

Hate Speech in India: Contemporary Legal Challenges and Responses

Dipshreeya Das

Gujarat National Law University, India

*Corresponding author: ddas@gnlu.ac.in

Abstract

This article examines the legal challenges of addressing hate speech in India, focusing on the balance between free speech and public order, as well as the protection of vulnerable communities. It reviews existing legal frameworks, including constitutional provisions under Articles 19(1)(a) and 19(2), statutory laws like Sections 153A, 295A, and 505(1) of the Indian Penal Code, and Section 69A of the Information Technology Act. Key judicial pronouncements, such as *Shreya Singhal v. Union of India*, are analyzed to highlight regulatory complexities. The study identifies social media proliferation, political exploitation, and legal ambiguities as critical contemporary issues. Using a qualitative research methodology, the article proposes reforms including clearer definitions of hate speech, stronger enforcement mechanisms, and new laws targeting online hate speech and political incitement. It also advocates for enhanced public awareness and education. These measures aim to curb hate speech while upholding democratic values and ensuring societal harmony, contributing to the discourse on digital governance and legal responses in India. In light of the increased incidents of hate speech in India and the legal challenges surrounding it, this article seeks to analyze the existing frameworks and propose necessary reforms.

Keywords: hate speech, legal framework, content moderation, freedom of expression.

Introduction

Hate speech in India refers to expressions that foster intolerance, discrimination, or violence against individuals or groups based on certain characteristics such as religion, caste, ethnicity, gender, or race (Arun & Nayak, 2016) (Modh, 2015). It is a critical legal and societal issue, given the country's diverse population and the potential for such speech to incite communal tensions and disrupt public order. The legal framework to address hate speech (*Explained What Is 'Hate Speech,' and How Is It Treated in Indian Law?*, 2022) in India includes constitutional provisions, statutory laws, and judicial decisions. The Indian Constitution guarantees the right to freedom of speech and expression under Article 19(a), but also allows for reasonable restrictions based on considerations such as public order, decency, and morality, as per Article 19.

Statutory provisions relevant to hate speech are found in the Indian Penal Code (*Indian Penal Code, 1860*, 2022), including Section 153A, which penalizes the promotion of enmity between different groups; Section 295A, which punishes acts intended to outrage religious feelings; and Section 505, which relates to statements conducing to public mischief. Furthermore, Section 69A of the Information Technology Act is pertinent to online hate speech, providing



mechanisms for the government to block access to content that could compromise public order and national security.

Judicial pronouncements have also clarified the scope and limitations of these legal norms. One landmark case is *Shreya Singhal v. Union of India*, which addressed issues related to freedom of speech on the internet and the provisions of Section 66A of the Information Technology Act, which was struck down for being vague and overbroad (*Hate Speech Laws in India*, 2009).

Common forms of hate speech in India typically manifest as derogatory or inflammatory statements made against individuals or groups based on their religion, caste, ethnicity, gender, or race. Here are some prevalent forms of hate speech encountered in India:

1. Religious Intolerance: India is a multi-religious society, and hate speech often targets religious communities, with anti-Muslim sentiment being particularly widespread.
2. Caste-Based Discrimination: Speech that demeans individuals based on their caste is still common, despite legal protections against caste discrimination. This can include derogatory terms and statements that reinforce caste-related stereotypes and prejudices (*Caste-Based Discrimination in Indian Education*, 2021) (Madheswaran & Singhari, 2016)
3. Ethnic Slurs: India's ethnic diversity sometimes leads to ethnic slurs against people from various regions of the country, furthering stereotypes and division.
4. Gender-Based Hate Speech: Women often face hate speech that is misogynistic in nature, including derogatory remarks and threats of gender-based violence.
5. Race-Based Offensive Language: Certain communities, particularly those from the Northeastern states of India and those of African descent, can be targets of racial slurs and discriminatory language (*'It's Just a Joke': The Subtle Effects of Offensive Language*, 2016).

It is crucial to understand that these forms of hate speech can occur both offline and online, with the rise of social media leading to a faster spread and potentially greater impact on communal harmony and public order. Efforts to combat hate speech in India must include both legislative measures and proactive societal initiatives. It is our responsibility as citizens to challenge and confront hate speech whenever we encounter it, whether online or in person.

