanctions, particularly given their chilling effect on lawful expression. The open-ended conceptualization of hate speech envisaged at the EU level, combined with the absence of core limiting principles and the automatic downstream effects triggered under the DSA, renders the initiative incompatible even with Strasbourg's permissive approach to the tripartite test under Article 10. The problems with legal precision and foreseeability would arguably be compounded by the harmonization of an online-only approach, creating a two-tier offense, depending on medium, over which speakers and users often have limited control.

III. EUROPE'S EXISTING HATE SPEECH FRAMEWORK: SCOPE, ENFORCEMENT, AND ESCALATORY DYNAMICS

The proposal to criminalize hate speech at the EU level cannot be assessed in isolation. It operates within an already expansive and intensifying legal and enforcement landscape in which European States exercise some of the most extensive powers over expression among open democracies. Far from revealing a regulatory

vacuum, the existing European framework demonstrates a persistent pattern of scope creep: criminal and other speech-intrusive measures initially justified by reference to extreme or violent expression have gradually expanded to encompass political dissent, satire, artistic expression, and minority advocacy.

Understanding this existing framework is essential to evaluating the proportionality and necessity of further EU-level criminalization. Viewed against Europe's post-war resistance to the criminalization of expression, it reveals a structural dynamic whereby broadly conceptualized hate speech frameworks, once embedded in law, erode enforcement thresholds and normalize increasingly speech restrictive measures.

A. *Historical Context: Europe's Post-War Resistance to the Criminalization of Expression*

The Commission's proposal is at odds with the prevailing sentiment underlying democratic Europe's own immediate post-war human rights tradition. During the drafting of the Universal Declaration of Human Rights (UDHR), the ECHR, and later the ICCPR, European democracies consistently resisted proposals to mandate broad prohibitions on expression in human rights instruments. These initiatives were most prominently advanced by the Soviet Union and its allies, who sought to embed expansive speech restrictions, often framed as anti-fascism or pro-social harmony, into international law.⁹⁵

In 1949, a Soviet representative asserted that,

[A]ll honest people throughout the world . . . were against any propaganda of racial and national superiority and hatred. There was more than obvious need to prohibit and suppress any such propaganda. Anyone attempting to propagate such views in the USSR would soon be suitably dealt with, and the same ought to be the case in all countries throughout the world.⁹⁶

Yet communist states including Bulgaria, Romania, and Hungary would rely on peace treaty provisions requiring the prohibition of fascist organizations and propaganda to effectively ban and suppress all political opposition, for which these countries were condemned by the UN General Assembly in 1950.

⁹⁵ Jacob Mchangama, *The Sordid Origin of Hate-Speech Laws*, HOOVER INST. (Dec. 1, 2011), <https://perma.cc/5Y5X-GMRE>.

⁹⁶ Comm. on Hum. Rts., *Summary Record of the Fifth Session*, at 4–5, U.N. Doc. E/CN.4/SR 123 (June 14, 1949).

When the provision that would later become Article 20(2) ICCPR was put to a vote, not a single member of the Council of Europe voted in favor.⁹⁷ The Nordic States were particularly outspoken. Speaking on their behalf, the Norwegian representative warned that freedom of expression was already “*endangered* by the wording of Article 19” and cautioned that Article 20 risked further erosion, noting that ‘advocacy of national hatred that constitutes incitement to hostility’ was so easy to misconstrue that those whom the provision was supposedly designed to protect might very well find themselves its victims.⁹⁸

Although Article 20(2) was ultimately adopted, several European States, including Belgium, Denmark, Finland, Luxembourg, and Malta entered reservations, explicitly declining to introduce domestic legislation to give it effect.⁹⁹

In relation to the ECHR, the UDHR and the draft version of the ICCPR formed the initial basis for the Convention’s drafting, with the *travaux préparatoires* pointing to “certain affinities” between Article 19 of the draft ICCPR and Article 10 of the ECHR.¹⁰⁰ During the drafting of Article 10 of the ECHR, some countries proposed measures to restrict the scope of freedom of expression.

For example, Turkey proposed that the right to freedom of expression should only be permissible to the extent it is reconcilable with the principles set forth in the preamble to the statute of the Council of Europe.¹⁰¹ The basis of its proposal was a Turkish law prohibiting the dissemination of extremist propaganda. The amendment was rejected because it was “too extensive” a restriction.¹⁰²

Greece likewise proposed introducing “special measures” to address individuals who, under the guise of expressing opinions, engaged in violence or sought to

⁹⁷ Mchangama, *supra* note 95.

⁹⁸ U.N. GAOR, 3d Comm., 21st Sess., 1084th mtg., ¶ 4, ¶ 5, U.N. Doc. A/C.3/SR.1084 (1961) (emphasis added).

⁹⁹ *Reservations and Declarations Made by State Parties of the International Covenant on Civil and Political Rights (ICCPR)*, CENTRE FOR CIV. & POL. RTS. (Mar. 31, 2017), <https://perma.cc/2NMG-7ZJP>.

¹⁰⁰ Council of Europe European Commission of Human Rights, *Preparatory Work on Article 10 of the European Convention on Human Rights*, DH (56) 15 (Aug. 17, 1956), <https://perma.cc/Z9AW-SQWG>.

¹⁰¹ *Id.* at 9.

¹⁰² *Id.*

incite it.¹⁰³ The explicit prohibition of incitement to violence envisaged in the Greek proposal was deliberately rejected, with the matter left to the discretion of individual States as to whether such forms of expression should be criminalized under domestic law.¹⁰⁴

This history is often forgotten in contemporary European policymaking, where broad criminal prohibitions on hate speech are presented as an inherent feature of democratic resilience. The current proposal thus represents a clear departure from the restraint that many European democracies once regarded as essential to safeguarding pluralism, dissent, and democratic contestation.

B. National Criminal Law: From Incitement to Open-Ended Prohibitions

Across the EU, hate speech provisions are typically embedded in criminal laws and frequently extend well beyond the narrow incitement paradigm reflected in Article 20(2) of the ICCPR, both in scope and enforcement. In many Member States, criminal liability applies not only to advocacy of violence or discrimination but also to insult, ridicule, symbolic expression, or speech deemed to violate the “dignity” of protected groups, even in the absence of intent to incite harm. In some countries, hate speech also includes the denial of certain historical events, such as the Holocaust, and overlaps with the criminalization of blasphemy or religious insult.

In Germany, the offense of incitement to hatred under section 130 of the Criminal Code is punishable by up to five years’ imprisonment or a fine.¹⁰⁵ France’s Law of 29 July 1881 provides for as much as one year’s imprisonment and a fine of €45,000 for the incitement to discrimination, hatred, or violence against a person or a group of people on the grounds of their origins, ethnicity, nationality, race, or religion.¹⁰⁶ In Denmark, threatening, ridiculing, or degrading groups based on race, color, origin, religion, sexual orientation, or sexual identity is punishable with up to 2 years’ imprisonment.¹⁰⁷

¹⁰³ Stephanie Farrior, *Molding the Matrix: The Historical and Theoretical Foundations of International Law Concerning Hate Speech*, 14 BERKELEY J. INT’L L. 1, 65 (1996).

¹⁰⁴ *Id.*

¹⁰⁵ Strafgesetzbuch [StGB] [Penal Code] § 130 (Ger.).

¹⁰⁶ Loi du 29 Juillet 1881 sur la Liberté de la Presse [Law on the Freedom of the Press of July 29, 1881].

¹⁰⁷ Straffeloven [Criminal Code] ch. 27, § 266b (Den.).

In the Czech Republic, publicly instigating hatred towards “any nation, race, ethnic group, religion, class or another group of people” carries as much as a 2-year prison sentence.¹⁰⁸ Hungary has created the offense of inciting hatred against “the Hungarian nation,” extending hate speech protections to abstract or majoritarian entities.¹⁰⁹ Article 510 of the Spanish Criminal Code punishes the direct or indirect incitement of hatred, hostility, discrimination, or violence against protected groups with up to four years’ imprisonment and a fine, where such groups are defined by characteristics such as ideology, religion, ethnicity, nationality, gender, or illness.¹¹⁰

Germany’s Criminal Code is emblematic. Section 130 criminalizes incitement to hatred but also penalizes speech that “violates the human dignity of others” by insulting or maliciously maligning segments of the population.¹¹¹ The law does not require proof of intent to incite violence, discrimination, or hostility. This provision operates alongside additional speech-related offenses, including bans on certain political symbols (Sections 86 and 86a), defamation and insult (Sections 185–189), and the criminalization of offending religious beliefs (Section 166).¹¹² In 2021, the German Criminal Code was amended to criminalize “hate-mongering insults” that fall below the threshold of incitement to hatred.¹¹³ These provisions create a layered system in which expressive acts may be reclassified as criminal based on shifting assessments of meaning, symbolism, or perceived social harm. In 2017, Germany adopted the Network Enforcement Act (NetzDG) (now superseded by the DSA), which obliged online networks to remove “manifestly unlawful” content within 24 hours or risk fines of up to five million euros.¹¹⁴

¹⁰⁸ *Trézi Zákoník* [Criminal Code] *zyklon č. 40/2009 Sb.*, § 356 (Czech.).

¹⁰⁹ 2012. évi C. törvény a Büntető Törvénykönyvről [Act C of 2012 on the Criminal Code] § 332(a) (Hung.); *see also* *MAGYARORSZÁG ALAPTÖRVÉNYE* [THE FUNDAMENTAL LAW OF HUNGARY].

¹¹⁰ *Código Penal* [C.P.] [Criminal Code] art. 510 (Spain).

¹¹¹ *Penal Code*, § 130(1) (Ger.).

¹¹² *Id.* §§ 86(1), 86a, 185–189, 166.

¹¹³ *Id.* § 192a.

