Peer – Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

DISCLAIMER

No part of this publication may be reproduced or copied in any form by any means without prior written permission of Editor-in-chief of White Black Legal – The Law Journal. The Editorial Team of White Black Legal holds the copyright to all articles contributed to this publication. The views expressed in this publication are purely personal opinions of the authors and do not reflect the views of the Editorial Team of White Black Legal. Though all efforts are made to ensure the accuracy and correctness of the information published, White Black Legal shall not be responsible for any errors caused due to oversight or otherwise.

EDITORIAL TEAM

Raju Narayana Swamy (IAS) Indian Administrative Service officer

Dr. Raju Narayana Swamy popularly known as Kerala's Anti-Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala. He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University. He also has an LLM (Pro) (with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and a professional diploma in Public Procurement from the World Bank.

Dr. R. K. Upadhyay

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB, LLM degrees from Banaras Hindu University & PHD from university of Kota. He has succesfully completed UGC sponsored M.R.P for the work in the Ares of the various prisoners reforms in the state of the Rajasthan.

Senior Editor

Dr. Neha Mishra

Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; PH.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St. Louis, 2015.

Ms. Sumiti Ahuja

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,
Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing PH.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.

Dr. Navtika Singh Nautiyal

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in 'Intercountry adoption laws from Uttranchal University, Dehradun' and LLM from Indian Law Institute, New Delhi.

Dr. Rinu Saraswat

Associate Professor at School of Law, Apex University, Jaipur, M.A, LL.M, PH.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

Dr. Nitesh Saraswat

E.MBA, LL.M, PH.D, PGDSAPM
Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.
More than 25 Publications in renowned National and International Journals and has authored a Text book on CR.P.C and Juvenile Delinquency law.

Subhrajit Chanda

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); PH.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

ABOUT US

WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal provide dedicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

HATE SPEECH VERSUS FREE EXPRESSION: WHERE SHOULD THE CONSTITUTIONAL BOUNDARY BE DRAWN?

AUTHORED BY - GAGAN TOMAR[1], RACHANA S[2], CLARIE RONALD JOHN[3]

ABSTRACT

This paper seeks to analyse the conflict between the guarantee of free expression under Article 19(1)(a) of the Indian Constitution and the state's authority to impose "reasonable restrictions" under Article 19(2) in the interest of "public order," given that the definitional lacuna still exists. It is essential to reassess our ongoing reliance on outdated provisions. This research aims to find out: What new jurisprudential principles should be adopted to establish a clear and constitutional boundary between protected, albeit offensive, expression and is hate speech something that can be subject to prohibition? This paper employs a doctrinal analysis of Indian jurisprudence, along with a comparative review of contemporary American and European models, and critically engages with Mill's harm principle. It argues that the existing legal framework is constitutionally unsustainable. The broad scope of statutes such as Sections 153A and 295A of the Indian Penal Code, which have historically been upheld in cases like Ramji Lal Modi[4], creates a significant "chilling effect" on legitimate speech.

This paper asserts that the constitutional boundary must be delineated by a judicially crafted standard based on a modified harm principle, specifically tailored to India's distinct socio-legal environment. This proposed standard would interpret the restriction related to "public order" in a narrow manner, limiting the categorisation of hate speech to expressions that directly and imminently incite violence or discrimination against a protected group. This approach builds upon the principles articulated in Kedar Nath Singh and is further crystallised in Shreya Singhal v. Union of India[5]. The principal contribution of this paper lies in the creation of a new three-part test for identifying hate speech that aligns with India's constitutional commitment to both free expression and substantive equality. This framework is designed to provide a coherent and

---

[1] The author is a student at St joseph's college of law, Bengaluru.
[2] The co-author is a student at St joseph's college of law, Bengaluru.
[3] The co-author is a student at St joseph's college of law, Bengaluru.
[4] *Ramji Lal Modi v. State of Uttar Pradesh*, A.I.R. 1957 S.C. 620 (India)
[5] *Shreya Singhal v. Union of India*, (2015) 5 S.C.C. 1 (India)

rights-protective methodology for addressing the ambiguity perpetuated by judicial deference in cases such as Pravasi Bhalai Sangathan.