Despite efforts to combat hate speech in India, significant research gaps persist, hindering effective intervention. One major issue is the lack of comprehensive data, necessitating systematic collection and analysis of hate speech incidents across different regions, communities, and platforms in India. Such data would help understand the prevalence, nature, and impact of hate speech (*Hate-Speech-Reading-List*, 2023). Additionally, the existing legal framework is inadequate, as current legal provisions may not fully capture the evolving nature of hate speech, particularly in online spaces. Limited research on the effectiveness of current laws and their impact on deterring hate speech further complicates the issue (*Do Indian Courts Face A Dilemma in Interpreting Hate Speech?*, 2020) (Lepoutre, 2019). Moreover, there is a need for more in-depth studies on how socio-political dynamics, including electoral politics and ideological campaigns, influence hate speech and communal relations (Zapata & Deroy, 2023). Furthermore, the role of media and social media platforms in the dissemination and amplification of hate speech requires further investigation, particularly concerning the responsibility and



accountability of platform providers(Omar & Hashem, 2022; Schwarzmänn, 2020)(Müller & Schwarz, 2020).

Understanding these impacts is crucial for developing effective interventions (New Study Explores the Impact of Hate Crimes on Victims - Tackling Hate 2024; The Harms of Hate 2021; The consequences of online hate speech – a teenager’s perspective 2019; Paterson et al. 2018; Simpson 2012). Responding to ‘hate speech’ with positive measures: A case study from six EU countries 2018). Addressing these research gaps could lead to a better understanding of hate speech patterns, inform policy and legal reforms, and enhance initiatives aimed at combating hate speech in India.

Method

This research paper employs an extensive literature-based qualitative methodology to explore the various facets of hate speech in India. The methodology focuses on conducting a thorough thematic analysis of the existing literature to understand the complexities of the issue within the Indian context. The literature search is conducted in several stages: Identification of relevant databases and journals that include peer-reviewed articles on hate speech, social media, legal studies, and Indian socio-political issues.

Employing a combination of keywords and Boolean operators to ensure a comprehensive search. Keywords will include "hate speech," "India," "online hate speech," "communal violence," "legal framework," "social media impact," "discrimination," and "media responsibility." Inclusion and exclusion criteria will be defined to filter the results based on the relevance to the research questions, publication date to ensure timeliness, and the credibility of sources. Selection of literature that provides diverse perspectives, including theoretical frameworks, empirical studies, case law analyses, socio-political commentaries, and reports from civil societies and human rights organizations.

In conducting a thematic analysis(Lochmiller, 2021) of literature on hate speech in India, the methodology will involve a comprehensive engagement with texts to gain a deep understanding of the content (Costa, 2023). From these extensive literature reviews the overarching themes that encapsulate the key aspects of the discourse on hate speech in India are developed. These themes are then meticulously reviewed and refined to ensure they accurately reflect the data and provide insightful answers to the research questions. Each theme is clearly defined to clarify its scope and the narrative it contributes to the research context. Finally, the results of the thematic analysis are reported, synthesizing the themes with the existing literature to address the research questions and achieve the study's objectives (Vimal, 2020).

Literature Review

Hate speech is a global issue with significant social ramifications, and in India, it takes on unique socio-political dimensions due to the country's diverse cultural fabric. This literature review aims to explore the complex phenomenon of hate speech within India, highlighting its manifestation across various mediums and its impact on the society. It has become particularly relevant in light of the increasing communal tensions and the rise of digital platforms that can



amplify such speech exponentially (Narrain, 2016). The scope of this review extends to examining the legal framework governing hate speech in India, the sociocultural factors that perpetuate it, the role of media in its dissemination, and the effectiveness of policies aimed at mitigating its effects.

The main themes are:

1. **Legal Framework:** Understanding the adequacy of existing laws and their enforcement in curtailing hate speech, including a review of statutes, judicial precedents, and the role of law enforcement agencies.
2. **Sociocultural Dynamics:** Exploring how hate speech is interwoven with India's complex social and cultural tapestry, accounting for the influence of caste, religion, and ethnicity on the articulation and reception of such speech.
3. **Media Influence:** Assessing the extent to which print, broadcast, and digital media serve as platforms for hate speech, and analyzing the responsibilities and actions of these entities in preventing or enabling hate speech.
4. **Policy and Countermeasures:** Reviewing government and non-governmental strategies in addressing hate speech, ranging from censorship to education initiatives, and their implications for democracy and freedom of speech.

The rationale for this study stems from the urgent need to understand and address the contributing factors and consequences of hate speech in India, particularly in an era where digital communication has blurred jurisdictional boundaries and has magnified the potential reach and harm of hate speech (Modh, 2015). By systematically reviewing the relevant literature, this study seeks to illuminate the current landscape of hate speech in India, identify the gaps in knowledge and research, and provide a structured synthesis that can serve as a foundation for future scholarship and policy-making (Dharmapala & McAdams, 2005).

Legal Framework

The effectiveness of a country's legal framework in addressing hate speech is a critical measure of its commitment to maintaining communal harmony and protecting the rights of its citizens. In examining the literature about India's legal apparatus relevant to hate speech, several themes emerge regarding its adequacy, enforcement, and the challenges faced (*Do Indian Courts Face A Dilemma in Interpreting Hate Speech?*, 2020).

