
RETHINKING HATE SPEECH REGULATION ON SOCIAL MEDIA WITH SPECIAL REFERENCE TO RELIGIOUS HATE SPEECH

Preethi H, Postgraduate, LLM (Constitutional Law and Legal Order),
Department of Legal Studies, University of Madras

ABSTRACT

The creation of the Internet and social media, along with various platforms, has made information and communication fast and easy to access. These social media platforms have become the new meeting places for public, social, and political discussions, as well as for sharing thoughts and ideas. As a result, social media has given the public a global audience for their messages and has boosted the participatory and representative aspects of democracy. However, there is always the downside of technological use, too. Digital communication is very quick and reaches many people, and can sometimes reach the specific audience, resulting in the proliferation and escalation of hate speech. Unlike other forms of communication, digital messages can be sent almost instantly to a large audience. The impact of a digital message is immediate and very real. According to the hate speech lab report Hate speech against religious minorities, especially Muslims, increased by 74%, Totaling 1,165 cases in 2024, compared to 668 cases in 2023.¹ The aim of this research is to determine whether the existing legislation and regulations are adequate to control religious hate speech, especially through social media.

Keywords: communication, social media, hate speech, religious hate speech, religious minorities, legal framework, regulation.

¹ Indian Hate Speech, “Report on Combating Hate Speech” (2025).

Introduction

Hate speech means “any speech disparaging a group on the grounds of colour, race, ethnicity, nationality, religion, sex, sexual orientation, gender identity, or disability, or a person who identifies with such a group”². There is no universally accepted definition of hate speech. Many scholars have tried to define what is hate speech or what constitutes hate speech. Richard Delgado³ during the 18th century proposed a definition of hate speech as a use of language with the intention to demean a group that reasonable people would consider it as a racial insult. Kenneth Ward⁴ gives the meaning of hate speech as a form of expression through which the speaker primarily intends to humiliate or incite hatred against their targets.

India has long been known as an ancient civilization with unique traditions; its diversity is evident in religious beliefs and customs. People who practice Hinduism differ from those who practice other religions, yet they remain united by the social equality they exhibit toward one another. In contrast to Western nations, which share a common culture, India is a nation characterized by diverse values and cultures among its people, and also follows the concept of Pluralism, which is embraced by the state via a secular type of government. The Constitution has secularism mentioned within it as well as through practice by its citizens. The existence of this harmony among the people is an essential component of the progress of the nation and living together peacefully. Besides being diversified in various aspects, one significant area of diversity is religion. India has various religions, with Hinduism being the dominant one, while Muslim, Buddhists, Christianity, Zoroastrianism, and Judaism are among the minorities.⁵ To maintain this pluralism, there are numerous challenges, one of which is communalism.⁶ Given that India is a nation with numerous religions practiced and where individuals view religion as integral to their lives, any disparaging remarks against a specific religion can lead to communal riots and national disintegration.⁷ Before independence, significant communal strife arose in the twentieth century due to the release of a pamphlet named ‘Rangila Rasul’ by Mahashay Rajpal, which was perceived as derogatory toward the Muslim faith and thus faced backlash

² Collins dictionary

³ Richard Delgado, “Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling” 17 Harvard Civil Rights- Civil Liberties Law Review 133 (1982).

⁴ Kenneth D. Ward, “Free Speech and the Development of Liberal Virtues: An Examination of the Controversies Involving Flag-Burning and Hate Speech” 52 *University of Miami Law Review* 733 (1988).

⁵ National Commission for Minorities (NCM) Act, (Act 19 of 1992), s.2(c).

⁶ Meena Ketan Sahu and Pankaj Singh, “Hate Speech and its Impact on the Democracy of India: A Critical Study” 7 *International Journal for Multidisciplinary Research (IJFMR)* (2025).

⁷ *Ibid.*

from the Muslim community. Besides being insulting, this sparked community conflict within society.⁸ Following this incident, the government highlighted the necessity for a legal structure to tackle religious speech and to avert communal violence stemming from it, leading to the introduction of Section 295A in the Indian Penal Code. Throughout history, hate speech has been made illegal.⁹

The hate speech impedes the constitutional values encompassed in the preamble i.e secular, fraternity, and dignity, because these values can exist only when there is reciprocal recognition and respect between the religious groups. The state must respond to this issue because this has the potential to violate the right to life with dignity as well. The hate speech is not merely offensive, but it is a misuse of freedom of speech and expression, rooted in the Constitution as a fundamental right under Part III of the Indian Constitution.

