

Criminal Liability for Hate Speech on Social Media: Problems and Solutions

Tileumuratova Madina Ilhom qizi

Karakalpak State University named after Berdaq, Faculty of Law, Jurisprudence Program, Undergraduate Student, Group 105, Uzbekistan

Received: 27 May 2026; **Accepted:** 23 June 2026; **Published:** 16 July 2026

Abstract: This article analyzes the phenomenon of hate speech on social media and the issues surrounding the establishment of criminal liability for it. Drawing on Article 156 of the Criminal Code of the Republic of Uzbekistan, international legal standards, the European Union's Digital Services Act, and comparative legal practice, the authors examine the boundaries of criminalizing hate speech. The article identifies the main problems in establishing criminal liability — definitional ambiguity, jurisdictional issues, and balancing with freedom of speech — and proposes solutions to them.

Keywords: Hate speech, hate speech, social media, criminal liability, freedom of speech, Criminal Code Art. 156, DSA, ICCPR.

Introduction: In the 21st century, the rapid development of social media has fundamentally transformed the sphere of communication. Facebook, Instagram, Telegram, TikTok, and other platforms today unite billions of users. However, this process has also created fertile ground for the unprecedented spread of a socially dangerous phenomenon — hate speech. Hate speech on social media based on national, racial, religious, or other characteristics not only violates human dignity but also creates the risk of instability, violence, and even incitement to genocide in society.

A vivid example of this is the events in Myanmar: the UN's Independent International Fact-Finding Mission found that hate speech by military and extremist groups on the Facebook platform played a significant role in supporting crimes committed against the Rohingya people, including acts of genocide. This situation necessitates studying the problem of criminal liability for hate speech at the intersection of international law, national legislation, and platform responsibility. In the Republic of Uzbekistan, digitalization processes are also developing rapidly: the number of Internet users among the country's population is growing year by year. Accordingly,

improving legislative mechanisms against hate speech on social media is becoming an urgent issue. Although Article 156 of the Criminal Code serves as the main legal source in this field, the specific characteristics of the digital environment leave many legal questions open.

DEFINITION OF HATE SPEECH AND INTERNATIONAL STANDARDS

There is still no single internationally accepted definition of hate speech. The UN Strategy and Plan of Action on Hate Speech, established under the leadership of the UN Secretary-General, defines hate speech as follows: "any kind of communication in speech, writing, or behavior that attacks or uses pejorative or discriminatory language with reference to a person or a group based on who they are — that is, on the basis of their race, ethnic origin, religion, gender, nationality, color, descent, sexual orientation, or other identity factor." The European Union, in the Council Framework Decision, defines hate speech as "public incitement to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, color, religion, descent, or national or ethnic origin." This definition is now also applied within the framework of the Digital Services Act.

International legal foundations: International law recognizes two main approaches to hate speech. The first is the prohibitive approach: Article 20(2) of the International Covenant on Civil and Political Rights obliges states to adopt laws prohibiting any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. The second is the balancing approach: Article 19 of the ICCPR and Article 10 of the European Convention on Human Rights guarantee freedom of expression and provide that it may be restricted only when necessary and proportionate.

The UN Human Rights Committee has established a three-part test for restricting freedom of expression: (1) the restriction must be provided by law; (2) it must serve a legitimate purpose; and (3) it must be necessary and proportionate. This standard serves as an important criterion for distinguishing between forms of hate speech that are subject to criminal liability and those that are not.

LEGISLATION OF UZBEKISTAN: ANALYSIS OF ARTICLE 156 OF THE CRIMINAL CODE

Article 156 of the Criminal Code of the Republic of Uzbekistan, entitled "Incitement of National, Racial, Ethnic, or Religious Hostility," is the main source of criminal liability for hate speech. This article establishes criminal liability for hate speech as follows: preparing, storing, or distributing materials promoting national, racial, ethnic, or religious hostility with the aim of dissemination, if committed after an administrative penalty has already been imposed for such acts, is punishable by a fine of up to six hundred times the base calculation amount, or correctional labor for up to three years, or restriction of liberty for one to three years, or deprivation of liberty for up to three years (Criminal Code of the Republic of Uzbekistan, 2020).

From a comparative perspective, Article 174 of the Criminal Code of the Republic of Kazakhstan provides for restriction of liberty or deprivation of liberty from 2 to 7 years for hate speech on social media. The Constitution of Kyrgyzstan directly prohibits the promotion of national, ethnic, racial, or religious hostility, or gender or other forms of social superiority that give rise to discrimination, hostility, or violence.