¹¹⁴ *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* [NetzDG] [Network Enforcement Act], Sept. 1, 2017, *BUNDESGESETZBLATT, TEIL I* [BGBL I] at 3352 (Ger.) [hereinafter *NetzDG*].

France's framework is also expansive. Under the 1881 *Loi Sur La Liberté de la Presse*, the state criminalizes insult, defamation, and "provocation" to discrimination, hatred, or violence.¹¹⁵ French courts have interpreted these provisions broadly, allowing liability to attach to satire, artistic expression, and political commentary.¹¹⁶ This criminal framework is reinforced by administrative powers under the Internal Security Code (expanded in 2022), discussed in the section below. These powers allow for the dissolution of organizations deemed to promote, justify, or incite hatred or terrorism, or to undermine "republican values."¹¹⁷ And the powers significantly expand the range of expression and association subject to criminal and quasi-criminal intervention, reinforcing the dynamics of escalation described above.

C. Enforcement Practices: From Exceptional Sanction to Routine Intrusion

The breadth of European hate speech laws is matched by the intensity and visibility of their enforcement. In several Member States, criminal provisions are routinely operationalized through police raids, device seizures, arrests, and publicized enforcement campaigns targeting online expression. Enforcement practices increasingly treat hate speech regulation not as an exceptional intervention, but as a routine instrument of public order and political governance.

In Germany, federal and state authorities have developed a coordinated enforcement apparatus for policing alleged hate speech online. Police conduct nationwide "days of action" involving early-morning searches of private homes and the confiscation of electronic devices affecting thousands of citizens.¹¹⁸ From 2021 to

¹¹⁵ Loi du 29 Juillet 1881 sur la Liberté de la Presse [Law on the Freedom of the Press of July 29, 1881].

¹¹⁶ *La Cour de Cassation Confirme Deux Condamnations d'Éric Zemmour pour Complicité d'Injure Publique, Provocation à la Haine et pour Diffamation* [The Court of Cassation Upholds Two Convictions of Éric Zemmour for Complicity in Public Insult, Incitement to Hatred, and Defamation], LE CLUB DES JURISTES (Dec. 3, 2025), <https://perma.cc/RK82-P8YC>; *Eon v. France*, App. No. 26118/10 (Eur. Ct. H.R. Mar. 14, 2013), <https://perma.cc/697W-JN2P> (overturning Eon's conviction by finding a violation of Article 10); *M'Bala M'Bala v. France*, App. No. 25239/13 (Eur. Ct. H.R. Oct. 20, 2015), <https://perma.cc/LV3E-3ZG6> (rejecting argument for speech protection under Article 17).

¹¹⁷ CODE DE LA SECURITE INTERIEURE [C. SEC. INT.] [INTERNAL SECURITY CODE] art. L212-1 (Fr.).

¹¹⁸ Alexander Hohlfeld, *Punishment Before Trial? 'Hate Speech' Raids in Germany*, BEDROCK PRINCIPLE (Oct. 24, 2025), <https://perma.cc/CT4T-YRPB>; Oliver Moody, *Raid on Germany's 'Digital Arsonists' Feeds Row over Free Speech*, TIMES (LONDON) (June 25, 2026, at 22:40 BST).

2024, the number of prosecutions for hate speech (broadly interpreted) rose from 2,411 to 10,732.¹¹⁹ As the head of the Federal Criminal Police Office stated, “We are making it clear that anyone who posts hate messages must expect the police to be at the front door afterward.”¹²⁰

These measures have not been confined to cases of explicit incitement or threats. A particularly illustrative example is the case of American playwright C. J. Hopkins, whose Berlin apartment was raided in November 2025. Police officers seized electronic devices and questioned Hopkins and his spouse in connection with a satirical book cover depicting a facemask faintly overlaid with a barely visible swastika.¹²¹ Hopkins had already been convicted the previous year for tweeting the image, deemed to constitute the dissemination of pro-Nazi propaganda, even though the book is a criticism of Germany's COVID policy rather than an endorsement of Nazism.¹²² The renewed investigation was based on the view that the mere act of publishing the satirical cover constituted a fresh criminal offense.¹²³ The case demonstrates how symbolic and artistic expression, even when intended as political critique, can be subsumed under hate speech and extremism provisions once interpretive discretion is expanded.

Similar dynamics are visible in the policing of pro-Palestinian expression. In Germany, numerous demonstrations have been preemptively banned or forcibly dispersed on the grounds of anticipated hate speech or antisemitic expression.¹²⁴ The artist and activist Stella Meris has been arrested multiple times in Germany in

¹¹⁹ Moody, *supra* note 118.

¹²⁰ Adam Satariano & Christopher F. Schuetze, *Where Online Hate Speech Can Bring the Police to Your Door*, N.Y. TIMES (Jan. 21, 2023).

¹²¹ Jacob Mchangama, *Brave New Restrictions*, INT'L POL. & SOC'Y J. (Dec. 9, 2025); *A Visit by the German Thought Police*, CJ HOPKINS (Nov. 26, 2025), <https://perma.cc/YAY8-2H52>.

¹²² See James Kirchick, *What Happens Where Free Speech Is Unprotected*, ATLANTIC (Jan. 3, 2024) (describing earlier conviction).

¹²³ Mchangama, *supra* note 121.

¹²⁴ See, e.g., Tom Wills, *Berlin Shutdown of Pro-Palestine Conference Was Unlawful*, COURT RULES, AL JAZEERA (Nov. 26, 2025); *Germany: Thousands March in Support of Gazans*, DEUTSCHE WELLE (Oct. 21, 2023); *Mehr als 93 Prozent der Anzeigen gegen Bürger stammen von Habeck und Baerbock* [More than 93 Percent of Complaints Against Citizens Stem from Habeck and Baerbock], FOCUS ONLINE (Nov. 24, 2024), <https://perma.cc/6HDX-U4ST>.

connection with pro-Palestinian demonstrations following October 7. Meris reports being detained merely for wearing a Palestinian flag, which police allegedly equated with prohibited extremist symbols, and for holding a sign stating, “from the river to the sea, we demand equality.”¹²⁵

In 2025, the Council of Europe Commissioner for Human Rights warned Germany that restrictions justified by public order or by concerns about hate speech or antisemitism must remain strictly limited, recalling that States have “little scope to impose restrictions on political speech or on debate on matters of public interest, unless the views expressed comprise incitements to violence.”¹²⁶ He further cautioned that the International Holocaust Remembrance Alliance’s working definition of antisemitism has been applied “in ways which lead to the blanket classification of criticism of Israel as antisemitic,” urging vigilance against it being “distorted, instrumentalised or misapplied to stifle freedom of expression and legitimate criticism.”¹²⁷

Since 2020, France has made extensive use of its Internal Security Code, which empowers the executive to dissolve organizations deemed to incite hatred, discrimination, violence, or acts undermining public order.¹²⁸ More than 40 associations and collectives have been dissolved since 2020, including activist groups, often on the basis of cumulative or indirect allegations rather than demonstrable incitement to violence.¹²⁹ In December 2020, the French government ordered the dissolution of the Collective against Islamophobia in France (CCIF), an organization whose stated mission was investigating anti-Muslim discrimination and providing legal

¹²⁵ Jackson, *supra* note 1.

¹²⁶ Letter from Michael O’Flaherty, Comm’r for Human Rights, Council of Eur., to Alexander Dobrindt, Federal Minister of the Interior, Ger. (June 6, 2025), <https://perma.cc/76RU-DLKW>.

¹²⁷ *Id.*

¹²⁸ See Internal Security Code art. L212-1 (Fr.).

¹²⁹ CIVICUS MONITOR, NATIONAL HUMAN RIGHTS INSTITUTION WARNS THAT CIVIC SPACE IN FRANCE IS UNDER THREAT (2025), <https://perma.cc/WP8U-35V6>; see also *Associations: Record de Dissolutions!* [Associations: Record Number of Dissolutions!], INSTITUT ISBL (Mar. 26, 2025), <https://perma.cc/4ELC-5AA4>.

support to victims.¹³⁰ The decree alleged that the CCIF provoked hatred, discrimination, and violence and that its leadership endorsed a radical approach to Islam (Islamist). In 2021, the Conseil d'État upheld the dissolution of the CCIF.¹³¹ The decision was built on a broad interpretation of "provocation" that held the group accountable for its institutional critique of state counter-terrorism policies and for anti-Semitic comments posted by independent third parties on its social media pages.¹³² Crucially, however, the text of the judicial decision itself acknowledged that "the evidence does not indicate that the CCIF or its members engaged in any activities aimed at inciting acts of terrorism."¹³³ Human rights organizations have pointed to the chilling effect of this decision on organizations working on racism and xenophobia and its reliance on vague concepts such as extremism to limit advocacy and anti-discrimination work.¹³⁴

Banishing an entire collective from public life based on undemonstrated claims of violence, or by holding an organization liable for the unmoderated comments of online users, establishes a dangerous precedent. When the state grants itself the power to legally erase an entire organization based on a subjective evaluation of its viewpoints, it effectively abandons the core principle of free expression, which requires a showing of direct causality or an imminent threat of violence. Treating offensive beliefs or an inability to police online forums as a justification to shut down an association fundamentally undermines open debate, transforming the state into an ideological gatekeeper that silences any groups critical of or opposed to mainstream social and political norms.

¹³⁰ 110. *Dissolution du Collectif Contre l'Islamophobie en France (CCIF)* [110. *Dissolution of the Collective Against Islamophobia in France (CCIF)*], OBSERVATOIRE DES LIBERTÉS ASSOCIATIVES [OBSERVATORY OF ASSOCIATIVE LIBERTIES], <https://perma.cc/YB38-UJJ2>; see also *Le CCIF c'Est quoi?* [What Is the CCIF?], ISLAMOPHOBIE [ISLAMOPHOBIA], <https://perma.cc/TGC9-Q86E>.