The **hypothesis** is that the existing legal framework is inadequate to effectively address the evolving nature of religious hate speech on social media.

Chapter 1- hate speech and religious hate speech

The internet and other digital tools have really changed the way people talk about things. It makes it easy for people to say negative things on the internet and other digital tools. These negative things can spread fast. Even though people say negative things on the internet, we do not have a clear idea of what exactly hate speech is on the internet, which makes it hard to stop hate speech. Hate speech is a big issue against religious minorities, and social media is making hate speech worse. When we talk about spaces, digital self-representation is really important because it is where people create and argue about their identities on the internet and other digital tools. These online representations of people and communities are not fair or neutral they are influenced by who has power and social status on the internet. Hate speech on the internet and other digital tools messes with this process of forming identities by attacking people because of their religion on the internet and other digital tools, which stops them from being able to express themselves and say who they are on the internet.¹⁰ Representation in

⁸ Yash Siha, "Blasphemy law in India: An Overview," *Manupatra*, 2021, available at <<https://articles.manupatra.com/article-details/Blasphemy-Law-in-India-An-Overview>> last visited on 5th March 2026.

⁹ *Ibid.*

¹⁰ Kyriaki Topida, "When Words Destroy Worlds: Online Hate Speech," *Language Science Press*, 2026, available at <<https://zenodo.org/records/18845596>> last visited on 1st April 2026.

spaces is shaped by power relations and socio-political hierarchies and hate speech targets individuals and groups based on their religious affiliation, undermining their capacity for self-identification and expression through digital self-representation, on the internet and other digital tools. Though social media are for anyone to share their thoughts, they have often been used as a tool to abuse minorities in the form of stigmatization or stereotyping them through visual messages, memes, harassment campaigns, etc. These acts tend to aggravate the violence and also affect the dignity of the person to whom it is being targeted. These acts marginalize minorities, and as a result, it creates a social inequality. Another concern is that this could bring an adverse effect where the victimized persons will rise up as a group of opposition. This can be solved by the existing legal framework but when the same happens on social media, there is a special concern is need because the spread of messages is so rapid on social media compared to the tradition methods of communication and also, the role of social media in intensifying this issue plays a crucial role in addressing this issue. Therefore, there is a need for special legislation to address this issue to balance the dignity of the minority and the protection of freedom of speech. This point can be seen in Justice Chauhan's remark made in the Supreme Court case about hate speech, where he said that hate speech cannot be evaluated merely based on its offensiveness or abusive content. Rather, its true impact lies in its capacity to systematically marginalize vulnerable groups and reinforce structures of inequality.¹¹

Meaning of hate speech

Hate speech lacks a universally agreed-upon definition. In the year 2020 united nations strategy and plan of action on hate speech describes the hate speech as “Any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group based on who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor” This definition was given by the united nation for countering the increase of hate speech. Since it is a policy framework rather than a treaty, it is not binding on the states.

Since there is no clear definition of what hate speech is on social media, each and every social media platform prescribes what constitutes hate speech in its terms and conditions.

¹¹ *Pravasi Bhalai Sangathan V. Union of India*, AIR 2014 SC 1591.

Hate speech incidents

The Centre for the Study of Organised Hates India Hate Lab project found out that there were 1,319 instances of hate speech against Christians and Muslims in 2025. The Centre for the Study of Organised Hate made the India Hate Lab. These things happened in 21 states and in Delhi, which is the National Capital Territory. This means that hate crimes against Muslims and Christians happened four times every day. The number of incidents for 2024 was 13 percent of the total hate crimes. From 2021 to 2022, there was a 97 percent increase in Muslim and Christian hate crime incidents and the majority occurred online. People uploaded 1,278 videos on Facebook, YouTube, Instagram and X. The India Hate Lab found out that more than 1,200 of these instances were online.¹² When the India Hate Lab looked at the cases they found that most of the things people said were posted on Facebook, then YouTube, Instagram and X. The India Hate Lab project by the Centre for the Study of Organised Hate is really important because it helps us understand hate speech against Christians and Muslims. The Centre for the Study of Organised Hate is doing a job, with the India Hate Lab project.