This paper examines the current digital ecosystem and critically analyses how the business models of social media platforms and their content moderation algorithms actively promote hateful content to increase engagement and profits. This section explores the effectiveness and constitutional implications of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[6], questioning whether these rules provide a solid solution or simply establish a new framework for state censorship. The research argues that any proposed legal standards must consider the unique role of platforms as facilitators of hate.

1. INTRODUCTION

One of the most crucial rights in a democracy is the freedom of speech and expression. This right is protected under Article 19(1)(a)[7] The Indian Constitution, permits people to freely express their thoughts, feelings, and convictions. This freedom is restricted, though. In some circumstances, such as when it impacts public order, the State may impose "reasonable restrictions" on free expression under Article 19(2)[8]. As a result, upholding public peace and defending individual rights are constantly at odds.

Determining the boundary between hate speech and free expression is one of the most difficult tasks in this field. Old colonial clauses like Section 153A and Section 295A of the Indian Penal Code serve as the foundation for the majority of India's present hate speech regulations. Due to their ambiguity and breadth, these laws have frequently been interpreted differently by judges. The judiciary has attempted to strike a balance between free expression and public order in cases such as *Pravasi Bhalai Sangathan v. Union of India*, *Kedar Nath Singh v. State of Bihar.[9]*, and *Ramji Lal Modi v. State of Uttar Pradesh*, but a clear legal norm is still lacking. Uncertainty results, and many individuals are afraid to talk openly, which has a "chilling effect."

---

[6] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (India).
[7] INDIA CONST. art. 19, cl. 1, subcl. (a).
[8] INDIA CONST. art. 19, cl. 2.
[9] Indian Penal Code, No. 45 of 1860, § 153A, Acts of Parliament, 1860 (India).

In the digital age, this problem has gotten even worse. Social media platforms facilitate the rapid dissemination of communication, especially offensive material that has the potential to cause actual harm. This issue has become even more serious in the digital age. Social media platforms allow speech to spread quickly, including hateful content that can incite real harm. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, were introduced to regulate online spaces, but they raise serious concerns about censorship and misuse of power.

This research aims to find a clear and fair legal standard that protects freedom of expression while allowing the State to act against real hate speech. It draws on Indian constitutional law, judicial decisions, and the harm principle developed by John Stuart Mill.[10]. It also looks at how other countries, like the United States and those in Europe, deal with similar issues. The goal is to propose a judicially crafted test that can help India balance free speech with public order in a way that is both clear and constitutionally sound.

2. RESEARCH PROBLEM

Indian courts are inconsistent when dealing with hate speech. Because they don't apply a single, clear standard, there is legal confusion about the line between protected free speech and illegal hate speech.

3. RESEARCH OBJECTIVE

- **Analyse the Constitutional Conflict:** To critically analyse the conflict between the fundamental right to free expression (Article 19(1)(a)) and the State's power to impose reasonable restrictions for public order (Article 19(2)).

- **Diagnose and Reform the Legal Framework:** To assert that the current, outdated legal provisions (like 153A[11] and 295A IPC[12]) are unsustainable and to establish new jurisprudential principles that define a clear constitutional boundary between protected, offensive expression and prohibitable hate speech.

- **Propose a New Three-Part Test:** To formulate a novel, judicially crafted three-part test for identifying hate speech. This test is based on a modified **Harm Principle**,

---

[10] Mill, J.S. (1978) *On Liberty*. 20th ed. Indianapolis: Hackett Publishing, pp. 9–15.
[11] Indian Penal Code, 1860, 153A.
[12] Indian Penal Code, 1860, 295A.

adapted to India's socio-legal context, and is designed to promote both free expression and substantive equality.