¹³¹ Conseil d'État, 10ème - 9ème chambres réunies, Sept. 24, 2021, 449215 (Inédit au recueil Lebon), ECLI:FR:CECHR:2021:449215.20210924, <https://perma.cc/FJZ2-G4D2>.

¹³² *Id.* ¶ 9.

¹³³ *Id.* ¶ 12.

¹³⁴ See, e.g., *France: Shutting Down Anti-Racist Organization Risks Freedoms*, AMNESTY INT'L (Nov. 20, 2020), <https://perma.cc/NXT9-WVUR>; Eva Cossé, *French Court Confirms Dissolution of Anti-Discrimination Group*, HUM. RTS. WATCH (Sep. 27, 2021, at 14:32 ET), <https://perma.cc/35RC-HFHG>.

This expansive approach to state intervention is mirrored in the extension of liability for speech uttered by third parties. A clear illustration is the case of politician Julien Sanchez, who was not convicted for his own statements, but for failing promptly to remove hateful comments posted by others on his Facebook page.¹³⁵ At the relevant time, Sanchez was Mayor of Beaucaire and a parliamentary candidate for the Front National (renamed Rassemblement National in 2018). When users posted anti-Muslim comments beneath one of his online posts, and he did not remove them once made aware, French courts held him criminally liable as the “producer” of a publicly accessible online service. He was convicted of incitement to hatred and discrimination, fined and ordered to pay damages, despite arguing that the volume of comments made effective moderation impracticable. This reasoning was endorsed by the ECtHR in 2023, which found that his conviction did not violate Article 10.¹³⁶

Read alongside the dissolution of civil society organizations under the Internal Security Code, *Sanchez* exemplifies a broader domestic trend toward indirect and omission-based liability, lowering the threshold for state interference with political participation and online public debate. These cases are not isolated ones. They show how broad hate speech frameworks, once embedded in criminal law and backed by robust enforcement, can expand beyond incitement to encompass dissenting political views, especially during periods of heightened geopolitical tension.

D. Scope Creep and the Targeting of Minorities and Internal Dissent

One of the most persistent and least acknowledged consequences of expansive speech regimes is that they frequently turn inward, affecting not only majoritarian or extremist speakers but ethnic, religious, racial, and political minorities and critical voices whose expression challenges prevailing political or moral norms.

Across several Member States, individuals belonging to historically marginalized groups have increasingly found their speech reclassified as hateful when it departs from officially endorsed narratives. Criticism of Israeli government policy by Jewish, Muslim, and Palestinian speakers offers a particularly salient example. In Germany, Jewish activists opposing Israel’s military actions in Gaza have repeatedly been accused of engaging in antisemitic hate speech, notwithstanding their

¹³⁵ *Sanchez v. France*, App. No. 45581/15, ¶ 3 (Eur. Ct. H.R. May 15, 2023), <https://perma.cc/C8NB-Y53A>.

¹³⁶ *Id.* ¶ 210.

identity, intent, or explicit rejection of violence. Iris Hefets, for instance, is an Israeli Jewish activist who has been repeatedly detained by police in Berlin during peaceful solo protests.¹³⁷ Her case illustrates how hate speech frameworks can eradicate the distinction between legitimate advocacy and protest and forbidden hostility, once symbolic meaning is assessed through the lens of militant democracy.

Similar dynamics have emerged in cultural and academic contexts. Palestinian and Arab artists have seen exhibitions cancelled, events defunded, and professional opportunities withdrawn following accusations of hate speech or extremism based on expressive content later acknowledged to be non-violent political critique. These measures often occur in the absence of judicial findings of incitement, relying instead on precautionary assessments by authorities or publicly funded institutions seeking to mitigate legal or reputational risk.¹³⁸ The effect is to reverse the protective logic of hate speech law. Instead of shielding minorities from abuse, expansive hate speech regimes enable the regulation of minority speech that challenges dominant political positions.

The problem is exacerbated when policies fail to account for intent, imminence, or harm thresholds. When hate speech is defined primarily through perceived offense or symbolic association, enforcement becomes detached from demonstrable risk. As a result, authorities are incentivized to intervene preemptively, particularly during periods of crisis and tension, when contested speech is treated as a trigger of social unrest rather than as a sign of societal discontent.

These dynamics extend beyond directly targeted speakers and affect the broader climate of public debate. For example, a recent study indicated that only 46% of the German population feels able to express their political views freely.¹³⁹ The study analyzed issues in the framework of recent political and social develop-

¹³⁷ Jackson, *supra* note 1.

¹³⁸ See, e.g., *Cancelling Palestinian Kids' Art Exhibit a Shameful Violation of Free Speech Principles*, NAT'L COAL. AGAINST CENSORSHIP (Sep. 12, 2011), <https://perma.cc/5R5H-2782>; *University Rector Reinstated to Ruling Body After Row*, BBC NEWS (Mar. 18, 2026).

¹³⁹ Thomas Petersen, *Die Deutschen und ihre Freiheit [The Germans and Their Freedom]*, FRANKFURTER ALLGEMEINE ZEITUNG (Oct. 17, 2025), <https://perma.cc/6URQ-E865>.

ments, including the COVID-19 pandemic, migration, and increasing polarization.¹⁴⁰ Strikingly, this erosion of perceived expressive freedom coexists with majority support for restricting or banning certain extreme or unpopular opinions. This tension suggests how expansive and precautionary approaches to speech can internalize self-censorship even in the absence of direct legal sanction. The result is a climate in which individuals refrain from speaking, not because their expression is unlawful, but because they anticipate social, professional, or institutional repercussions.

From a proportionality perspective, this represents a profound failure of tailoring. Criminal and administrative measures are applied not because expression poses a concrete threat, but because it is politically sensitive, controversial, or unsettling. The chilling effect is concentrated precisely where democratic discourse is most vulnerable and acutely endangered: within minority communities and among dissenting voices whose speech contests state policy.

E. EU-Level Pressure and the Escalation of National Criminalization

The above-described national dynamics cannot be separated from the role of EU institutions in shaping the trajectory of hate speech regulation. While criminal law formally remains a Member State competence, EU-level instruments and enforcement mechanisms exert continuous upward pressure toward broader and more punitive national frameworks.

The 2008 Framework Decision already obliges Member States to criminalize a wide range of expression, including public incitement to violence or hatred and the denial or trivialization of genocide and crimes against humanity.¹⁴¹ Although framed as a minimum harmonization measure, the Framework Decision has functioned in practice as an escalatory mechanism, encouraging progressive expansion rather than convergence around a narrow core, favoring the expansion of EU-wide criminal restrictions over a narrow, incitement-based approach to expression.

Since 2020, the European Commission has launched at least 14 infringement proceedings against Member States for allegedly incomplete or incorrect transposition of the Framework Decision.¹⁴² These proceedings often focus on perceived

¹⁴⁰ *Id.*

¹⁴¹ Council Framework Decision 2008/913/JHA, *supra* note 14, art. 1.

¹⁴² *Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Report on the Implementation of the*

gaps in coverage or enforcement of hate speech. The practical effect has been an attempt to compel national legislatures to broaden speech offenses even where domestic democratic processes have expressed hesitation or resistance.

Ireland provides a particularly illustrative example. In 2022, this country introduced the Incitement to Violence or Hatred and Hate Offences Bill, which expanded existing hate speech prohibitions.¹⁴³ The proposal provoked sustained public opposition across civil society, academia, and the political spectrum, with critics warning that it inverted the presumption of innocence and chilled lawful expression.¹⁴⁴ Following extensive debate, the government ultimately removed the hate speech component from the bill, which was adopted in December 2024.¹⁴⁵

Meanwhile, in 2024, the European Commission sent a letter of formal notice to Ireland for only partially and incorrectly transposing the Framework Decision.¹⁴⁶ In other words, the Commission did not believe that Ireland's existing law went far enough. By May 2025, the Commission gave Ireland two months to comply with the letter or face referral to the CJEU.¹⁴⁷

The episode shows how EU-level enforcement can displace national democratic judgments formed through open and sustained debate. This shift weakens democratic accountability by transferring decisions about the limits of permissible expression from national institutions to EU-level bodies that are less directly attuned to affected communities and the sources of social conflict. The result is sustained pressure toward broader definitions of hate speech and intensified enforcement, even in the absence of domestic consensus or demonstrated effectiveness.

EU Anti-Racism Action Plan 2020-2025 and on National Action Plans Against Racism and Discrimination, COM (2024) 419 final (Sep. 25, 2024), <https://perma.cc/B575-J7PA>.

¹⁴³ Criminal Justice (Incitement to Violence or Hatred and Hate Offences) Bill 2022 (No. 105 of 2022) (Ir.) (as initiated Nov. 1, 2022), <https://perma.cc/55RU-CKGX>.

¹⁴⁴ See "Hate Speech" Element Dropped from Censorial Irish Bill, ADF INT'L (Sep. 24, 2024), <https://perma.cc/SD6F-KKJS>; Barbara McCarthy, *Why Is Ireland's Hate Crime Bill Attracting So Much Hate of Its Own?*, AL JAZEERA (June 17, 2024).

¹⁴⁵ Criminal Justice (Hate Offences) Act 2024 (Act No. 41/2024) (Ir.), <https://perma.cc/YM4P-LUHT>.

¹⁴⁶ *European Commission Calls on Ireland, Bulgaria and Estonia to Correctly Transpose EU Law Combating Racism and Xenophobia*, EUROPEAN COMM'N (Oct. 3, 2024), <https://perma.cc/2WN7-UFTT>.

¹⁴⁷ European Commission Infringement Decisions Package INF/25/982, *supra* note 82.