1. During the Delhi riot 2020 The police in Delhi found out that some things people posted on Facebook and sent on WhatsApp made people very angry. These messages made the Muslim and Hindu communities fight with each other. Over 1,000 of these messages were examined in court. This shows that what people post online can cause a lot of trouble in the world. Some people are saying mean things about other people online because of their caste.. Between 2023 and 2025 people made over 2,500 mean posts about Dalit communities on Facebook. The Delhi riots and these hateful posts are examples of how online hate speech can lead to real-world violence in Delhi riots and violence in places like Uttar Pradesh and Bihar.¹³

2. The terrorist attack in Pahalgam that happened in April 2025 was one of the examples where the. The India Hate Lab found out that there were 98 hate speech demonstrations across the country in just sixteen days. These demonstrations had a lot of talk, including people saying we should hurt others and using bad words about Muslim communities. People used media to spread this kind of talk really fast. Instagram accounts from groups helped spread the talk by posting together, and big Facebook groups made it easy to share quickly. YouTube also showed

¹² Indian Hate Speech, "Report on Combating Hate Speech" (2025).

¹³ Nirmal Ripathi, "Policies in Social Media to Control Religious and Caste Hate Speech: Need vs Misuse Risks" 8, *International Journal for Multidisciplinary Research (IJFMR)* (2026).

these events live so people could see what was happening away.¹⁴

Chapter 2- existing legal framework and international policies governing the hate speech on social media

i) International framework and policies

International Covenant On Civil And Political Rights (ICCPR)

Article 20(2) of the Covenants states that every state has the obligation on States to prohibit “any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence.

Rabat plan of action

The Rabat plan of action adopted in the year 2012. It was initially raised due to the emerging concern over xenophobia, hate speech, and religious intolerance. The Rabat plan of action is the recommendation given by several experts from the OHCHR (Office of the United Nations High Commissioner for Human Rights) during the workshop held to discuss the international human rights law. The main object of the workshop is to discuss the legislative and judicial practices and implementation of policies of the state towards the religious and racial incitement in consistency with Articles 19 and 20 of the ICCPR. Some of the key recommendations are

- a. to establish a threshold to distinguish between protected speech and prohibited speech
- b. Any restriction on freedom of speech must comply with the restrictions provided under article 20(2) of the ICCPR
- c. The state has the obligation to prohibit the speech that can cause hatred, discrimination, violence, etc
- d. The state must enact a balanced civil and legal framework regarding hate speech, and also, the minorities must be protected in this process.

¹⁴ David Sathuluri, “India’s Hate Speech Crisis and the Myth of Neutral Platforms” Tech Policy Press, 2014, available at < <https://www.techpolicy.press/indias-hate-speech-crisis-and-the-myth-of-neutral-platforms/>> last visited on 2nd April 2026.

ii) United Nations strategy and Plan of action on Hate Speech

The United Nations developed a strategy and Action Plan in 2019 under the auspices of the Secretary-General's strategy and plan of action for Hate Speech, launched by António Guterres, in response to the growing rate of hate speech, primarily in the digital era. The strategy is the outcome of the coordinated work of 16 entities under the UN Working Group on hate speech, and also this plan of action was developed by the United Nations Office on genocide. The strategy defined hate speech as “any form of communication, whether speech, writing, or behaviour and attacks or uses discriminatory language against individuals or groups based on identity factors such as religion, race, ethnicity or nationality”. The strategy provided thirteen commitments to the state to address hate speech. Some of its important commitments are

- a. It emphasises the state's need to monitor and analyse hate speech to have an immediate response.
- b. It also urges to support the victims who are vulnerable and also to provide speedy access to remedies.
- c. The state must counter the algorithmic issue in the digital platforms that help spread hate speech.

Indian Constitutional Framework

Article 19(1)(a)

The freedom of speech and expression is a fundamental right regardless of its content or expression or expressed through any medium, which includes social media. This right was expanded through the judicial interpretation, which indirectly includes the right to circulate, the right to dissent, right to commercial speech, right to compelled speech. Hate speech falls under restricted under Article 19(2).

Article 19(2)

Article 19(2) passed through several changes on the recommendation of various sub-committee reports. Later on the date of adoption it had the restriction related to i) slander ii) libel iii) defamation iv) contempt of court v) any matter which offends against decency and morality iv)

any matter undermines the security of the state or tends to overthrow the state. This restriction underwent changes after the 1st and 16th amendment currently the restriction for the freedom of the speech can be made on the grounds of i) interest of sovereignty and integrity of india, ii) the security of the state, iii) friendly relations with foreign sates, iii) public order v) decency or morality vi) contempt of court vii) defamation and viii) incitement to an offence. These restrictions gives the power to the state to enact any law which can reasonably restrict the freedom of speech and expression in order to protect the any i) individuals ii) groups of persons iii) any section or class of society. Under this provision, hate speech is being restricted; besides, there is no clear boundary regarding the circumstances under which it can be restricted.