India has various laws that are used to combat hate speech, such as Sections 153A and 295A of the Indian Penal Code (*Section 295A of the Indian Penal Code - Wikipedia*, 2022), which criminalize acts that promote enmity between different groups or acts intended to outrage religious feelings. Moreover, other legal provisions like the Information Technology Act address electronic forms of hate speech. Despite these statutes, scholars and activists have raised concerns about their effectiveness and implementation. Literature often points out the vagueness of legal language, which could both lead to misuse and hinder enforcement against genuine instances of hate speech (*Hate Speech Laws in Democratic Countries*, 2021).



Judicial precedents in India also shed light on the balance the judiciary attempts to strike between freedom of speech and curbing hate speech. The Supreme Court and various High Courts have laid down judgements interpreting the constitutional limits of free speech when it comes to incitement to violence or communal disharmony. Studies analyzing these judgements often highlight the delicate precedent-based navigation between legal restrictions and democratic liberties.

The role of law enforcement agencies is also a pivotal area of study within this legal framework. The literature suggests that the efficacy of these laws is significantly influenced by the swiftness and impartiality of law enforcement. However, research also indicates a degree of reticence and bias in the enforcement of these laws, affected by political pressures and societal norms, which can favour certain groups over others.

Research gaps may include a lack of comprehensive studies on the practical enforcement of hate speech laws across diverse regions of India and a shortage of empirical data concerning the legal processes from report to prosecution. Moreover, there is a need for more scholarly work that critically evaluates the balance struck by courts between upholding free speech and sanctioning hate speech, especially in the rapidly evolving digital landscape (*How Do Prosecutors Collect and Use Data in Decisionmaking?*, 2018).

In India, key legislative provisions addressing hate speech predominantly come from the Indian Penal Code and the Representation of the People Act, along with more recent legislation that deals with electronic communication:

1. Section 153A of IPC: Section 153A penalizes 'Promoting enmity between different groups on grounds of religion, race, place of birth, residence, language, etc., and doing acts prejudicial to maintenance of harmony.' This includes acts of hate speech that might incite disharmony or feelings of enmity, hatred, or ill-will between different religious, racial, language, or regional groups or castes and communities (*Will India Criminalise Online Hate Speech? Here Are Some Answers*, 2018)
2. Section 295A of IPC: This section is for punishing 'Deliberate and malicious acts, intended to outrage religious feelings of any class by insulting its religion or religious beliefs.' It targets hate speech acts that insult or attempt to insult the religion or religious beliefs of any class.
3. Section 505 of IPC: This section punishes statements conducive to public mischief, including statements that promote or are likely to promote, on grounds of religion, race, place of birth, residence, language, caste or community, feelings of enmity, hatred, or ill-will between different religious, racial, language or regional groups or castes or communities (*Hate Speech and Incitement to Hatred or Violence*, 2023).
4. The Representation of the People Act, 1951: Under this act, Section 123(3A) deems it to be a corrupt practice if a candidate or someone on their behalf promotes or attempts to promote feelings of enmity or hatred between different classes of the citizens of India on grounds of religion, race, caste, community, or language, via public meetings, or by acts, or by articles published during election time (Zargar, 2024)
5. Information Technology Act, 2000: While not specifically addressing hate speech, the Information Technology Act of 2000, which was last amended in 2008. The IT Act provides



the legal framework for electronic governance by giving recognition to electronic records and digital signatures. It also defines cybercrimes and prescribes penalties for them.

The IT Act 2008 introduced several changes, including provisions against identity theft, cyber terrorism, and child pornography. It also introduced Section 66A, which criminalized sending offensive messages through communication services. However, this section was struck down by the Supreme Court of India in 2015 in the landmark *Shreya Singhal v. Union of India* case for being unconstitutional and a violation of the freedom of speech. The Information Technology Act of 2000, which was last amended in 2008. The IT Act provides the legal framework for electronic governance by giving recognition to electronic records and digital signatures. It also defines cybercrimes and prescribes penalties for them.

The IT Act 2008 introduced several changes, including provisions against identity theft, cyber terrorism, and child pornography. It also introduced Section 66A, which criminalized sending offensive messages through communication services. However, this section was struck down by the Supreme Court of India in 2015 in the landmark *Shreya Singhal v. Union of India* case for being unconstitutional and a violation of the freedom of speech. The Information Technology Act, 2000, although not explicitly framed as a hate speech law, contributes to the regulation of online communication and provides mechanisms to address cybercrimes, which can include hate speech disseminated through electronic channels. Since Section 66A was struck down, other provisions of the act, as well as other sections of the IPC, are used to combat hate speech online. For instance, Section 69A of the IT Act empowers the government to block public access to content in the interest of sovereignty, integrity, security of the state, and maintaining public order. This can include blocking access to online hate speech.