- **Advocate for Narrower Restrictions:** To advocate for a narrow interpretation of the "public order" restriction, limiting its application only to speech that directly and imminently **incites violence or discrimination**, while ensuring the proposed standards are applicable and effective in regulating contemporary social media platforms.

4. RESEARCH HYPOTHESIS

The Indian judiciary's reliance on antiquated colonial provisions for hate speech regulation undermines constitutional free expression, necessitating reinterpretation in light of India's democratic principles.

5. RESEARCH METHODOLOGY

- Doctrinal Analysis of Indian Law: The central method is a doctrinal analysis focusing on Indian constitutional provisions, statutes, and relevant case law.
- Comparative Legal Review: This analysis is enriched by a comparative review of contemporary American and European legal models.
- Critical Theoretical Engagement: The research critically engages with legal and philosophical theories, specifically by adapting John Stuart Mill's harm principle for the Indian context.
- Digital Ecosystem Examination: The study includes a thorough examination of the current digital ecosystem,

6. RESEARCH QUESTIONS

1. How should the scope of reasonable restrictions under Article 19(2), particularly in relation to public order, be reinterpreted in light of the chilling effect caused by broad statutory provisions and judgments like Ramji Lal Modi and Pravasi Bhalai Sangathan?

2. Can a judicial framework built on Mill's harm principle, refined through Indian precedents such as Shreya Singhal v. Union of India, and enriched by American and European models, provide a constitutionally consistent boundary between offensive speech and hate speech?

3. In what ways do social media platforms and their algorithm-driven business models exacerbate hate speech, and how should constitutional and regulatory standards evolve to address this challenge while safeguarding free expression?

7. THE TENDENCY DOCTRINE: AN ANALYSIS OF RAMJI LAL MODI AND ITS CONSTITUTIONAL VALIDATION OF THE CHILLING EFFECT.

The judiciary's first significant interpretation of the new "public order" clause clearly established its flexible nature. In the landmark case of Ramji Lal Modi v. State of U.P. (1957), a five-judge bench decisively upheld the constitutionality of a specific provision in the Indian Penal Code (IPC), which makes it illegal to provoke religious feelings intentionally and maliciously. The Court made two important decisions that negatively impact free expression. First, it interpreted the First Amendment's phrase "in the interests of public order as having a very broad meaning. This interpretation is wider than "for the maintenance of" public order, allowing laws to be valid even if they aren't specifically aimed at maintaining order. Secondly, the Court introduced the "calculated tendency" test. The petitioner argued that Section 295A was overly broad, as it could penalise speech that merely offended emotions without leading to public disorder. The Court rejected this argument, stating that the state does not need to prove that speech caused disorder; it only needs to demonstrate that the speech tended to do so. Consequently, any speech deemed malicious could result in punishment. This "chilling effect" is not just a theoretical idea; it is a direct procedural outcome of the Modi doctrine. The vague "tendency" test provides constitutional cover for overly broad statutes like Section 295A[13] And Section 153A, which addresses promoting enmity. Because these are cognizable offences, the law is readily weaponised at the point of initiation. Any individual can file a complaint, and police, guided by a low standard of "tendency," can make arrests. The "chilling effect" describes this punitive process, which silences speech long before a court has the chance to assess actual harm.

I. The Doctrinal Counter-Narrative: The Rise of the Proximity and Imminence Standards

The Modi doctrine established a standard that favours the state, but a conflicting approach developed in the Supreme Court. This created two different views on free speech law. This counter-narrative started with the 1965 case of Dr Ram Manohar Lohia v. State of Bihar[14],

---

[13] Indian Penal Code (1860) s. 295A.
[14] *Dr. Ram Manohar Lohia v. State of Bihar*, AIR 1966 SC 740

where the Court moved away from the "tendency" test. It decided that the connection between restricted speech and the need for "public order" must be "close" and "direct," not "remote" or "far-fetched." This "proximity" test corrected the Modi "tendency" test, which was less strict. In S. Rangarajan v. P. Jagjivan Ram (1989), the Court reinforced this speech-protective standard. It introduced the "spark in a powder keg" test, stating that the anticipated danger must not be remote or unsure. Instead, it should be closely linked to the speech. The speech must pose a real threat to the public interest, like a "spark in a powder keg." This standard contradicts the Modi standard, as the "tendency" test allows for restrictions based on distant possibilities of harm, while the Rangarajan test requires Immediate and certain