F. Interaction with the Digital Services Act: Platform Enforcement as an Accelerant of Scope Creep

The dynamics described above are substantially amplified by the EU's digital regulatory architecture, particularly the DSA.¹⁴⁸ Under Article 3(h) DSA, "illegal content" is broadly defined as any information that is unlawful under EU or national law.¹⁴⁹ Regardless of whether the Commission pursues full harmonization of hate speech or merely an EU-level definition of online hate offences, that definition would automatically feed into the DSA's "illegal content" category, triggering a cascade of platform obligations. Even if framed as "online-only," it would thereby reshape the practical boundaries of permissible expression at scale.

The Commission has emphasized that "the DSA does not censor content," but, instead, "creates efficient mechanisms for the removal of illegal content."¹⁵⁰ In practice, where illegality is vague or context-dependent, platforms face strong incentives to remove content rather than risk regulatory exposure, especially under audit and enforcement pressure that can result in massive fines.¹⁵¹

Articles 34 and 35 of the DSA are especially significant. They require Very Large Online Platforms (VLOPs) to identify, assess, and mitigate "systemic risks," including the dissemination of illegal content and negative impacts on fundamental rights.¹⁵²

In countries where hate speech is defined broadly, platforms face pressure not just to react to complaints but to prevent harmful content from spreading in the first place. In practice, this means automated filters that routinely misclassify satire,

¹⁴⁸ Digital Services Act, *supra* note 14.

¹⁴⁹ *Id.* art. 3(h).

¹⁵⁰ Henna Virkkunen, Exec. Vice-President, Eur. Comm'n, *Need to Enforce the Digital Services Act to Protect Democracy on Social Media Platforms Including Against Foreign Interference and Biased Algorithms* (Jan. 20, 2025), <https://perma.cc/G5QH-RUDF>.

¹⁵¹ See Barrie Sander, *Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation*, 43 FORDHAM INT'L L.J. 939 (2020); Marcella Atzori, *The Digital Services Act and the Freedom of Expression in the European Union: A Political Perspective* (June 21, 2024) (unpublished manuscript), <https://perma.cc/B5TN-GPFQ>; Natalie Alkiviadou, *Moderating Hate or Moderating Rights? The Paradox of the European Approach to Online Hate Speech and Platform Liability* (CELE Research Paper No. 69, 2025), <https://perma.cc/246C-86UV>.

¹⁵² Digital Services Act, *supra* note 14, arts. 34, 35.

reclaimed slurs, minority dialects, political advocacy, and human rights documentation as hateful. Studies show that content using African American English, LGBTQ+ terminology, or pro-Palestinian political framing is disproportionately flagged or removed. Once criminal law provides a broad, indeterminate definition of hate speech, these errors are no longer internal policy choices but compliance-driven responses to legal risk.¹⁵³

A 2024 Future of Free Speech (FoFS) study found that between 87.5% and 99.7% of user comments removed by Facebook and YouTube in France, Germany, and Sweden were legally permissible under national law.¹⁵⁴ The study indicates systematic over-removal and highlights regulatory-risk incentives as a plausible driver. Over-removal can chill public debate, distort community trust, and displace contentious speech into less visible spaces; and the DSA's audit incentives may reinforce these dynamics.

EU-level criminalization of hate speech or harmonization of online hate offences would not merely add another offence to national criminal codes; it would materially expand the capacity of national authorities and the European Commission to exercise top-down control over the online public sphere through legally compelled private enforcement. In practice, this risks a pervasive restriction on lawful but controversial expression, especially where automated tools are calibrated for risk avoidance rather than contextual judgment.

G. *Synthesis: From National Scope Creep to EU-Level Entrenchment*

Taken together, Parts III.A–III.F show that Europe already operates within an expansive and escalating framework of hate-speech regulation. Offence definitions and enforcement practices have broadened beyond a narrow incitement paradigm. Intrusive interventions have become routine. Minorities and dissenting voices are increasingly drawn into the net. And EU-level pressure has typically pushed rules to become broader in scope and stricter in their penalties.

¹⁵³ Thiago Dias Oliva et al., *Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online*, 25 *SEXUALITY & CULTURE* 700 (2021); *Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook*, HUM. RTS. WATCH (Dec. 21, 2023), <https://perma.cc/5U5E-3C84>; Megan McCluskey, *These TikTok Creators Say They're Still Being Suppressed for Posting Black Lives Matter Content*, TIME (July 22, 2020, at 18:04 ET).

¹⁵⁴ *Preventing "Torrents of Hate" or Stifling Free Expression Online?*, FUTURE OF FREE SPEECH (May 28, 2024), <https://perma.cc/5N2Z-VS5S>.

The DSA then amplifies these dynamics by translating legal expansion into platform enforcement at scale, heightening incentives for over-removal and chilling lawful expression. Against that background, the next Part tests the core empirical premise that further criminalization will reduce intolerance or violence, rather than producing counterproductive dynamics. Moreover, while many Member States already criminalize hate speech, EU-level harmonization would not merely replicate existing rules. It would establish a minimum criminal baseline across all Member States, expand the scope of “illegal content” under the Digital Services Act, and generate legal and political pressure toward broader national alignment.

IV. SPEECH-INTRUSIVE APPROACHES TO HATE SPEECH: EMPIRICAL CHALLENGES AND UNINTENDED EFFECTS

The empirical analysis in this section tests whether the assumptions underpinning necessity and proportionality claims are supported by available evidence.

Hate speech can cause real harm. In certain contexts, such as Rwanda and Myanmar, organized, dehumanizing propaganda has contributed to mass violence and genocide.¹⁵⁵ These cases involved explicit calls for extermination or violence and would meet even the strictest incitement thresholds under the most speech-protective legal systems.¹⁵⁶ They offer no support for criminalizing much broader and indeterminate categories of expression in what Eric Heinze describes as “longstanding, stable and prosperous democracies.”¹⁵⁷

In the EU context, the relevant question is not whether hateful speech exists, but whether expansive criminalization of expression is an effective, necessary, and proportionate response. On this point, the empirical record is far more ambivalent than EU policy discourse suggests.

¹⁵⁵ Alexander Tsesis, *Dignity and Speech: The Regulation of Hate Speech in a Democracy*, 44 WAKE FOREST L. REV. 497 (2009); David Yanagizawa-Drott, *Propaganda and Conflict: Evidence from the Rwandan Genocide*, 129 Q.J. ECON. 1947 (2014); Anamika Modok, *Role of Social Media in Inciting the Genocidal Acts: A Case Study on Myanmar’s Rohingya*, 4 CONTEMPORARY CHALLENGES (2023).

¹⁵⁶ See, e.g., *Brandenburg v. Ohio*, 395 U.S. 444, 447 (1969) (holding that speech can only be forbidden as inciting speech if it is directed to inciting or producing “imminent lawless action” and is likely to do so).

¹⁵⁷ Eric Heinze, *How We Should—and Shouldn’t—Ban Extremist Speech*, at 1 (Queen Mary L., Research Paper No. 462/2025), <https://perma.cc/Y9JP-SA7W>.

A recurring feature of that discourse is the tendency to treat hateful expression not merely as socially harmful, but as causally linked to offline violence. Commissioner McGrath's address marking "No Hate Speech Week" in June 2025 exemplifies this framing. He asserted that,

[T]hroughout Europe and the wider world, we are witnessing an increase in expressions of hatred—both online in our social media feeds, and offline on our streets. Hateful words expressed online can motivate harmful actions offline, and—disproportionally—these target people belonging to minority groups.¹⁵⁸

This rhetoric suggests a direct causal relationship between speech and violence. Yet the Commission does not acknowledge the contested nature of that claim or the substantial empirical debate surrounding it. Nor does it engage with the second-order effects of expansive speech regulation, including the chilling of lawful expression, the reinforcement of grievance and victimhood narratives, or the displacement of contentious discourse into less visible and less accountable spaces.

As this Part shows, the academic literature does not support a general direct causal chain between hateful expression and harmful outcomes, much less the idea that prohibiting hate speech is a cure for hatred and violence. Where links are identified, they are highly context-specific and typically confined to extreme forms of incitement rather than applying to the broad categories of speech targeted by contemporary European regulation.

A. *The Missing Causal Link between Criminalization and Reduced Harm*

A central assumption underpinning European hate speech policy is that restricting expression will reduce prejudice, discrimination, and violence. Yet decades of experience with stringent hate speech laws across Europe have not produced clear or consistent evidence supporting this claim.

National criminal statistics illustrate this pattern. In Germany, officially recorded hate crimes rose from 10,240 in 2020 to 21,773 in 2024, even as speech-intrusive laws, including the NetzDG, were adopted and enforcement escalated.¹⁵⁹ In France, police recorded 2,672 hate crime incidents in 2020; by 2024, that figure had

¹⁵⁸ McGrath, *supra* note 24.

¹⁵⁹ Germany Hate Crime Report, OSCE OFF. FOR DEMOCRATIC INSTS. & HUM. RTS., <https://perma.cc/RT6T-6LKZ>.

increased to 6,096.¹⁶⁰ Following the attacks of October 7, antisemitic incidents rose sharply across multiple Member States, despite the existence of extensive criminal laws targeting antisemitism, including Holocaust denial and incitement provisions.¹⁶¹

In 2024, the European Parliament's research service published a briefing supporting the Commission's proposal, asserting that "[o]ver the last 20 years, there has been a marked increase in hate speech and hate crime in Europe."¹⁶² Recent reports (2021–2025) by the Fundamental Rights Agency, an independent human rights body established by the EU, similarly report demonstrate levels of antisemitic and anti-Muslim hate speech and hate crime.¹⁶³

Yet these assessments do not engage with a central tension at the heart of EU policy. Official reports show increasing levels of hate speech and hate crime even as the EU and its Member States have steadily expanded speech-intrusive laws and enforcement measures, as documented in Part III.