Challenges arise when regulating electronic communication:

- a) The vastness and anonymity of online spaces,
- b) Determining jurisdiction and applicability of laws,
- c) Balancing regulation with digital rights and freedom of speech.

Indian laws generally strive to balance the prohibition of hate speech with the protection of freedom of speech and expression, as embodied in Article 19(a) of the Indian Constitution. However, this freedom is not absolute. Article 19 allows for reasonable restrictions in the interests of sovereignty, integrity, state security, friendly relations with foreign states, public order, decency, morality, in relation to contempt of court, defamation, or incitement to an offense.

Sections 153A and 295A of the IPC are indicative of this balance (Indian Penal Code, 1860). They define and limit speech that could potentially lead to public disorder or communal strife, while attempting to respect the democratic principle of free expression (*IPC Section 153 - Wantonly Giving Provocation with Intent to Cause Riot*, 2022). Yet, the application of these laws can be contentious, and enforcing them in a manner that is fair, consistent, and in line with the principles of democracy remains a significant challenge, particularly in the digital age where the reach and impact of speech are magnified. The deployment of these laws against hate speech



requires careful and considered application to ensure that regulation does not encroach upon the essential democratic value of free speech.

The definitions provided by these sections are not exhaustive and are often subjected to interpretation by the judiciary. Legal language may sometimes be broad or vague, which has resulted in criticisms over potential misuse and challenges during enforcement. To regulate hate speech effectively, law enforcement and the judiciary must work within these.

Sociocultural Dynamics

The historical context of caste, religion, and ethnicity in India significantly contributes to both the articulation and reception of hate speech, fostering divisions that are deeply rooted in the country's social fabric. The caste system is a millennia-old structure of social stratification in India. Historically, it has distinguished individuals based on their hereditary occupations and dictated their social status, often leading to systemic discrimination against those deemed lower in the hierarchy, particularly the Dalits. This entrenched system continues to influence interpersonal and societal interactions, leading to hate speech that reinforces caste prejudices and perpetuates the marginalization and exclusion of lower caste individuals (Kumar, 2021). India has a diverse religious landscape with Hinduism being the majority religion, alongside significant populations of Muslims, Christians, Sikhs, Buddhists, and others. Historical events, such as periods of religious rule, colonialism, and partition, have left a legacy of inter-religious tensions. These dynamics often surface in hate speech, with minority religions sometimes targeted through speech that exacerbates communal divisions and raises fears of the 'other'. The wide array of ethnic and linguistic groups in India has its roots in the subcontinent's long history of migrations and conquests. This diversity, while enriching the nation's cultural landscape, has sometimes led to competing nationalisms and cultural chauvinism. Linguicism, for example, can lead to hate speech targeting those who do not conform to the dominant language or dialect in a region, impacting their socio-economic and cultural wellbeing (Kumar, 2021).

The articulation of hate speech often leverages these historical schisms to create a sense of 'us versus them', drawing on past grievances, stereotypes, and fears to fan contemporary conflicts. Conversely, the reception of hate speech is also deeply influenced by this history, as communities that have historically experienced discrimination may be more sensitive to such speech, perceiving it as a continuation of past injustices. On the other hand, dominant groups may use historical narratives to legitimize hate speech as a form of cultural or religious preservation. India's historical context of caste, religion, and ethnicity not only provides a fertile ground for the articulation of hate speech but also shapes its reception and impact on different communities (Adcock, 2014).

The historical diversity of ethnic and linguistic identities in India has had a profound impact on both the articulation and reception of hate speech, which in turn has significant implications for social cohesion in the country.

1. Articulation of Hate Speech: The rich tapestry of India's ethnic and linguistic diversity stems from a long history of migrations and conquests. However, this diversity has led to certain



regional and nationalistic sentiments that sometimes marginalize non-dominant ethnic and linguistic groups. In such contexts, hate speech (*What You Need to Know about Hate Speech*, 2023) can be a tool to assert dominance or to resist perceived cultural threats. For example, individuals or groups might use hate speech to target those who do not speak the dominant language or dialect in a particular region (Kumar, 2021), thereby attempting to protect the hegemony of their own ethnic or linguistic identity (UCU, 2023).

2. Reception of Hate Speech: The reception of this kind of hate speech varies based on one's ethnic or linguistic identity. Individuals belonging to a linguistic minority, for instance, may experience hate speech as a direct attack on their cultural and social identity. This can lead to feelings of alienation, exclusion, and fear, which are detrimental to the sense of belonging and unity necessary for a cohesive society.
3. Implications for Social Cohesion: When hate speech exacerbates existing ethnic and linguistic tensions, it threatens the fabric of social cohesion. It can hinder integrative processes and fuel segregation or the disturbing idea of 'otherness'. Such developments are often counterproductive as they impede the ability of different groups to work together, live in harmony, and contribute to a shared national narrative.
4. Linguicism and ethnic discrimination manifest as hate speech impede social integration and the overall development of the nation. They reinforce stereotypes and create an environment of mistrust and animosity, disrupting the potential for inclusive growth and mutual respect. To address these challenges, it is crucial for society to promote inclusivity, respect for diversity, and equal opportunities for all individuals regardless of their ethnic or linguistic background (Lysa, 2007).