II. The Pravasi Bhalai Sangathan Enigma: A Diagnostic Success, A Doctrinal Failure

The 2014 case, Pravasi Bhalai Sangathan v. Union of India[15], highlighted the dangers of hate speech, rightly identifying it as an attack on the dignity and equality of groups. However, the Court failed to resolve the legal conflict. Instead of addressing the flawed Modi standard that allows the misuse of hate speech laws, it pointed to issues with enforcement and referred the matter to the Law Commission. This was problematic because the Court effectively called for stricter enforcement of existing laws that follow the flawed Modi doctrine. It recognised the problem but suggested more of the same approach, not fixing the core issue that leads to a "chilling effect" on speech.

III. Conclusion: A Proposed Framework for Reinterpretation

The "chilling effect" in this analysis is not just a theory; it results directly from the Supreme Court's 1957 decision in Ramji Lal Modi v. State of UP. By upholding Section 295A of the Indian Penal Code (IPC), the five-judge bench set a low standard for judging speech, known as the "calculated tendency" test. This test does not require proof that speech has caused or is likely to cause public disorder. It only requires that the speech has the potential to cause such disorder, like the simple act of insulting a religion. This "tendency" standard is the main reason for the chilling effect. It allows vague laws like Sections 153A and 295A to be misused. Since these laws are cognizable offences, anyone can file a complaint, and the police can arrest based on the low threshold set by the Modi test. The resulting legal process can take years, often serving as punishment itself. This misuse of the law to target legitimate speech makes people fear reprisal, leading to self-censorship. These standard hurts free speech and are not sound

---

[15] *S. Rangarajan v. P. Jagjivan Ram*, (1989) 2 SCC 574

legal thinking. It goes against the speech-protective rules the Court has developed. This alternative view began with the "proximity" test in Dr Ram Manohar Lohia and fully emerged in the "spark in a powder keg" test from S. Rangarajan. This perspective demands clear evidence and immediate danger. The court confirmed this standard in Shreya Singhal v. Union of India (2015).

In Shreya Singhal, the Court clearly separated "discussion" and "advocacy" from "incitement." It ruled that Article 19(2) allows restrictions only on incitement. This aligns with the Brandenburg v. Ohio.[16] Standard, which states that speech can only be limited when it is (1) aimed at inciting immediate illegal action and (2) is likely to cause that action.

This situation creates a serious conflict in legal precedent. The two-judge bench in Shreya Singhal could not officially overrule the five-judge bench in Modi. The Shreya Singhal decision tried to resolve this by claiming that Modi should only apply to true incitement. However, this view is logically and legally flawed. "Calculated tendency," which suggests a small chance of harm, is the opposite of "incitement to imminence," which indicates a high chance of immediate harm. Because of this confusion, while the legal standard has advanced, police and lower courts can still use the outdated Modi test, continuing the chilling effect on free speech.

Therefore, a reinterpretation of Article 19(2) to resolve this conflict is imperative. This requires two clear and definitive steps:

1. **Formally Overrule *Ramji Lal Modi***: A Supreme Court bench of sufficient strength (a seven-judge bench, as it would need to overrule the five-judge *Modi* bench) must formally declare that the "calculated tendency" test from *Ramji Lal Modi v. State of U.P.* is incompatible with Article 19(1)(a). The Court must hold that this test is unconstitutionally vague and "over-broad" precisely because it fails to distinguish between protected advocacy and unprotected incitement.

2. **Consolidate the *Shreya Singhal* Standard**: The Court must then hold that *all* legislative restrictions on speech grounded in "public order" or "incitement to an offence" under Article 19(2) must be tested *only* against the high-threshold *Shreya Singhal* standard: *incitement to imminent lawless action*. This would make the "spark in a powder keg" a binding and universal test, not just one of two conflicting options.