At the same time, the figures implicitly assume a positive correlation between hate speech and hate crime without accounting for multiple intervening variables that could distort this relationship. Increases may instead reflect improved reporting mechanisms, heightened public awareness, or expanded legal definitions rather than an actual rise in harmful conduct. Without disaggregating these factors, en-

¹⁶⁰ France Hate Crime Report, OSCE OFF. FOR DEMOCRATIC INSTS. & HUM. RTS., <https://perma.cc/FN8K-ADP3>.

¹⁶¹ *Antisemitic Acts Have 'Exploded' in France Since 7 October, Interior Minister Says*, GUARDIAN (Nov. 5, 2023, at 20:14 ET); Nette Nöstlinger, *Germany Records Sharp Rise in Antisemitic Incidents*, POLITICO (June 25, 2024, at 17:02 ET); Liv Stroud, *Antisemitic Incidents Surge Across Europe and the World, ADL's J7 Task Force Report Shows*, EURONEWS (Aug. 5, 2025, at 11:24 GMT), <https://perma.cc/MZQ9-PK7K>.

¹⁶² IMMENKAMP ET AL., *supra* note 10, at 1.

¹⁶³ See EUR. UNION AGENCY FOR FUNDAMENTAL RTS., FUNDAMENTAL RIGHTS REPORT—2025 (2025), <https://perma.cc/2HZ8-3MBR>; EUR. UNION AGENCY FOR FUNDAMENTAL RTS., FUNDAMENTAL RIGHTS REPORT—2024 (2024), <https://perma.cc/44Q8-92LU>; EUR. UNION AGENCY FOR FUNDAMENTAL RTS., FUNDAMENTAL RIGHTS REPORT—2023 (2023), <https://perma.cc/N9LH-VSX7>; EUR. UNION AGENCY FOR FUNDAMENTAL RTS., FUNDAMENTAL RIGHTS REPORT—2022 (2022), <https://perma.cc/433U-TG4A>; EUR. UNION AGENCY FOR FUNDAMENTAL RTS., FUNDAMENTAL RIGHTS REPORT—2021 (2021), <https://perma.cc/L4S6-RUWE>; see also *Who We Are*, EUR. UNION AGENCY FOR FUNDAMENTAL RTS., <https://perma.cc/K2EC-FNCV>.

forcement data cannot be treated as reliable evidence of legislative failure or support for the underlying assumption that hate speech laws are an effective remedy for hate crimes.

Recent developments in Australia further illustrate the limits of reflexive reliance on speech repression as a response to social harm. Following antisemitic incidents after October 7, both the Australian federal government and the government of New South Wales adopted stricter hate speech laws.¹⁶⁴ These measures did not prevent subsequent acts of lethal antisemitic violence, including the deadly Bondi Beach attack on December 14, 2025. In the aftermath, authorities nonetheless announced additional restrictions on speech and protest, prompting civil liberties groups to warn of further encroachments on political expression and peaceful assembly.¹⁶⁵ In April 2026, Queensland police arrested more than 20 people over a weekend of protests in Brisbane against new state laws banning the phrases “from the river to the sea” and “globalise the intifada,” which carry a maximum sentence of two years in jail.¹⁶⁶

Recent developments in Canada illustrate both the expansion of hate speech regulation and its contested effectiveness.¹⁶⁷ In 2022, the Criminal Code was amended to specifically criminalize the “wilful[] promot[ion of] antisemitism by condoning, denying or downplaying the Holocaust.”¹⁶⁸ This reform was introduced against the backdrop of longstanding hate speech prohibitions dating back to 1970, originally aimed at countering neo-Nazi activity.¹⁶⁹ However, the persistence, and indeed escalation, of antisemitism raises questions about the effectiveness of such measures. Following the October 7, 2023 attacks by Hamas and the subsequent war in Gaza, Canada witnessed what officials described as an “unprecedented wave” of

¹⁶⁴ Criminal Code Amendment (Hate Crimes) Act 2025 (Cth) (Austl.); Crimes Legislation Amendment (Racial and Religious Hatred) Bill 2025 (NSW) (Austl.).

¹⁶⁵ Saadet Gokce, *Australian Premier Vows Tougher Hate Speech Laws After Bondi Beach Attack*, AA [ANADOLU AGENCY] (Dec. 18, 2025), <https://perma.cc/8E4E-NA86>.

¹⁶⁶ Joe Hinchliffe, *More Than 20 Arrested over Queensland's Controversial 'from the River to the Sea' Laws During Weekend of Protests*, GUARDIAN (Apr. 19, 2026, at 03:11 ET).

¹⁶⁷ Jacob Mchangama & Jeff Kosseff, *Opinion, After Oct. 7, Hate-Speech Laws Looked Like the Answer. Europe Shows Why They Aren't*, GLOBE & MAIL (TORONTO) (Apr. 9, 2026).

¹⁶⁸ Budget Implement Act, 2022, No. 1, S.C. 2022, c. 10, § 332 (Can.).

¹⁶⁹ See Kenneth Grad, *A Gesture of Criminal Law: Jews and the Criminalization of Hate Speech in Canada*, 59 OSGOODE HALL L.J. 375 (2022).

antisemitism, including harassment, threats, and a sharp increase in hate crimes.¹⁷⁰ This surge occurred despite both the longstanding criminal framework and its recent expansion in 2022.

In response, lawmakers have proposed further legislative intervention. Bill C-9, the “Combating Hate Act,” which introduces additional restrictions, including the criminalization of the public display of certain symbols, such as the swastika or Hamas insignia, when used to promote hatred.¹⁷¹ Yet, there is limited evidence that successive expansions of hate speech law have achieved their intended aims. Earlier criminalization did not prevent rising antisemitism in the decades that followed, nor did the 2022 reforms avert the post-October 7 surge.¹⁷² As noted by Kenneth Grad, prosecutions for hate speech “work in theory, but not in practice,” failing to effectively protect targeted communities or fulfil the broader objectives of criminal law.¹⁷³

B. *Repression, Radicalization, and the “Martyr Effect”*

A growing body of interdisciplinary research suggests that speech repression risks producing unintended effects. Rather than marginalizing extremist ideas, criminalization may reinforce narratives of persecution, increase polarization, and contribute to radicalization.

Comparative research on right-wing terrorism in Western Europe identifies extensive repression of radical right actors and opinions as a variable correlated with increased violent activity. An extensive cross-national analysis concludes that heavy-handed repression may contribute to militant escalation by fragmenting movements and pushing actors into smaller, more extreme networks.¹⁷⁴ A 2023

¹⁷⁰ See *Memory, Truth, and Responsibility: Canada’s Role in Confronting Antisemitism and Holocaust Distortion After October 7th*, GOV’T OF CAN., <https://perma.cc/K2KK-X9P4>.

¹⁷¹ Combating Hate Act, Bill C-9, 1st Sess., 45th Parl. (as passed by House of Commons, Mar. 25, 2026) (Can.); see also *Canada Introduces Legislation to Combat Hate Crimes, Intimidation, and Obstruction*, GOV’T OF CAN., <https://perma.cc/3B93-ZJC4>.

¹⁷² HC Deb (Apr. 27, 2022) (151) 4453 (Can.), <https://perma.cc/EB2L-MDDR>.

¹⁷³ Kenneth Grad, *Should Hate Speech Be Criminalized? Lessons from the Canadian Experience in R. v. Zundel and R. v. Keegstra*, 19 FIU L. REV. 661, 707 (2025).

¹⁷⁴ Jacob Aasland Ravndal, *Explaining Right-Wing Terrorism and Violence in Western Europe: Grievances, Opportunities and Polarisation*, 57 EUR. J. POL. RSCH. 845 (2017).

study “suggests that legal repression of controversial or extremist actors can generate counter-mobilization effects, including increased polarization.”¹⁷⁵ Moreover, a 2020 study found that “discussion freedom is unambiguously associated with less terrorism in democracies.”¹⁷⁶

Similar findings emerge from digital research. A large-scale study found that Telegram’s explosive growth among “far-right”¹⁷⁷ actors “coincides in time with the mass bans of the far-right actors on mainstream social media platforms.”¹⁷⁸ The study demonstrates the potential ineffectiveness of deplatforming for curbing the influence of far-right and other extremists.

Beyond organized extremism, repression can fuel broader anti-government sentiment. A 2022 cross-national study drawing on 101 nationally representative samples found a positive association between perceived repression and individuals’ willingness to engage in anti-government violence.¹⁷⁹ Follow-up analyses in countries experiencing political unrest showed that personal experiences of censorship or repression strongly correlated with support for political violence.¹⁸⁰

Criminal prosecutions can also produce a “martyr effect.” Research on hate speech prosecutions in the Netherlands found that the trial of Geert Wilders increased public support for his party.¹⁸¹

A later study showed that media coverage of hate speech cases amplified the visibility of far-right issues and, in some contexts, increased electoral support for anti-immigration parties. When prosecutions concern issues “owned” by extremist

¹⁷⁵ Pramukh Nanjundaswamy Vasist et al., *The Polarizing Impact of Political Disinformation and Hate Speech: A Cross-country Configurational Narrative*, 26 INFO. SYS. FRONTIERS 663 (2024).

¹⁷⁶ Lasse Skjoldager Eskildsen & Christian Bjørnskov, *Does Freedom of Expression Cause Less Terrorism?*, 70 POL. STUD. 1, 1 (2020).

¹⁷⁷ The authors do not define the far-right *per se* but instead use it as an umbrella term that includes different subgroups such as the extreme right and radical right. For more on the definitional framework see work by Cas Mudde. See, e.g., CAS MUDDE, *THE FAR RIGHT TODAY* (2019).

¹⁷⁸ Aleksandra Urman & Stefan Katz, *What They Do in the Shadows: Examining the Far-Right Network on Telegram*, 25 INFO., COMM’N & SOC’Y 904, 904 (2022).