For societies to thrive amidst diversity, there must be concerted efforts to understand, respect, and celebrate differences. Addressing hate speech actively through legal means, education, and public discourse is essential in promoting an inclusive society where every ethnic and linguistic identity can contribute to and benefit from the nation's collective progress.

Media Influence

The influence of media, encompassing print, broadcast, and digital platforms, is considerable when it comes to the dissemination of hate speech. Media channels often reflect and amplify social and political sentiments, and as such, they can either act as conduits for hate speech or as forces that challenge it.

In print and broadcast media:

Such media have historically had editorial gatekeepers who can filter content. However, editorial biases or the pursuit of sensationalism can lead to irresponsible reporting that may stoke divisions or normalize discriminatory narratives. (*Bias, Media Bias, Bias in the Media, News Bias*, 2019) Misrepresentation and lack of context in reporting can inadvertently disseminate hate speech. Responsible journalism practices, on the other hand, can counteract hate speech by refusing to provide a platform for inflammatory rhetoric and by emphasizing fact-based, respectful discourse. (Jasper, 1993)



In the realm of digital media:

The absence of traditional gatekeepers and the viral nature of content mean that hate speech can spread rapidly and widely online. Social media platforms may be used to organize and amplify hate campaigns, sometimes with algorithmic reinforcement through echo chambers and confirmation biases. The anonymity afforded by digital platforms can embolden individuals to express hate speech more freely than they might in other settings. In response to these challenges, there is a growing body of fact-checking organizations working to debunk false and inflammatory claims. Furthermore, efforts are being made to hold online service providers accountable for the content on their platforms, pressuring them to be more transparent and proactive in managing and deplatforming hate speech.

The responsibilities of media entities in preventing or enabling hate speech are profound and multifaceted: They must balance the imperative to report news honestly and openly with the need to avoid amplifying messages of hate. There is an ethical obligation to consider how reporting might impact vulnerable communities. They need to implement and enforce content moderation policies that strike a careful balance between freedom of speech and preventing harm. As purveyors of public discourse, media entities carry the weight of shaping societal norms and debates, which necessitates an approach grounded in journalistic integrity and social responsibility. (Parker & Ruths, 2023)

The extent to which media platforms serve as amplifiers or mitigators of hate speech depends on their editorial policies, content moderation practices, and willingness to engage critically with the content and its potential societal impacts. As such, the role of media in preventing or enabling hate speech is crucial and requires constant vigilance and ethical decision-making. (Müller & Schwarz, 2020) From a legal and ethical perspective, media entities have the responsibility to prevent the dissemination of hate speech while upholding the principles of freedom of expression. The nature of their responsibilities varies depending on the type of media platform due to differences in their operational dynamics and reach.

Responsibilities across media types differ due to:

1. Scale and Reach: Digital media have a larger scale and reach, which makes content moderation more challenging compared to the more controlled environments of print and broadcast media. (Pozniak, 2023)
2. Speed of Dissemination: Digital media can spread content much more quickly, which requires faster responses to potential hate speech incidents. (Gillespie, 2020)
3. Interactivity: Unlike traditional media where content is static once published or broadcast, digital media involve a constant interaction of users, where content can be shared, commented on, and modified.
4. Anonymity: Digital platforms often provide users with greater levels of anonymity than traditional media, which can embolden individuals to express hate speech.
5. Editorial Control: Traditional media have more direct editorial control over content, while digital media must manage user-generated content, which is more complex and less predictable.



Regardless of these differences, all media entities share the ethical responsibility to foster an informed and respectful public discourse and to avoid contributing to societal divisions through the dissemination of hate speech. They must constantly evaluate their content and moderation policies to effectively balance freedom of speech with the need to prevent harm. Responsibilities across media types differ due to their unique characteristics and challenges. (López, 2014)

The development of critical thinking skills, media literacy, and a deepened understanding of the impact that hate speech can have on individuals and society (Why tackle hate speech? 2021; Ubangha, 2016). These initiatives aim to equip individuals with the tools to critically evaluate the information they encounter, identify and challenge hate speech, and foster a culture of open dialogue and tolerance. The core components often include:

Media Literacy: Education programs teach how to critically assess media sources, discerning bias and manipulative narratives. This skill is crucial in an age where hate speech can quickly spread through digital media (Office, 2021).