---

[16] *Brandenburg v. Ohio*, 395 U.S. 444 (1969)

This reinterpretation would, by necessity, force a re-evaluation of the constitutionality of Sections 153A and 295A of the IPC. As currently worded, these are "tendency" based statutes. They punish acts like "promoting enmity" or "outraging religious feelings" —which are acts of advocacy—without requiring any proof that these acts are *likely* to incite *imminent* violence. They are, therefore, "over-broad" under the *Shreya Singhal* test and would likely be struck down or read down significantly.

This framework would not legitimise hate speech. On the contrary, it would distinguish genuine threats from mere dissent. It would end the systemic "chilling effect" of procedural harassment and compel Parliament to draft new, narrow, and precise laws that are aimed at *actual, imminent incitement to violence and discrimination*. This is the only way to finally align India's statutory framework with its own highest constitutional standards and resolve a conflict that has, for over 60 years, undermined the foundation of free expression in the republic.

Introduction

Free speech is one of the core values of India's democracy. The constitution of India places limits when speech that threatens public order or harms others by allowing the state to control such actions. The difficulty is that these limits are not applied consistently, which sometimes leads to a chilling effect where people hesitate to speak openly in public. Because of this tension between social harmony and liberty, it is important to find a clear and principled way to separate offensive speech from hate speech, which can genuinely harm society.

Mill's harm principle [17]is a liberty-limiting principle that draws boundaries of governmental or societal interference with individual freedom. The principle states that a state can rightfully exercise its power over any individual or a member of a civilised community against their will only when his/her actions cause harm to others. Any Action that impacts or harms others can warrant intervention. Speech can only be restricted to prevent actual harm, but not mere offence. The key distinction is that offensive speech is merely disagreeable or insulting, while hate speech attacks or targets a person or a group based on identity factors such as race, religion, or gender, which includes threat or incitement to harm. Hate speech is a communication that uses discriminatory language that promotes hatred and violence, while offensive speech is General disagreement or insult. Offensive speech is generally protected by law. Hate speech is

---

[17] JOHN STUART MILL, ON LIBERTY 14–15 (Elizabeth Rapaport ed., Hackett Publ'g Co. 1978) (1859).

also protected unless it gives way to violence or brings in threat to society.

Why does the issue matter?

Speech should be free unless it causes real harm. This helps to draw boundaries crucial for maintaining social order, preventing violence and protecting vulnerable groups from targeted abuse.

To balance between free speech and public order[18] The law allows restrictions on speech that incites violence or hatred, which leads to a chilling effect where fear of legal repercussions causes individuals to stop speaking openly. These laws can be misused to silence criticism. Protecting public order is very much essential, but it should not be used to suppress controversial or uncomfortable ideas, and the real focus should be on encouraging counter speech and public debate.

Indian Law and Practice

The Supreme Court[19] Interprets Article 19(1)(a) as a fundamental right essential to democracy, individual autonomy, and intellectual progress, while also ruling that restrictions under Article 19(2) must be **exhaustive, reasonable, and directly related to the specific grounds listed in Article 19(2)**. This means the government cannot add new reasons for restricting free speech, and any laws imposing restrictions must pass the test of reasonableness and be in the interest of public order, security, defamation, contempt of court, or morality, among other grounds.

In the Tehseen Poonawalla judgement (2018)[20] The Supreme Court brings in a principle that Hate speech is not a fundamental right and can be restricted to uphold public order, dignity and the rule of law.

Keywords: Free speech, Constitutional limits, Public order, Chilling effect, Offensive vs hate speech, Social harmony, Democratic values.