Other statutes

1. Information Technology Act 2000,

Section 69A empowers the central government or an authorized officer to mandate the blocking of any online content that threatens national security, sovereignty and integrity, India's defense, cordial relations with other states, public order, or prevents incitement to criminal activity. Section 79 mandates that intermediaries like social media platforms eliminate content deemed to incite violence upon receiving notification from the relevant authorities under section 69A of the act.

2. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) rules, 2021

Rule 3(1)(d), which requires the intermediary to remove the unlawful content within three hours upon receiving a notice from an authorised government officer. Rule 3(2) the central government is given the power to establish the grievance appellate committee to hear the appeal cases against the order of the grievance officer. Along with this, Rule 4(1) mandates the intermediaries to appoint certain officers, such as the chief compliance officer, nodal officer, and resident grievance officer, to handle the complaints. Rule 4(2) requires the intermediaries, including the social media platforms, to identify the first originator of the information, which is considered to be an allegation upon the notice from the competent court or authority.

3. Indian Penal Code 1860 (IPC) (repealed)

Section 153A of the IPC, which criminalizes the act of promoting enmity or showing hatred

between different groups on grounds such as religion, race, place of birth, residence, language, or community.

Section 295A of the IPC, which specifically punishes acts intended to outrage the religious feelings of any group of persons in society.

4.Barathiya Nyaya Sanhita 2023

Section 196 makes it illegal to incite hostility among various groups due to race, birthplace, residence, language, or community. This segment is activated when an action is performed via electronic communication. Section 299 criminalises the act of intentionally outraging religious sentiments or beliefs. Section 353: This section is invoked against any person who circulates any information or rumours that are likely to create enmity or hatred between the different classes of groups in society.

Chapter 3-Supreme court approach towards identification of Hate Speech

Since there is no explicit line laid by the legislature between free speech and restrictive hate speech, the judiciary interpreted and laid down various tests and the things to be considered for the act to fall under the purview of restrictive hate speech.

Shreya Singhal vs union of India¹⁵

The court made three distinctions for the speech. The distinctions are discussion, advocacy, and incitement. In this case, the court held that mere discussion and advocacy were protected under the free speech. The speech that provokes the incitement is considered to be a disturbance of public order, so such speech is said to be restricted. The court in this case made a threshold for the restriction of speech as ‘incitement’.

Pravasi Bhalai Sangathan vs Union of India¹⁶

In this case justice dr. S. B Chauhan remarked that since hate speech as a close relationship with the discrimination, the impact of which is not be measured based on its abusive value alone but rather it should be seen on how it affects the marginalised people. The judiciary in

¹⁵ AIR 2015 SC 1523.

¹⁶ AIR 2014 SC 1591.

this case requested the law commission to examine about the hate speech and the need of its definition in deciding the cases related to the hate speech so that it can be recommended to the legislature for curbing the hate speech issue along with its effective implementation.

The law commission recommended to insert section 153C (Prohibiting Incitement to Hatred) and 505A (Causing fear, alarm, or provocation of violence in certain cases) in the Indian Penal Code, however it was not implemented in the Indian Penal Code as well as the new penal code, i.e., Bharatiya Nyaya Sanhita (BNS).

Amish Devgan vs union of India¹⁷

In this case the supreme court expounded that there are three elements that the legislature and the court can follow to identify and define the hate speech. firstly the content-based element which means the content of the speech has to be offensive enough to target a particular community. This can be analysed from the side of a reasonable and courageous man. Secondly the intention-based approach means the intention of the speaker must be checked whether he has the intention to target any particular community or particular class of people. Lastly, the impact based which means the speech should have the capacity to cause harm or violence in society and also affect the victim it can be either physical or mental well-being.

Ashwini Kumar Upadhaya vs Union of India¹⁸

This case did not prescribe any new threshold or expand the threshold set in the Amish Devgan case; rather, the court examined the existing legal framework governing hate speech and held that the judiciary does not have the power to create a new criminal offense or to expand the penalties under the existing legal framework due to the doctrine of separation of powers. The court also clarified that, concerning hate speech, there is no legislative vacuum; the matter is not unoccupied; instead, it is not effectively implemented.