¹⁷⁹ Henrikas Bartusevičius et al., *Political Repression Motivates Anti-Government Violence*, 10 ROYAL SOC’Y OPEN SCI. 221227 (2023).

¹⁸⁰ *Id.*

¹⁸¹ Joost van Spanje & Claes de Vreese, *The Good, the Bad and the Voter: The Impact of Hate Speech Prosecution of a Politician on Electoral Support for His Party*, 21 PARTY POL. 115, 115 (2015).

actors, such as immigration or national identity, legal action may inadvertently reinforce their agenda rather than undermine it.¹⁸² The Alternative for Germany (AfD) doubled its national vote share between 2021 and 2025, becoming the second-largest party in the Bundestag.¹⁸³ Its Thuringia leader, Björn Höcke, twice convicted for using prohibited Nazi-era slogans, cast himself as a persecuted dissident, likening German authorities to inquisitorial censors.¹⁸⁴ His party's subsequent electoral success illustrates how punitive speech regimes can be reframed as evidence of elite suppression, reinforcing rather than weakening extremist narratives.

These findings do not imply that hateful expressions should be ignored. They do suggest, however, that criminal law is an ineffective instrument that may exacerbate the very dynamics it is intended to counter, while undermining freedom of expression.

C. *Free Expression as a Safety Valve and Predictor of Tolerance*

Importantly, empirical literature does not merely cast doubt on whether criminalization of hate speech decreases hatred; it increasingly suggests that it actually increases hatred. A growing body of cross-national research indicates that strong protections for freedom of expression correlate with higher levels of social tolerance and minority protection.

A 2023 large-scale dataset analysis found a negative correlation between freedom of expression and social conflict in democratic systems.¹⁸⁵ A 2025 study drawing on data across more than 100 countries found that societies with stronger free speech and media independence exhibit higher levels of racial and ethnic tolerance,

¹⁸² Laura Jacobs & Joost van Spanje, *Prosecuted, yet Popular? Hate Speech Prosecution of Anti-Immigration Politicians in the News and Electoral Support*, 18 COMPAR. EUR. POL. 899 (2020).

¹⁸³ See Bundestag Election Results, FED. RETURNING OFFICER, <https://perma.cc/65UV-4JMR>.

¹⁸⁴ Liv Stroud, *German Far-Right AfD Party Scores Historic Win in Eastern State Election*, EU-RONEWS (Sep. 2, 2024, at 10:54 GMT); *German Far-Right Politician Björn Höcke Guilty of Using Nazi Slogan*, BBC NEWS (May 14, 2024); *German Court Fines AfD's Höcke over 2nd Use of Nazi Slogan*, DEUTSCHE WELLE (July 1, 2024); Alex Dziadosz, *The Trial of Björn Höcke, the 'Real Boss' of Germany's Far Right*, GUARDIAN (Aug. 29, 2024, at 00:00 ET).

¹⁸⁵ Christian Bjørnskov & Jacob Mchangama, *Freedom of Expression and Social Conflict* (Rsch. Inst. of Indus. Econ., Working Paper No. 1473, 2023), <https://perma.cc/9V34-36HM>.

greater minority access to political power, and lower levels of government discrimination. Notably, the study found no comparable positive effects associated with hate speech or defamation laws.¹⁸⁶

These findings support the view that free expression functions as a “safety valve” allowing grievances to be aired, contested, and countered in public rather than suppressed and displaced into more extreme or violent forms. In democratic societies, open contestation enables counterspeech, social norm formation, and peaceful conflict resolution, mechanisms that criminal law cannot replicate and may actively undermine.

D. Trust, Legitimacy, and Reporting Gaps

Even if criminalization were an effective remedy, its practical success depends on trust in institutions. Here, the evidence is troubling. FRA data consistently show that many victims of hate-motivated incidents, particularly Roma, Muslims, Jews, people of African descent, and LGBTQ+ individuals, do not report incidents to authorities. Common reasons include the beliefs that “nothing would change,” that incidents are “too trivial,” or that the police cannot be trusted.¹⁸⁷

This raises a fundamental issue, namely that the alleged protective function of these laws is undermined. While underreporting is a common feature across various categories of ordinary crime, the alleged protective function of hate speech legislation is severely undermined when targeted communities systematically distrust the state apparatus. Expanding criminal provisions under these conditions risks reinforcing existing inequalities rather than alleviating them. If hate speech laws are perceived as selectively enforced or politically instrumentalized, civic trust erodes further, transforming a tool intended for minority protection into a mechanism that lacks the baseline legitimacy required to function effectively.

III Criminal law that lacks legitimacy cannot function as an effective tool for social protection.

¹⁸⁶ Claudia W. Kramer, *Spewing Hate or Speaking Tolerance: How Free Speech Affects Social Norms and Minority Rights* (Liberty & L. Ctr., Research Paper, 2025), <https://perma.cc/S9NB-NHBB>.

¹⁸⁷ EUR. AGENCY FOR FUNDAMENTAL RTS., FRA BRIEF: CRIMES MOTIVATED BY HATRED AND PREJUDICE IN THE EU (2013), <https://perma.cc/5RPH-W3AY>.

E. Implications for Necessity and Proportionality

Empirical evidence does not support the assumption that expanding hate speech criminalization reduces harm in democratic societies. At best, the effects are uncertain; at worst, repression may intensify polarization, fuel radicalization, and legitimize extremist narratives.

Under IHRL, such uncertainty matters. The HRC has repeatedly emphasized that restrictions on expression must be justified by demonstrable necessity, not speculation. Where evidence of effectiveness is weak or contested, criminal sanctions, being the most intrusive response, cannot satisfy the proportionality requirement.

This is especially salient in the EU context, where criminalization interacts with the DSA to produce large-scale, automated enforcement with documented chilling effects on lawful speech. When combined with scope creep, democratic displacement, and global adoption of similar measures through the “Brussels Effect,” the empirical case for further criminalization becomes weaker still.¹⁸⁸

V. THE EXTERNAL DIMENSION OF EUROPEAN HATE SPEECH REGULATION

When liberal democracies legislate against certain forms of expression, authorities elsewhere can portray analogous measures as consistent with European standards. This does not imply direct causation. Rather, it highlights the interpretive space that opens when the EU or its Member States adopt speech restrictions that can later be recontextualized in jurisdictions with weaker constitutional safeguards.

Within this broader landscape, the so-called “Brussels Effect,” the EU’s capacity to externalize its internal regulatory standards through market size, legal credibility, and regulatory coherence, cannot be ignored. While the concept is often associated with positive spillovers in areas such as competition or data protection, its application in speech regulation raises more complex concerns. As Anu Bradford notes, global actors frequently adopt EU rules not because they are encouraged to do so, but because European frameworks are perceived as credible regulatory benchmarks.¹⁸⁹ In the speech context, however, that credibility may be invoked in political and legal environments markedly different from those in which EU rules were originally designed.

¹⁸⁸ See ANU BRADFORD, *THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD* (2020).

¹⁸⁹ *Id.*

A central illustration of this dynamic is Germany's NetzDG.¹⁹⁰ Adopted in 2017 and later superseded by the DSA, the NetzDG introduced accelerated takedown deadlines, substantial administrative fines, and detailed reporting obligations for major social media platforms.¹⁹¹ Within two years of its adoption, *Justitia* documented that more than 20 countries, mostly authoritarian or hybrid regimes, had cited NetzDG in formal proposals or informal legislative discussions or had enacted laws reflecting its core features.¹⁹² Some governments adopted the procedural architecture of strict deadlines and platform liability, while others explicitly invoked the German model to justify broader State intervention in online expression.¹⁹³

Less than two weeks after the NetzDG entered into force, for instance, members of President Putin's party introduced legislation requiring social media platforms to delete unlawful content within 24 hours of notification, explicitly referencing the NetzDG.¹⁹⁴ Reporters Without Borders described the proposal as essentially a "copy-and-paste" version of the German law.¹⁹⁵

In 2020, Turkey's parliament passed a law further regulating online publications.¹⁹⁶ A government spokesperson stated that "we looked at Germany as our starting point for all of it, from ensuring our freedom in using social platforms, to combating disinformation, to protecting the right to privacy and the protection of our data."¹⁹⁷

¹⁹⁰ NetzDG, *supra* note 114.

¹⁹¹ *Id.*

¹⁹² JACOB MCHANGAMA & NATALIE ALKIVIADOU, JUSTITIA, THE DIGITAL BERLIN WALL: HOW GERMANY (ACCIDENTALLY) CREATED A PROTOTYPE FOR GLOBAL ONLINE CENSORSHIP—ACT TWO 21 (2020), <https://perma.cc/EN7W-5EUN>.

¹⁹³ *Id.*

¹⁹⁴ See *Russian Bill Is Copy-and-Paste of Germany's Hate Speech Law*, REPS. WITHOUT BORDERS (July 19, 2017), <https://perma.cc/94CN-L9WV>.

¹⁹⁵ *Id.*

¹⁹⁶ İnternet Ortamında Yapılan Yayınların Düzenlenmesi ve Bu Yayınlar Yoluyla İşlenen Suçlarla Mücadele Edilmesi Hakkında Kanunda Değişiklik Yapılmasına Dair Kanun [Law Amending the Law on Regulation of Publications on the Internet and Combating Crimes Committed by Means of Such Publications], No. 7253, July 29, 2020, Official Gazette No. 31202, <https://perma.cc/MEK4-7BXG>.

¹⁹⁷ Gul Tuysuz, *Critics Fear Turkey's New Social Media Law Could Hurt Freedom of Expression. Here's How*, CNN (July 29, 2020), <https://perma.cc/NDJ5-NRDM>.