The supreme court brings in various tests on hate speech though its judicial precedents where

---

[18] S. Rangarajan v. P. Jagjivan Ram, (1989) 2 SCC 574
[19] India Const.art.19(1)(a), art. 19(2).
[20] Tehseen S. Poonawalla v. Union of India, (2018) 9 SCC 501.

in **Shreya Singhal vs union of india[21]** the test of Proximate and direct nexus emerges and the test of ''Incitement''. the Supreme court held that The Supreme Court held that any restriction on free speech, to be considered "reasonable" under Article 19(2) of the Constitution, must have a direct, proximate, and imminent connection to public disorder or other grounds mentioned in Article 19(2). Laws that are vague, overbroad, or have a remote or fanciful connection to a potential harm (like "annoyance" or "inconvenience") are unconstitutional because they criminalise a vast amount of innocent and protected speech, leading to a "chilling effect" on free expression. The test of "incitement" distinguishes between three types of speech: discussion, advocacy, and incitement[22]. The Court held that mere discussion or even advocacy of a particular cause, no matter how unpopular, is protected free speech. It is only when speech reaches the level of **incitement,** which tends to cause imminent public disorder or other constitutionally valid restrictions, that Article 19(2) applies and the speech can be curtailed.

In the case of Ram Lal Modi, the test "Deliberate and Malicious intent "emerged, where the Supreme Court ruled that only acts done with a **deliberate and malicious intention** to outrage religious feelings, which have a "calculated tendency" to disrupt public order, fall outside the protection of free speech.

Comparative models

In *Brandenburg v. Ohio* (1969)[23]The U.S. Supreme Court held that speech can only be punished when it is directed to inciting imminent lawless action and is likely to produce such action.

This sets a high threshold, protecting even offensive speech unless it triggers immediate violence. In the US, Hate speech is mostly protected protected in the U.S. because the First Amendment's guarantee of freedom of speech is broadly interpreted to protect even offensive ideas. Hate speech is protected unless it directly incites violence and crime. In the US, it is difficult to define hate speech. Laws against hate speech are challenged because it is difficult to define "hate speech" in a way that is not overly broad or vague, which would risk chilling protected speech.

---

[21] Shreya Singhal v. Union of India, (2015) 5 SCC 1 (striking down vague speech restrictions for creating a chilling effect on free expression)
[22] Id. at 22–24 (distinguishing discussion, advocacy, and incitement).
[23] *Ramji Lal Modi v. State of Uttar Pradesh*, AIR 1957 SC 620.

The European court of human rights[24] Allows for restrictions on expression if they are prescribed by law, which must have a clear legal basis. The ECHR involves several key rules and considerations. The core principle is that it recognises freedom of expression as vital and protects hate speech when it attacks the dignity and safety-based identity. The ECHR balances the right to free expression with the need to protect people from harm, but Article 10(2) allows the government to limit this freedom when speech threatens public order or the rights of others. The Court strongly protects open debate, even when opinions are unpopular or offensive.

The judicial framework built on Mill's Harm Principle and the refined Indian precedents like *Shreya Singhal,* which are guided by American and European models, can provide a constitutionally consistent boundary between offensive and hate speech.

Such a framework ensures that restrictions under Article 19(2) are reasonable, precise, and proportionate, protecting the spirit of free expression while preventing genuine harm to individuals and society. In essence, freedom must prevail until speech turns into harm, not merely when it offends.

Keywords: Real harm, Public order, Vulnerable groups, Counter-speech, Legal restrictions, Misuse of laws, Free expression vs security, Liberty-limiting principle, Harm vs offence, Government interference.

Social media platforms and their algorithm-driven business models

Hate speech on social media is basically any kind of communication through offensive language, violent messages, images, videos, symbols, and memes that is said to attack an individual or group of people based on their race, gender, religion, ethnicity, disability, or sexual orientation. The United Nations to addressed the issue of Hate speech globally, defined Hate speech through the UN Strategy and Plan of Action on Hate Speech[25], defined Hate speech is defined 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." Social media platforms exacerbate hate speech as their core business

---

[24] European Convention on Human Rights art. 10(2)
[25] United Nations, *United Nations Strategy & Plan of Action on Hate Speech* (2019).

model is built to reward what captures the attention of the public. Engagement-based algorithms are designed in a way to maximise likes, shares and comments, which leads to sensational, controversial, and hate content over accuracy.[26].