Noteworthy are the amendments introduced into the Criminal Code and the Code on Administrative Liability of Uzbekistan by the law adopted on October 31, 2023: liability was strengthened for "promoting the superiority or inferiority of one category of citizens over another based on national, racial, ethnic, or religious affiliation." These amendments are a clear example of the legislative response to the growing spread of hate speech on social media.

INTERNATIONAL EXPERIENCE: The European Union has formed the most advanced regulatory framework for hate speech. The Digital Services Act, adopted on October 19, 2022, came into full force on February 17, 2024. The DSA imposed on Very Large Online Platforms — giant platforms such as Meta, Google, TikTok, and X — an obligation to assess systemic risks and remove illegal content such as hate speech. Fines of up to 6 percent of annual global turnover are established for violations.

On January 20, 2025, the European Commission and the European Board for Digital Services integrated the "revised Code of Conduct on Countering Illegal Hate Speech Online" into the DSA framework. This step transformed voluntary commitments into legal obligations and further strengthened platform accountability. The European Court of Human Rights (ECtHR) has also created important precedents on this issue. In *Sanchez v. France* (2023), the Court for the first time recognized that holding a Facebook user criminally liable for failing to remove hate-speech comments left by third parties on their wall does not necessarily violate Article 10 of the ECHR. This ruling signifies an important legal shift regarding the liability of platform users.

In the United States, the regulation of hate speech differs fundamentally from the European model. The First Amendment to the Constitution provides broad protection for freedom of speech, such that "pure" hate speech is covered by constitutional protection. Section 230 of the Communications Decency Act, adopted in 1996, exempts social media platforms from liability for user-generated content, granting these platforms broad freedom of action.

However, U.S. platforms are also being forced to make changes due to the requirements of European regulations. This phenomenon is known as the "Brussels Effect," referring to the global spread of European standards. In 2016, Facebook, Microsoft, Twitter, and YouTube signed a code of conduct on countering hate speech with the European Commission, obliging the platforms to review reported hate-speech content within 24 hours.

Precedents in international criminal law: The link between hate speech and the most serious of crimes — incitement to genocide — is also reflected in international criminal law. In the Nuremberg trials, the publisher of the newspaper *Der Stürmer* was convicted for supporting genocide through hate speech. The 1948 Convention on the Prevention and Punishment of the Crime of Genocide established direct and public incitement to genocide as a crime. At the International Criminal Tribunal for Rwanda, the heads of two media

outlets were convicted specifically on the basis of incitement to genocide.

MAIN PROBLEMS IN ESTABLISHING CRIMINAL LIABILITY

The absence of a single, universally accepted definition of hate speech is the most fundamental problem in establishing criminal liability. Different states have adopted different definitions based on their own historical, philosophical, and constitutional traditions. As a result, the key elements required for criminal liability — intent to incite hostility, potential harm, and the identifiability of the target group — often remain unclear. The EU's Digital Services Act has itself been criticized for not clearly defining concepts such as "hate speech" and "illegal content": this ambiguity may push platforms toward over-removal of content.

Jurisdictional problem: The cross-border nature of social media greatly complicates the practical exercise of criminal jurisdiction. Because different states take different approaches in their hate speech laws, content that is legal in one country may give rise to criminal liability in another. Moreover, platform companies are often registered in other countries, making it legally complex to apply relevant national laws to them. State attempts to regulate hate speech through centralized national or regional criminal laws have shown limited effectiveness.

Balancing with freedom of speech: Balancing criminal liability for hate speech with freedom of expression (ICCPR Art. 19, ECHR Art. 10) is one of the most severe tests facing democratic societies. As stated by the Inter-American Court of Human Rights in *Kimel v. Argentina* (2008), the use of criminal prosecution for statements on matters of public interest should be exceptional and a measure of last resort. In Finland, a 2019 case in which a member of parliament was charged with criminal hate speech for posting a Quranic verse on Twitter, together with a case involving a church pamphlet, was reviewed by the Supreme Court of Finland, illustrating just how delicate this issue is. Technologies for automatically detecting hate speech remain insufficiently advanced. AI-based moderation systems face problems in fully grasping context, properly assessing subtle meanings, and producing false positives. A 2024 study by Vanderbilt University's Future of Free Speech center found that, due to regulatory pressure from the DSA, platforms are also removing content that is harmful but not illegal. This creates the risk of violating the rights of innocent users.

Improving legislation:

First, it is advisable to introduce clarifying additions to Article 156 of the Criminal Code of the Republic of Uzbekistan that take into account the specific

characteristics of the digital environment and social media. In particular, it is necessary to introduce a definition of "distribution" that covers online publication and sharing, and to establish criteria that clearly distinguish hate speech from incitement to genocide, from forms leading to mass unrest, and from content that may be assessed as opinion. It is recommended that the UN Rabat Plan of Action (2012), which proposes a six-part test for such assessments, be integrated into national legislation.