Critical Thinking: By encouraging critical thinking, educational programs aim to help individuals question and analyze the motives behind hate speech, reducing the likelihood of its acceptance and replication.

Empathy Development: Education can also focus on developing empathy and understanding across different groups, facilitating a respectful exchange of views and reducing prejudiced attitudes that can give rise to hate speech (*Diversity and Expression*, 2017).

Historical Context: Educating about the historical consequences of hate speech can help underline the seriousness of its impact and discourage its perpetuation. (Davids, 2018)

Civic Education: Programmes that educate about rights, responsibilities, and the pillars of democracy can reinforce the value of respecting free expression while recognizing the harm hate speech can inflict. (Reichert & Print, 2017)

Contemporary legal challenges

Enforcing hate speech laws in the online domain poses a constellation of challenges due to the internet's inherent global nature. Jurisdictional issues complicate enforcement as content deemed illegal in one country might be perfectly legal in another and could be stored on servers located in a region with different legal standards on hate speech. This raises complex questions about which laws should be applied and how different jurisdictions can assert their legal authority internationally. An added layer of complexity comes from the anonymity and pseudonymity provided by online platforms, which can make it difficult to track down the individuals behind hate speech, hindering legal repercussions. The sheer quantity of content being posted online every day makes it a Herculean task for law enforcement and platform moderators to police hate speech effectively. Adding to the burden are the varying legal standards worldwide, which create an inconsistent patchwork of regulations that can be perplexing to navigate and enforce uniformly. Meanwhile, ever-advancing technologies further muddy the waters. The rise of



encryption and decentralized platforms can shield personal privacy but at the same time can obstruct the monitoring and regulation of hate speech.

The rapid and continuous evolution of online speech means that modes of communication can change faster than laws can be updated, potentially rendering existing regulations obsolete. Moreover, many existing laws lack clarity when applied to the online context as they were drafted before social media became ubiquitous, creating enforcement difficulties and grey areas. Furthermore, private companies, through their community standards and terms of service, create their own set of policies that may not necessarily align with local laws, leading to inconsistent enforcement and confusion over the boundaries of acceptable speech.

Responses

To tackle these complex challenges, international legal cooperation is crucial, which can take the form of treaties and agreements. National laws are increasingly aiming to target cross-border hate speech, and there's growing emphasis on technological solutions that enable more refined detection and reporting. Collaborative efforts between governments and online platforms are also vital to enhance accountability and transparency in how hate speech is moderated, with the goal of creating a more uniform approach to combatting this global issue without infringing on the cherished value of free speech.

Collaborative efforts between governments and online platforms to enhance accountability and transparency in the moderation of content involve navigating a range of challenges. Governments and platforms often have differing objectives, with the former concerned with public safety and compliance, and the latter focused on user engagement and privacy. By engaging in dialogue and forming partnerships, all parties can work towards aligning their goals, recognizing that a healthy digital environment benefits everyone. (*How Can Digital Public Technologies Accelerate Progress on the Sustainable Development Goals?*, 2022)

Protecting freedom of speech while regulating hate speech is a delicate balance. To avoid encroaching on free speech, it's crucial to have clear definitions of what is prohibited and to ensure transparency in moderation decisions and appeals processes. Privacy concerns also arise, as collaborative monitoring could lead to extensive data collection. Implementing robust privacy controls and independent oversight can help protect user information.

Platforms are global entities, often operating under vastly different legal regimes. Developing international legal frameworks that accommodate these differences is necessary to enable coherent action. Technology presents its own set of challenges, as automated moderation tools may erroneously identify content. Investing in better AI and integrating human reviews can improve accuracy, and shared databases can assist with content tracking. Economic considerations are also at play, as platforms may fear losing revenue with stricter content rules. Governments can offer incentives for compliance and address financial repercussions. Abuse of regulations by governments, potentially to curb dissent, is another concern that necessitates checks and balances and multi-stakeholder oversight. (Haigh, 2023)

Inconsistency in applying content policies can undermine public trust. (*Illegal Content on Online Platforms*, 2023) To counter this, standardization of enforcement procedures is needed,



complete with transparent guidelines and reporting. Addressing these challenges requires sustained conversation and collaboration among governments, civil society, and platforms Policies Transparency Centre.(Weispfenning, 1994) Crafting flexible cooperation frameworks that protect individual rights and leverage empirical research can lead to more effective management of content moderation issues.

Conclusion

In conclusion, addressing the legal challenges posed by hate speech in modern India demands a multifaceted approach that balances regulation with the preservation of free speech. The complex nature of the internet, which transcends national boundaries, alongside technological innovations and varying cultural standards, makes this task particularly daunting. Effective strategies should potentially include the harmonization of international laws, technological advancements to aid in identification and moderation of hate speech, and safeguarding the principles of democracy through transparency and protecting individual rights. Collaborative efforts between state actors, online platforms, civil society, and international bodies are essential in crafting and implementing these solutions. In recognizing the real harm that hate speech can inflict upon societies, India must navigate these contemporary legal challenges with both firmness and respect for the fundamental rights that underpin its democracy.