The algorithm presents users with more emotionally charged content when they interact with it, perpetuating a cycle that polarises narratives and strengthens extreme viewpoints. Although social media platforms did not set out to be hate speech engines, the system they created for financial gain ended up being ideal amplifiers for it. Anything that keeps users clicking, scrolling, and responding is rewarded by their algorithms. Rather than legitimacy or public health, their business models rely on views. The foundation of the issue lies in the engagement-based algorithms; the system is trained to identify posts that are most likely to receive the greatest number of comments, shares, and emotional responses. The objective is straightforward: maximise engagement on the platform. The system continues to push these posts once it recognises that they maintain user interaction. Whether the content is harmful to the public or beneficial is irrelevant to the algorithm. It only evaluates the responses and awards prizes to the finest performers.[27].

The business model of the media is strongly linked to the incentive system. The majority of significant social media businesses make money through advertising. They can show more commercials and earn greater revenues as long as they can keep people engaged. Because of this, there is a clear financial gain to promote anything that keeps people engaged, regardless of whether the information incites rage or terror. The algorithm starts to favour hate content as it typically performs better than the normal content. Outrage generates interaction, interaction gets revenue, and revenue pushes the platform to adjust its mechanisms to produce more of the same. An additional dimension of risk is by way of recommendation engines. Platforms tend to lead users through a path that encourages more hate over time by recommending "related" content. If someone starts with a mainstream political video, they may quickly receive recommendations for content that is more divisive, provocative, or conspiratorial, relating to the particular political video. The recommendation engine determines which posts are most likely to keep visitors interested and clicks "optimise." The algorithm will favour the provocative video if the data suggests viewers of a moderate video often continue to watch a

---

[26] European Comm'n, *Assessment of the Code of Practice on Disinformation* (2020).
[27] Frances Haugen, *The Facebook Files: Testimony Before the U.S. Senate Comm. on Commerce* (2021).

stronger, more controversial one. The said dynamic should be balanced by moderation mechanisms; however, this is rarely the case. Because of the size at which platforms operate, consistent enforcement is quite challenging. Every hour, millions of posts are published, and automatic moderation technologies frequently overlook coded language, irony, and context. Human moderators are frequently unsupported and overworked. Even worse, when businesses erase material, there is a clear financial consequence. Each post that is removed reduces the amount of content that may generate interaction and display advertisements. Because of this, moderation happens to be more generally reactive than preventive. Usually, platforms only take action when an issue attracts public notice or when advertisers share their concerns. By then, the harmful content would have already multiplied, shared, and recirculated throughout multiple platforms.

Opacity is another significant problem; the majority of social media businesses do not reveal their algorithms, content guidelines, or moderation processes. They keep their systems private from the public, regulators, and researchers, viewing them like trade secrets. It is hard to track how hate content spreads or why certain comments are taken down while others are left up due to a lack of transparency.[28]. In addition, it makes it difficult for outsiders to comprehend the direct relation between algorithmic behaviour and profit incentives. It is easy to escape accountability when the system operates like a black box. The platform may argue that the algorithm is too complicated to explain if hate speech grows in number. The business blames user behaviour rather than its own design judgments. Anything that keeps consumers interested is rewarded by the advertising model. Users are directed toward more extreme content by recommendation systems. Moderation stays inconsistent and often too slow to be significant. It is challenging to hold the platform responsible due to the lack of transparency.

CONCLUSION & SUGGESTIONS

The constitutional tension between the right to freedom of speech and expression, as outlined in Article 19(1)(a), and the regulation of hate speech under Article 19(2) is one of the most contentious issues in Indian democracy. This research has revealed that the current legal framework reflects a significant lack of coherence in its jurisprudence. While the Supreme Court has gradually embraced a more liberal standard of "imminent incitement" in

---

[28] U.N. Hum. Rts. Council, *Report on Disinformation and Freedom of Expression*, U.N. Doc. A/HRC/47/25 (2021).

contemporary rulings, such as *Shreya Singhal v. Union of India* and *S. Rangarajan*, the enduring influence of the 1957 *Ramji Lal Modi* "calculated tendency" test continues to cast a repressive shadow over civil liberties.