Hate speech on social media cannot be curbed by criminal law instruments alone. International law and comparative legal research show that an effective approach must be multifaceted: combining lawful prosecution by state authorities, self-regulation by platform companies, and oversight by civil society organizations. In accordance with international human rights standards and business principles, platform companies should assume primary responsibility for addressing hate speech. It is recommended that a "localization" strategy be applied in detecting and assessing hate speech. This approach involves adapting platform companies' terms of service and guidelines to the local context and to cultural and linguistic characteristics, and involving local experts and civil society groups in the moderation process. This is consistent with international human rights standards and reduces the ambiguities that arise in detecting hate speech.

International cooperation mechanisms: It is necessary to strengthen international cooperation mechanisms to curb the cross-border nature of hate speech. Within the framework of relevant international bodies and the Council of Europe's Convention on Cybercrime, it is recommended that additional protocols on hate speech be used to develop mutual legal assistance, information exchange, and joint prosecution mechanisms among partner states. The Republic of Uzbekistan should intensify its participation and cooperation within these international treaties.

CONCLUSION

Hate speech on social media refers to insults, discrimination, and threats directed at someone based on their national, racial, religious, or other characteristics. Today this phenomenon is a global problem that undermines citizens' dignity and gives rise to violence and extremism in society.

Legislation and liability: In the Republic of Uzbekistan, Article 156 of the Criminal Code establishes liability for hate speech, providing for deprivation of liberty of up to 3–5 years. Under international law, the International Covenant on Civil and Political Rights and the European Convention on Human Rights protect freedom of

expression and provide that it may be restricted only when necessary and proportionate.

International experience: The European Union's Digital Services Act obliges internet platforms to remove hate speech and establishes fines for violations. The European Court of Human Rights issues rulings on hate speech cases and recognizes incitement to genocide as the most serious of crimes.

Main problems: The absence of a single definition of hate speech, the cross-border nature of social media, and the imbalance between freedom of speech and criminal liability. Platform companies may take excessive measures in removing content, which restricts free speech.

Solution: Combating hate speech requires state laws, self-regulation by internet platforms, civil society oversight, and international cooperation. Raising users' media literacy and maintaining transparent moderation systems are essential.

Conclusion: Hate speech threatens the security of society, but combating it can be achieved not only through law but also through interethnic harmony, religious tolerance, and honest dialogue among citizens.

REFERENCES

International documents and legal acts:

1. International Covenant on Civil and Political Rights, December 16, 1966, UN Treaty Series, Vol. 999, p. 171.
2. UN Human Rights Committee, General Comment No. 34 on Article 19 (freedoms of opinion and expression).
3. UN Strategy and Plan of Action on Hate Speech. United Nations, 2019.
4. UN Office on Genocide Prevention. Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred. Geneva, 2012.
5. UN Independent International Fact-Finding Mission on Myanmar.
6. Convention on the Prevention and Punishment of the Crime of Genocide, December 9, 1948.
7. Council Framework Decision 2008/913/JHA on Combating Certain Forms and Expressions of Racism and Xenophobia. Official Journal of the European Union, 2008.
8. Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act — DSA), October 19, 2022.
9. European Commission and European Board for

Digital Services. Revised Code of Conduct on Countering Illegal Hate Speech Online, January 20, 2025.

Court decisions and precedents:

10. ECtHR, Sanchez v. France, Application No. 45581/15, Grand Chamber Judgment, May 15, 2023.
11. ECtHR, Delfi AS v. Estonia, Application No. 64569/09, Grand Chamber Judgment, June 16, 2015.
12. ECtHR, Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary, Application No. 22947/13, February 2, 2016.
13. Inter-American Court of Human Rights, Kimel v. Argentina, Judgment of May 2, 2008, Series C No. 177.
14. International Criminal Tribunal for Rwanda, Prosecutor v. Nahimana, Barayagwiza and Ngeze (Media Case), Case No. ICTR-99-52-T, December 3, 2003.

Legislative sources of the Republic of Uzbekistan:

15. Criminal Code of the Republic of Uzbekistan, September 22, 1994. Latest revision: 2026. Lex.uz
16. Constitution of the Republic of Uzbekistan, 1992 (as amended in 2023).
17. Law of the Republic of Uzbekistan on Amendments and Additions to the Criminal Code and the Code on Administrative Liability, No. O'RQ-874, October 31, 2023.
18. Law of the Republic of Uzbekistan on Amendments and Additions to the Criminal Code, the Criminal Procedure Code, and the Code on Administrative Liability, No. O'RQ-899, January 19, 2024.