References

- Adcock, C. S. (2014). DEBATING CONVERSION, SILENCING CASTE: THE LIMITED SCOPE OF RELIGIOUS FREEDOM. In *Cambridge University Press* (Vol. 29, Issue 3, pp. 363–377). <https://doi.org/10.1017/jlr.2014.19>
- Arun, C., & Nayak, N. (2016). *Preliminary Findings on Online Hate Speech and the Law in India*. <https://doi.org/10.2139/ssrn.2882238>
- Bias, Media Bias, Bias in the Media, News Bias.* (2019). <https://www.imediaethics.org/topics/bias/>
- Bilewicz, M., & Soral, W. (2020). *Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization*. <https://doi.org/10.1111/pops.12670>
- Caste-based discrimination in Indian education.* (2021). https://en.wikipedia.org/wiki/Category:Caste-based_discrimination_in_Indian_education
- Costa, A. P. (2023). Qualitative Research Methods: Do digital tools open promising trends? *Revista Lusofona de Educacao*, 59(59), 67–76. Scopus. <https://doi.org/10.24140/issn.1645-7250.rle59.04>
- Dauids, N. (2018). *On the (in)tolerance of hate speech: Does it have legitimacy in a democracy?* <https://doi.org/10.1080/17449642.2018.1477036>
- Dharmapala, D., & McAdams, R. H. (2005). *Words That Kill? An Economic Model of the Influence of Speech on Behavior (with Particular Reference to Hate Speech)*. <https://doi.org/10.1086/425599>



- Diversity and Expression*. (2017). <https://freespeechcenter.universityofcalifornia.edu/programs-and-resources/resource-materials/ask-the-experts/what-steps-can-campuses-take-in-order-to-enhance-both-free-expression-and-diversity-and-inclusion/>
- Do Indian Courts Face A Dilemma in Interpreting Hate Speech?* (2020). <https://www.epw.in/engage/article/do-indian-courts-face-dilemma-interpreting-hate>
- Education and Literacy—Freedom of Expression*. (2023). <https://www.coe.int/en/web/freedom-expression/education-and-literacy>
- Explained | What is 'hate speech,' and how is it treated in Indian law?* (2022). <https://www.thehindu.com/news/national/explained-what-is-hate-speech-and-how-is-it-treated-in-indian-law/article64851585.ece>
- Gillespie, T. (2020). *Content moderation, AI, and the question of scale—Tarleton Gillespie, 2020*. <https://journals.sagepub.com/doi/10.1177/2053951720943234>
- Haigh, G. (2023). *Global Overview of Censorship Law & Policy: Mapping the Censorship Industrial Complex—Censorship Industrial Complex*. <https://censorshipindustrialcomplex.org/resources-2/globalcensorshipreport>
- Hate speech and incitement to hatred or violence*. (2023). <https://www.ohchr.org/en/special-procedures/sr-religion-or-belief/hate-speech-and-incitement-hatred-or-violence>
- Hate Speech Laws in Democratic Countries*. (2021). <https://compassjournal.org/hate-speech-laws-in-democratic-countries/>
- Hate speech laws in India*. (2009). https://en.wikipedia.org/wiki/Hate_speech_laws_in_India
- Hate-Speech-Reading-List*. (2023). <https://github.com/hate-alert/Hate-Speech-Reading-List>
- Hechiche, A. (2004). The Ethics and Culture of Peace. In *Taylor & Francis* (Vol. 29, Issue 4, pp. 495–502). <https://doi.org/10.1080/03797720500083864>
- How can digital public technologies accelerate progress on the Sustainable Development Goals?* (2022). <https://www.brookings.edu/research/how-can-digital-public-technologies-accelerate-progress-on-the-sustainable-development-goals/>
- How do prosecutors collect and use data in decisionmaking?* (2018). <https://www.urban.org/urban-wire/how-do-prosecutors-collect-and-use-data-decisionmaking>
- Illegal content on online platforms*. (2023). <https://digital-strategy.ec.europa.eu/en/policies/illegal-content-online-platforms>
- Indian Penal Code, § 375 (1860).
- Indian Penal Code, 1860*. (2022). <https://devgan.in/ipc/index.php>
- IPC Section 153—Wantonly giving provocation with intent to cause riot*. (2022). <https://devgan.in/ipc/section/153/>
- 'It's just a joke': The subtle effects of offensive language*. (2016). https://www.sydney.edu.au/news-opinion/news/2016/07/20/_it_s-just-a-joke--the-subtle-effects-of-offensive-language.html
- Jasper, K. (1993). Monitoring and responding to media messages. In *Taylor & Francis* (Vol. 1, Issue 2, pp. 109–114). <https://doi.org/10.1080/10640269308248278>