The persistence of the "tendency" doctrine enables the State to criminalize speech based on speculative and remote possibilities of public disorder rather than concrete threats. This low threshold creates a systemic "chilling effect," whereby the mere potential for offending religious or communal sentiments—regardless of intent or tangible outcome—can result in criminal prosecution under Sections 153A and 295A of the Indian Penal Code (now retained as Sections 196 and 299 of the Bharatiya Nyaya Sanhita, 2023).

The findings of this research indicate that this statutory vagueness, in conjunction with the algorithmic amplification of outrage facilitated by social media platforms, has led to a "Heckler's Veto," where the intolerance of the mob dictates the boundaries of permissible speech.

Ultimately, the preservation of India's constitutional morality requires a decisive shift from the colonial logic of "subjective hurt" to the democratic logic of "objective harm." A democracy cannot survive if it silences all offensive speech to maintain a superficial order; equally, it cannot thrive if it permits dehumanising hate speech to dismantle the dignity of its citizens. The boundary must be drawn not at the point of offence, but at the precipice of incitement to violence and discrimination.

Suggestions for Reform: To resolve this doctrinal conflict and address the challenges of the digital age, this paper proposes the following structural interventions:

- **Legislative Reform - Decoupling and Definition:**
    - *Codification of "Hate Speech":* The Bharatiya Nyaya Sanhita (BNS) must be amended to introduce a specific statutory definition of Hate Speech that focuses on **harm to dignity** rather than **hurt to sentiment**. The standard must be strictly defined as an expression that advocates incitement to imminent violence, discrimination, or hostility against a protected group, distinct from blasphemy or insult.
    - *The "Safe Harbour" Proviso:* A specific exception should be added to BNS Sections 196 and 299 protecting academic, artistic, and political speech made

in good faith. This would ensure that historical critique, satire, or dissent are not prosecuted as "promoting enmity," codifying the distinction between *advocacy* and *incitement*.

- **Judicial Reform - The "Seven-Judge" Correction:**
  - *Overruling Ramji Lal Modi:* The Supreme Court must constitute a seven-judge Constitution Bench to formally review the *Ramji Lal Modi* (1957) judgment. The Bench should declare the "calculated tendency" test unconstitutionally vague and explicitly adopt the **"Spark in a Powder Keg"** (proximity and imminence) test as the universal standard for restricting speech under Article 19(2).
  - *Harmonising Precedents:* The Judiciary must authoritatively rule that the high threshold for incitement established in *Shreya Singhal* (for online speech) applies equally to offences under the Penal Code, removing the current dichotomy where online speech is theoretically more protected than offline speech.

- **Procedural Safeguards:**
  - *Mandatory Preliminary Enquiry:* To curb the weaponisation of the legal process, the Supreme Court should mandate a "Preliminary Enquiry" (PE) by an officer of the rank of ACP/DSP before registering any FIR for hate speech.
  - *Strict Mens Rea:* Courts must strictly enforce the requirement of "malicious intent." Proof of a specific intent to incite violence must be a prerequisite for arrest, preventing the detention of individuals merely for sharing offensive content without malicious motive.

- **Digital and Societal Measures:**
  - *Algorithmic Accountability:* Regulation must move beyond simple content takedowns. The Information Technology Rules should be amended to mandate "circuit breakers" during times of communal tension—throttling the *virality* of content rather than deleting the content itself—to allow time for fact-checking and counter-speech.
  - *Media Literacy:* Finally, the long-term antidote to hate speech is resilience. A national "Digital Citizenship" curriculum should be introduced to train citizens to distinguish between "hate speech" (incitement) and "offensive speech" (dissent), reducing the social volatility that fuels the demand for censorship.