In late 2017, Venezuela adopted the “Constitutional Law Against Hatred, for Tolerance and Peaceful Coexistence,” imposing fines on social networks that fail to remove alleged hate speech within six hours and criminal penalties of up to 20 years’ imprisonment for individuals who publicly promote or incite hatred or discrimination based on a wide range of characteristics, including political identity.¹⁹⁸ The first vice-president of the National Constitutional Assembly stated,

Germany passed a law that requires companies, such as Twitter or Facebook, to delete messages that are reported as offensive by users and will have penalties of up to 50 million euros if they do not rectify in 24 hours or less. Likewise, YouTube and other social networks apply these severe sanctions and have had positive results: they have lowered the promotion of hate on social media.¹⁹⁹

The ECtHR’s case law on blasphemy discussed in Part II.A above is another example of how speech-intrusive norms originating in Europe can severely influence speech-restrictive legal frameworks in authoritarian countries. The Organization of Islamic Cooperation’s Independent Permanent Human Rights Commission “recalled and reaffirmed . . . the jurisprudence emerging from the European Court of Human Rights, which validates restrictions on freedom of expression criticizing religious beliefs where such expression constitutes incitement to hatred and is deemed offensive to the adherents of a particular religion.”²⁰⁰

The consequences of this cross-fertilization of speech-intrusive norms have been highlighted by civil society actors and dissidents in authoritarian states. In an interview with FoFS, Alexander Verkhovsky, Director of the SOVA Center in Moscow, observed,

The balance between protecting liberties and preventing incitement has moved far beyond any reasonable position in our country and across Central Asia. . . . What matters most for us, however, is the comparison with Western Europe, because most of

¹⁹⁸ Ley Constitucional Contra el Odio, por la Convivencia Pacífica y la Tolerancia [Constitutional Law Against Hatred, for Peaceful Coexistence and Tolerance], art. 20, Gaceta Oficial No. 41.276, Nov. 10, 2017 (Venez.), <https://perma.cc/B6JQ-QM9V>.

¹⁹⁹ Comisión Revisará Ley que Regular las Redes Sociales [Commission to Review Law Regulating Social Media], EL NACIONAL (Aug. 23, 2017), <https://perma.cc/4LX6-4ASQ>.

²⁰⁰ Aswad & Kaye, *Convergence & Conflict*, *supra* note 39, at 170 n.21 (quoting *On the Occasion of ‘International Human Rights Day’ IPHRC Welcomes the Adoption of ‘Cairo Declaration of the OIC on Human Rights’ and Calls upon Member States to Apply Human Rights Based Approaches in Dealing with the Socio-Economic and Health Related Challenges of COVID-19 Pandemic*, ORG. OF ISLAMIC COOPERATION INDEP. PERMANENT HUM. RTS. COMM’N (Dec. 10, 2020)).

our repressive norms and practices are justified by pointing to supposedly similar measures in places like Germany or France. So any problematic decision in Western Europe is immediately invoked in our region.²⁰¹

Tarık Beyhan, a human rights defender from Turkey, likewise noted:

In Türkiye, the offense of incitement to hatred and hostility . . . should logically function as a safeguard for minorities who may be targeted by hate crimes. Instead, it is widely used to shield the majority and even the government itself from criticism and to suppress political groups that are already in the minority. . . . Many regulations that restrict fundamental rights and freedoms and function as tools of repression in Türkiye are often justified by claiming that similar rules exist in Western Europe.²⁰²

Similar concerns arise in EU candidate countries, where regulatory alignment is encouraged through accession conditionality. In the July 2021 EU-Georgia annual human rights dialogue, both entities emphasized “the importance of making further efforts to address hate speech, incitement to hatred, and intolerance,” including through “[e]ffective . . . prosecution of hate speech.”²⁰³ However, the application of these commitments has raised concerns under Article 10 of the ECHR. The Council of Europe has warned that certain uses of Georgian hate speech provisions are incompatible with European freedom of expression standards, noting that “[t]o ensure clarity, Article 55(2) [the hate speech clause] should be amended to expressly state that critical and/or offensive speech is not to be considered hate speech.”²⁰⁴

Transparency International Georgia has echoed these concerns, cautioning that,

²⁰¹ Discussion with Alexander Verkhovsky, Sept. 16, 2025. Text on file with authors.

²⁰² Discussion with Tarık Beyhan, Nov. 14, 2025. Text on file with authors.

²⁰³ See TALANDER JANSEN & HANNAH AHAMAD MADATALI, EUR. PARLIAMENTARY RSCH. SERV., PE 699.489, ASSOCIATION AGREEMENT BETWEEN THE EU AND GEORGIA: EUROPEAN IMPLEMENTATION ASSESSMENT (UPDATE) 10, 71 (2022), <https://perma.cc/8G4B-BFZ6>.

²⁰⁴ Council of Europe, Directorate General Human Rights and Rule of Law, Information Society Department, *Opinion on the Law of Georgia on Broadcasting* 35, DGI (2023)1 (Feb. 21, 2023), <https://perma.cc/KJ9P-VKPT>; see also Giorgi Lomsadze, *Georgia's Hate Speech Legislation Spurs Censorship Concerns*, EURASIANET (Oct. 27, 2023), <https://perma.cc/VM2H-78YR>.

Legal regulation of hate speech contains a threat against the background of the authorities' attitude to critical opinion, the lack of strong and independent regulatory bodies, and the low confidence in the judicial system.²⁰⁵

These examples illustrate how EU-driven standards, even when adopted with rights-protective intentions, can be selectively appropriated in more fragile environments. In such contexts, European hate speech frameworks risk serving not as safeguards for pluralism but as legitimizing reference points for censorship and the suppression of dissent under the banner of European alignment.

VI. ALTERNATIVE APPROACHES TO THE EUROPEAN CRIMINALIZATION OF HATE SPEECH

If the objective is to address the harms associated with hateful expression while safeguarding democratic pluralism, criminal law should remain a measure of last resort, and one largely exhausted by existing legislation. The EU already possesses tools that allow it to respond to discrimination and hostility in more targeted and rights-compatible ways. Before expanding the scope of EU criminal law, institutions should assess whether existing civil, regulatory, and participatory mechanisms offer more proportionate solutions.

This section highlights two such pathways. First, counterspeech provides a non-coercive means of challenging harmful expression and strengthening democratic resilience. Second, the EU's non-discrimination framework addresses concrete harms linked to discriminatory expression within defined contexts. Together, these approaches illustrate that effective action need not depend on the harmonization of expansive criminal prohibitions.

A. *The Case for Counterspeech*

Counterspeech has been defined as “any direct response to hateful or harmful speech that seeks to undermine it.”²⁰⁶ At its core, counterspeech aims to reduce the influence and persuasive force of harmful expression through non-coercive means, including factual correction, empathetic engagement, humor, solidarity, or principled disagreement. It may be spontaneous or organized and can target both the original speaker and the wider audience that observes the exchange

²⁰⁵ *By the Force of Law: The Georgian Media Faces the Danger of Censorship*, TRANSPARENCY INT'L GEOR. (Oct. 27, 2023), <https://perma.cc/7YTW-S94G>.

²⁰⁶ *Why Talk Back?*, DANGEROUS SPEECH PROJECT, <https://perma.cc/84GW-2VZL>.

Unlike censorship, counterspeech seeks to persuade rather than silence. It enters public discourse to contest harmful ideas rather than removing them from view. As such, counterspeech is not merely an instrument but a democratic practice. It reflects a commitment to pluralism, moral contestation, and civic agency, which aligns closely with freedom of expression and deliberative democratic theory. Rather than relying on coercive state power, it empowers individuals and communities to respond to harmful expression through speech itself.

For these reasons, counterspeech offers a rights-compatible alternative to expanding criminal prohibitions. At the same time, counterspeech should not be understood as a comprehensive solution to hate speech or as a substitute for all forms of legal intervention. It does not neutralize structural discrimination, does not reliably change the behavior of committed ideologues, and cannot by itself address coordinated harassment, power asymmetries, or offline violence. Its relevance lies not in offering a universal remedy, but in illustrating how non-coercive, speech-based measures can operate alongside narrowly tailored legal responses without reproducing the chilling effects associated with expansive criminalization.

There is a growing body of evidence showing that counterspeech can influence discursive environments even when it does not alter the behavior of the original speaker.²⁰⁷ Organized hateful expression can shape public debate, and organized counterspeech can limit the diffusion of organized hate speech, reduce its normalization, and foster more constructive exchanges among observers.²⁰⁸ Empirical research underscores both the promise and the limits of counterspeech. Studies indicate that effectiveness depends on multiple factors, including the perceived credibility of the counterspeaker, tone, emotional intelligence and context.²⁰⁹ Empathy-

²⁰⁷ Jozef Miškolci, Lucia Kováčová & Edita Rigová, *Countering Hate Speech on Facebook: The Case of the Roma Minority in Slovakia*, 38 SOC. SCI. COMPUT. REV. 128 (2018).

²⁰⁸ Carla Schieb & Mike Preuss, *Governing Hate Speech by Means of Counter Speech on Facebook* (June 2016) (paper presented at the 66th International Communication Association Annual Conference), <https://perma.cc/GRM7-UBSR>.