- Kumar, D. (2021). *Journey with Rural Identity and Linguicism* (Vol. 2, Issue 1, pp. 202–218). <https://doi.org/10.26812/caste.v2i1.285>
- Lepoutre, M. (2019). *HATE SPEECH LAWS: EXPRESSIVE POWER IS NOT THE ANSWER*. <https://doi.org/10.1017/s135232522000004x>
- Lochmiller, C. R. (2021). *Conducting Thematic Analysis with Qualitative Data*. <https://doi.org/10.46743/2160-3715/2021.5008>
- López, A. (2014). Greening the Mediapolis with Media Literacy. In *IGI Global* (Vol. 5, Issue 2, pp. 1–12). <https://doi.org/10.4018/ijsesd.2014040101>
- Lysa, H. (2007). Singapore and its Tensed Pasts: History and Nation-Building. In *RELX Group (Netherlands)*. <https://doi.org/10.2139/ssrn.1317164>
- Madheswaran, S., & Singhari, S. (2016). *Social exclusion and caste discrimination in public and private sectors in India: A decomposition analysis*. <https://doi.org/10.1007/s41027-017-0053-8>
- Modh, K. (2015). *Controlling Hate Speech on the Internet: The Indian Perspective*. <https://doi.org/10.2139/ssrn.2783447>
- Müller, K., & Schwarz, C. (2020). *Fanning the Flames of Hate: Social Media and Hate Crime*. <https://doi.org/10.1093/jeea/jvaa045>
- Narrain, S. (2016). *Hate Speech, Hurt Sentiment, and the (Im)Possibility of Free Speech*. <https://www.epw.in/journal/2016/17/special-articles/hate-speech-hurt-sentiment-and-impossibility-free-speech.html>
- Office, R.-G. I. P. (2021). *EU research and innovation to foster and enhance European democracy*. <https://op.europa.eu/en/publication-detail/-/publication/dc161934-771b-11eb-9ac9-01aa75ed71a1/>
- Omar, A., & Hashem, M. E. (2022). *An Evaluation of the Automatic Detection of Hate Speech in Social Media Networks*. <https://doi.org/10.14569/ijacsa.2022.0130228>
- Parker, S., & Ruths, D. (2023). *Is hate speech detection the solution the world wants?* <https://doi.org/10.1073/pnas.2209384120>
- Pozniak, H. (2023). *Tackling the impossible problem of content moderation*. <https://eandt.theiet.org/content/articles/2023/04/tackling-the-impossible-problem-of-content-moderation/>
- Reichert, F., & Print, M. (2017). *Civic participation of high school students: The effect of civic learning in school*. <https://doi.org/10.1080/00131911.2017.1316239>
- Schwarzmann, W. (2020). *Traditional Knowledge on Modern Milling Robots How CNC-joinery machines promote a renaissance to lost techniques in the profession of a carpenter*. 2, 597–604. Scopus. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85128162770&partnerID=40&md5=02df7496980636a918d0ff556ca26c33>
- Section 295A of the Indian Penal Code—Wikipedia*. (2022). https://en.wikipedia.org/wiki/Section_295A_of_the_Indian_Penal_Code
- Subba, D. (2014). *Democratic Values and Democratic Approach in Teaching: A Perspective*. <https://doi.org/10.12691/education-2-12a-6>



- UCU, E. (2023). *Define “Academic Freedom” and “Freedom of Expression.”* <https://www.ucuedinburgh.org.uk/blog/define-academic-freedom>
- Vimal, J. (2020). *Thematic Content Analysis in Literary Studies.* <https://doi.org/10.34256/irjt2118>
- Weispenning, J. T. (1994). *Television Self-Regulation: Organizational Processes and the Network Censors.* <https://doi.org/10.1177/107769909407100312>
- What you need to know about hate speech.* (2023). <https://www.unesco.org/en/countering-hate-speech/need-know>
- Will India Criminalise Online Hate Speech? Here Are Some Answers.* (2018). <https://www.thebetterindia.com/135099/india-criminalise-online-hate-speech-smriti-irani/>
- Zapata, J., & Deroy, O. (2023). *Ordinary citizens are more severe towards verbal than nonverbal hate-motivated incidents with identical consequences.* <https://doi.org/10.1038/s41598-023-33892-8>
- Zargar, H. (2024). *India elections: Vying for a third term, Modi is intensifying voter polarisation.* <https://www.middleeasteye.net/opinion/india-elections-modi-third-term-intensifying-voter-polarisation>