Is the existing legal framework adequate? critical analysis

The legislative framework in India regarding online hate speech is not unoccupied, or there is no legislative vacuum; instead, there is a legal framework that exists to address the religious hate speech on social media platforms. The existing legal framework is not adequate to address

¹⁷ (2021) 3 SCC 306.

¹⁸ Writ Petition (civil) No. 943 of 2021.

the proliferation of hate speech. The inadequacies in the current legislation are explained as follows:

1. The new criminal law Bharatiya Nyaya Sanhitha (BNS) establishes a term called ‘electronic communication’ in section 196, which criminalises the act of promoting enmity between groups on the grounds of religion. This does not exist in the Indian Penal Code. This addition of electronic communication will make the act of spreading hatred on social media punishable. The inadequacy in this section is that it applies to the person who possesses the intention to do such an act. Since social media has the widest coverage of rapid dissemination of information, even the act of doing so without the intention to spread hate can cause a major problem to the individual or the group to whom it is going to be targeted, and also erodes the harmony within the religious group. So the criminalisation of hate speech under BNS will not protect persons from the spreading of hate speech, without intention, so there is a need for a framework to address even the act without the intention as well.

2. The algorithm issue in social media that needs to be addressed. The algorithm shows content that goes viral to users. Even the inflammatory content gets amplified through this algorithm used by the social media platform, which makes the information spread faster than traditional media of communication. In India, the Information Technology rules 2021 mandate the social media intermediaries to deploy themselves the automated tools to review and label the AI-generated content and also the content prohibited by any law in force in the country.¹⁹ If these tools are found to be inaccurate or biased, then the intermediary will lose its protection under section 79 of the Information Technology Act as a safe harbour immunity. Apart from this, India does not possess an effective regulation on algorithms, unlike the EU’s Digital Services Act, which requires online platforms to do an annual risk assessment²⁰, including the algorithmic systems. Additionally, they must disclose the parameters used by their platforms in providing a choice to the users on content recommendations²¹ and mandate an independent audit of algorithms²².

4. Though the Indian legal framework provides a regulatory mechanism towards the removal of unlawful content by the social media intermediary after notice from the central government

¹⁹ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (2021), rule.3

²⁰ EU’s Digital Services Act 2022, art 34.

²¹ *Ibid*, art 38.

²² *Ibid*, art 40.

or its authorised officer, here, there is no obligation on the part of the intermediary to remove the hateful content. In case of non-removal of such content without the interference of the government, no penalty was charged to the intermediary. This framework needs to be changed for the effective implementation of the removal of unlawful content on social media. This change can be seen in the German Network Enforcement act where it prescribes that the non-removal of such content on the platform within 24 hours by the social media attracts a financial penalty of up to 50 million.²³ On taking inspiration from Germany, Singapore adopted the anti-hate speech law to combat misinformation on social media. The success of providing such an obligation to the intermediary can be seen in the report submitted by Facebook to the NetzDG between January 1, 2022 to June 30, 2022, total of 22,826 contents were removed as hate speech.²⁴ Facebook did this by hiring the 1200 German-speaking moderators to review the content on the social media, which contains hate speech.²⁵ This acts as a progress due to the huge financial liability imposed upon the social media platforms.

conclusion and suggestions

As freedom of speech and expression is a fundamental right, it cannot be arbitrarily taken away. At the same time, it cannot be extended to the act of spreading hatred. Although restrictions on hate speech are not explicitly mentioned under Article 19(2), due to technological advancement and the rapid dissemination of information, there is a need to effectively regulate the amplification of religious hate speech on social media, which dehumanizes and marginalises the people in society. Though we have the existing legislation to address hate speech, which creates an inadequacy in addressing hate speech in social media concerning the dignity of the victim, algorithm amplification, and penalty for non-compliance by the intermediary. Responding to this issue, the private member bill named “online hate speech bill 2024” was proposed by the parliament by Punjab MP Vikramjit Singh Sahaney in the Rajya Sabha; however, it was not passed until now.

²³ Netzwerkdurchsetzungsgesetz (NetzDG) 2017, s.3(3)

²⁴ Facebook, “Facebook NetzDG Transparent Report” (2022).

²⁵ *Ibid.*