²⁰⁹ Gina M. Masullo et al, *Effects of a High-Person-Centered Response to Commenters Who Disagree on Readers' Positive Attitudes Toward a News Outlet's Facebook Page*, 10 DIGIT. JOURNALISM 439 (2022).

based and fact-based responses tend to reduce hostility and foster deliberation, while confrontational or defensive approaches can escalate conflict.²¹⁰

An analysis of counterspeech to anti-Roma comments in Slovakia found little change in the behavior of original perpetrators, but threads containing counterspeech attracted more pro-Roma statements, encouraged alternative voices, and altered the broader discursive environment.²¹¹ Specifically, the study found that pro-Roma comments encouraged other participants with a pro-Roma attitude to become involved in online discussions. Counterspeech can also provide reassurance and recognition to targeted groups and increase the confidence and agency of those who engage with it.²¹²

A prominent example is the Swedish initiative #jagärhär (*#iamhere*), which later became #iamhereinternational, an international collective that intervenes in online discussions marked by racist or xenophobic content. Operating across several countries, members enter toxic comment threads with accurate, empathetic, and solidarity-based messages, guided by internal rules that prohibit personal attacks. Their explicit aim is to reach the “movable middle,”²¹³ the silent or undecided readers whose views may be influenced by what they observe. Research suggests that such interventions can shift tone, disrupt hateful narratives, and empower participants who might otherwise remain silent.²¹⁴

Counterspeech is already acknowledged, at least formally, within the EU’s digital governance framework. The Code of Conduct on countering illegal hate speech

²¹⁰ Dominik Hangartner et al., *Empathy-Based Counterspeech Can Reduce Racist Hate Speech in a Social Media Field Experiment*, PNAS, Dec. 6, 2021, art. e2116310118; Marc Ziegele & Pablo B. Jost, *Not Funny? The Effects of Factual Versus Sarcastic Journalistic Responses to Uncivil User Comments*, 47 COMM’N RSCH. 891 (2016).

²¹¹ Jozef Miškolci et al., *supra* note 207.

²¹² CATHY BUEGER, *WHY THEY DO IT: COUNTERSPEECH THEORIES OF CHANGE* (Dangerous Speech Project 2022), <https://perma.cc/98RC-EQCP>.

²¹³ Catherine Buerger, *#iamhere: Collective Counterspeech and the Quest to Improve Online Discourse*, 7 SOC. MEDIA + SOC’Y (2021).

²¹⁴ *Id.*

online recognizes the importance of counternarratives and civil society engagement, though in practice, implementation has often prioritized rapid removals. Under the DSA, the enhanced Code+ (2025) institutionalizes these commitments.²¹⁵

Rather than expanding criminal liability, the EU could build on this foundation by investing more systematically in counterspeech as a central pillar of its approach to hate speech. The EU should develop a clear roadmap for the role of counterspeech within its broader hate speech strategy, grounded in a robust empirical assessment of existing practices across Member States and online environments. Based on these findings, the EU could then designate targeted structural and funding support to civil society, scale effective models, facilitate cross-border cooperation, and ensure that counterspeech initiatives are embedded within a coherent, rights-respecting policy framework. This would empower counterspeakers and encourage platforms to facilitate and amplify counterspeech without converting these commitments into content-based enforcement obligations or accelerated removal requirements.

B. The EU Non-Discrimination Framework

EU non-discrimination law offers a structured framework for addressing some of the concrete harms that are frequently associated with hate speech, without transforming freedom of expression into a general object of regulation. The Race Equality Directive and the Employment Equality Directive operate within clearly delineated contexts such as employment, education, and access to goods and services in the case of the former, and employment and vocational training in the case of the latter.²¹⁶ Neither applies to the public sphere at large. They are confined to spaces where power asymmetries and patterns of exclusion are most acute, and they do not prohibit expression as such. Rather, they target discriminatory conduct, unequal treatment, and the creation of hostile or degrading environments within clearly bounded contexts, that avoids general prohibitions of political speech.

²¹⁵ *The Code of Conduct on Countering Illegal Hate Speech Online* +, EUR. COMM'N (Jan. 20, 2025), <https://perma.cc/F8XM-H8YY>.

²¹⁶ Council Directive 2000/43/EC of 29 June 2000 Implementing the Principle of Equal Treatment Between Persons Irrespective of Racial or Ethnic Origin, 2000 O.J. (L 180) 22, <https://perma.cc/8ZSV-G2FM>; Council Directive 2000/78/EC of 27 November 2000 Establishing a General Framework for Equal Treatment in Employment and Occupation, 2000 O.J. (L 303) 16, <https://perma.cc/U5RE-GED4>.

As demonstrated by the case law of the CJEU, particularly in cases such as *Feryn* (racism), *ACCEPT* (homophobia), and *NH* (homophobia), discriminatory statements may fall within the scope of EU equality law not because of their expressive content as such but because of their material effects on access, dignity, and equal participation.²¹⁷ Public statements that, for example, reinforce exclusion or contribute to hostile environments can, in specific settings, amount to discrimination even where no individual complainant is identified. This does not imply a general duty to police offensive speech. Instead, it underscores the importance of context: Equality law intervenes only where expression operates within institutional or market settings and produces tangible exclusionary effects.²¹⁸

Viewed in this way, EU non-discrimination law provides a valuable point of reference for a less speech-restrictive policy architecture on hate speech. Rather than expanding criminal law or moving prematurely toward harmonization, the EU could draw on the logic of its existing equality acquis to examine how harms linked to hateful expression are already addressed through civil and administrative mechanisms in specific contexts, to identify good practices and safeguards, and to assess whether targeted, non-binding guidance is warranted. Framed in this manner, EU action would offer constructive and positive alternatives to criminalization, while setting clear boundaries against expansive national implementation and remaining faithful to the principles of legality, necessity, and proportionality under IHRL.

CONCLUSION

The Commission's proposal to make hate speech an EU crime or harmonize definitions of online hate speech offences under Article 83(1) TFEU would mark a major shift in Europe's speech governance. Even if framed as "online only," it would push criminal law deeper into the regulation of public discourse across 27 Member States with distinct constitutional traditions, and it would do so on the

²¹⁷ Case C-54/07, *Centrum voor Gelijkheid van Kansen en voor Racismebestrijding v. Firma Feryn*, 2008 E.C.R. I-05187; Case C-81/12, *Asociația Accept v. Consiliul Național pentru Combaterea Discriminării*, ECLI:EU:C:2013:275 (Apr. 25, 2013); Case C-507/18, *NH v. Associazione Avvocatura per I Diritti LGBTI—Rete Lenford*, ECLI:EU:C:2020:289 (Apr. 23, 2020).

²¹⁸ For more on the issue, see Uladzislau Belavusau, *Fighting Hate Speech Through EU Law*, 4 AMSTERDAM L. FORUM 20 (2012); Natalie Alkiviadou & Uladzislau Belavusau, *Rien que des Mots: Counteracting Homophobic Speech in European and U.S. Law*, 21 INT'L J. DISCRIM. & L. 374 (2021).

basis of concepts that remain legally contested and difficult to operationalize with precision.

First, it falls short on legality. The Commission's communications frame hate speech through broad moral and sociological descriptions rather than the constituent elements required for criminal liability. Parliament's call for an open-ended expansion of protected characteristics would make the boundaries of the offence responsive to shifting political and social pressures. Criminal law cannot be built on a moving target without sacrificing foreseeability, inviting selective enforcement, and chilling lawful speech, particularly on contested matters of public interest.

Second, the Commission has not demonstrated necessity. Europe already possesses extensive tools addressing incitement and discriminatory harm, including national criminal codes, the Framework Decision, and a far-reaching digital enforcement architecture under the DSA. The proposal does not identify a concrete enforcement gap that EU-level criminalization uniquely fills. Nor does it show that less intrusive measures, civil remedies, administrative enforcement in bounded contexts, counterspeech, and non-coercive democratic responses are insufficient. Shifting political rationales may create a sense of urgency; they do not substitute for evidence of indispensability.

Third, the proposal is disproportionate. Criminal sanctions carry uniquely severe consequences for speakers and predictable chilling effects for public debate. Yet the initiative, as presently framed, omits limiting criteria central to rights-compliant incitement regulation: intent, contextual assessment, and a credible threshold of likelihood and imminence of harm. Without those safeguards, enforcement will not remain confined to the narrow category of expression that international standards require states to prohibit. It will predictably expand to cover polemical political speech, satire, artistic expression, and minority advocacy, exactly the speech most vital to democratic contestation and accountability.

These concerns are amplified by the proposal's systemic spillovers. Under the DSA, new criminal prohibitions would automatically widen what counts as "illegal content," increasing takedown pressure and pushing platforms toward risk-averse over-removal, often through automated systems ill-suited to context and nuance. The result would be a larger, quieter restriction on lawful expression: less visible than prosecutions, but more pervasive in shaping what citizens can say and hear online.

The empirical case for such a tradeoff is also weak. The available literature does not support confident claims that expanding criminal hate-speech prohibitions reduces intolerance or violence in stable democracies. It suggests, instead, that punitive restrictions can generate counterproductive dynamics, polarization, grievance narratives, and displacement into more radical spaces, while undermining trust and legitimacy.

Finally, the proposal's implications extend beyond Europe. EU speech governance exerts global influence, and expansive European models are frequently invoked elsewhere to legitimize suppression of dissent. Creating an EU-level regime harmonizing hate speech or online hate offences would strengthen the portability and apparent legitimacy of broad criminal restrictions in legal environments that lack Europe's safeguards.

A credible democratic strategy against hatred must begin with precision and restraint. International human rights law does not treat criminalization as the default response to offensive or prejudiced expression. It reserves criminal sanctions for a narrowly defined category of intentional incitement and places the burden on states to demonstrate necessity and proportionality.

Within the EU's own legal order, equality law and targeted civil and administrative mechanisms already address concrete harms in bounded contexts without converting freedom of expression into a general object of penal regulation. Counterspeech and civic resilience measures, while imperfect, offer rights-compatible tools that strengthen democratic agency rather than substituting coercion for persuasion.

At a moment of heightened polarization and geopolitical pressure, the EU faces a choice. It can continue down a path that normalizes expansive criminal governance of speech, or it can reaffirm pluralism, open debate, and restraint as core democratic commitments. The Commission's proposal, as currently framed, risks weakening those commitments and, in doing so, becoming a cure worse than the disease.