

A Study of Regulation of Hate Speech in India: Constitutional Boundaries, Criminal Liability, And the Impact of Digital Technology

Shaurya Jaiswal

Amity Law school

Amity university, Uttar Pradesh

Abstract - The study investigates the regulation of hate speech in India, focusing on its constitutional basis, criminal law provisions, digital aspects, theoretical underpinnings, and comparative approaches. The core premise of this work is that hate speech presents a significant challenge within constitutional democracies, necessitating a careful balance between upholding freedom of speech and safeguarding public order, individual dignity, and equality. Constitutionally, the study investigates the parameters of free speech and expression as enshrined in Article 19(1)(a) of the Indian Constitution, considered alongside the permissible limitations outlined in Article 19(2). This involves a critical assessment of how free speech jurisprudence has evolved judicially, particularly concerning the differentiation between advocacy and incitement, the notion of public order, and the principle of proportionality. The analysis reveals that while the judiciary has endeavored to establish guiding principles, a definitive and consistently applied doctrinal standard remains elusive, leading to ambiguity regarding the precise threshold for punishable speech. Turning to the criminal law framework, specifically under the Bharatiya Nyaya Sanhita, 2023, the study indicates that hate speech is managed through various broadly formulated provisions targeting the promotion of enmity, religious insult, and threats to national cohesion. This statutory language, employing terms such as “hatred,” “enmity,” and “ill-will,” is inherently indeterminate, which the paper argues permits considerable subjective interpretation. This ambiguity consequently fosters both excessive criminalization and inconsistent enforcement. Although judicial rulings have sought to clarify these matters by focusing on intent and context, they have not entirely resolved the inherent ambiguities, thus perpetuating doctrinal inconsistencies. The central argument of this study is that India's legal framework for regulating hate speech is characterized by structural fragmentation and conceptual incoherence. This is primarily attributed to the lack of a clearly articulated and narrowly defined threshold for restricting speech. Consequently, the study advocates that substantive reform would necessitate implementing an incitement-based standard rooted in the principle of

proportionality, alongside enhancing the precision of statutory language and establishing more robust procedural safeguards to mitigate potential misuse. The study concludes by proposing a set of legal and institutional reforms aimed at developing a more coherent, consistent, and constitutionally compliant framework for the regulation of hate speech. It emphasizes that the objective of regulation should not be the suppression of offensive or unpopular expression, but the prevention of speech that causes real and demonstrable harm. Ultimately, the study underscores the need for a balanced approach that preserves the fundamental right to freedom of expression while addressing the challenges posed by hate speech in an increasingly complex and digital society.

I. INTRODUCTION

Hate speech has emerged as a particularly difficult and contentious issue in modern constitutional and criminal jurisprudence, more so in societies that are socially, religiously and culturally diverse. In India, where plurality is a cornerstone of the constitutional order, the restriction of speech aimed at or disparaging specific groups presents an especially tough legal challenge. The core problem is not only to identify damaging expression but also to regulate it in a manner that respects the fundamental right to freedom of speech and expression, a sine qua none of democratic governance. The restriction on speech attacking or denigrating certain populations is a particularly hard legal task in India because plurality is a keystone of the constitutional order. The central challenge is not just identifying harmful expression but regulating it in a way that respects the fundamental right to freedom of speech and expression, the bedrock of democratic governance. The constitutional framework of the Indian Constitution tries to reconcile these conflicting interests. Article 19(1)(a) guarantees freedom of speech and expression to all persons, which is indispensable for a healthy exchange of ideas, the

formation of public opinion, and the functioning of democratic institutions. However, Article 19(2) grants the State the power to impose reasonable restrictions on this freedom, especially in the interest of public order, morality, decency, and other specified reasons. This bifurcated framework recognises that while free speech is vital, its exercise is not absolute, particularly when it endangers social harmony or violates the rights of others. Article 19(1)(a) guarantees to all people the right to freedom of speech and expression which conduces to a lively exchange of ideas, the development of public opinion and the smooth functioning of democratic institutions. At the same time, Article 19(2) allows the State to impose reasonable restrictions on this freedom, especially in the interest of public order, morality, decency and other specified grounds. This divided framework acknowledges that free speech is important, but its use is not without limits, especially when it threatens social harmony or impinges on the rights of others.

1.1 Statement of Problem

The prohibition of hate speech in India continues to be characterised by doctrinal vagueness and practical inconsistency, even in the face of statutory provisions and constitutional safeguards. The lack of a coherent and unambiguous definition of hate speech has led to a disjointed legal environment in which many statutes operate in parallel without a common criterion for determining culpability. This vagueness inevitably renders it difficult to distinguish between speech that is lawfully subject to limitation and speech that is protected under the constitution. The absence of a clear and consistent definition of hate speech has led to a fragmented legal landscape where numerous laws operate simultaneously without a common standard for determining guilt. This ambiguity cannot but complicate the task of distinguishing speech subject to limitation from speech constitutionally protected.

By creating concepts like the distinction between advocacy and incitement and by highlighting the necessity of a close relationship between speech and public disorder, judicial interpretation has attempted to close this definitional gap. However, there are different thresholds for banning identical types of communication because these rules have not been consistently applied rigorously across cases. Such

inconsistency weakens the law's predictability and, consequently, reduces its overall efficacy.

Judicial interpretation has sought to fill this gap in the definition through the development of concepts such as the distinction between advocacy and incitement and the need for a close nexus between speech and public disorder. Yet, identical types of communication are prohibited at different thresholds because the rules have not been applied in a rigorous and consistent manner across cases. Such inconsistency detracts from the predictability of the law and thus also its overall efficacy. But there are different standards for banning the same kind of communication, as these rules have not been strictly and consistently applied from case to case. Such inconsistency undermines the predictability of the law and therefore its general effectiveness.

1.2 Research Questions

There are number of important legal issues are raised by hate speech law in India. The following are the research question of this study-

1. What is the scope and inherent nature of the constitutional framework governing freedom of speech in India, and how does this framework apply specifically to hate speech?
2. To what extent do existing criminal law provisions effectively regulate hate speech, and what are their limitations?
3. How has the emergence and proliferation of digital technology fundamentally transformed the nature and impact of hate speech?
4. What specific reforms are necessary to develop a more coherent, consistent, and constitutionally sound framework for regulating hate speech in India?

1.3 Objectives of the Study

The paper aims to conduct an analysis of the legal framework which are governing hate speech in India. The main goal is to critically analyse the complex interactions between judicial interpretations, statutory regulations, and constitutional principles in the restriction of speech that has the potential to cause social harm.

This study analyses the constitutional foundations of free speech, particularly as defined by Article 19(1)(a) of the Indian Constitution, and looks at the acceptable range of limitations outlined in Article 19(2), with an emphasis on how they relate to hate speech. Evaluating the impact of criminal law provisions—particularly those found in the *Bharatiya Nyaya Sanhita, 2023*—in reducing different types of damaging expression is another goal. The study also looks into how digital technology affects hate speech's nature and dissemination, as well as how well-suited the current legal systems are to these changes in technology.

The paper also aims to contextualise the Indian regulatory framework by contrasting it with hate speech policies implemented in other countries. The purpose of this comparison analysis is to find substitute models and extract possible lessons relevant to the Indian setting. In the conclusion, this study aims to offer suggestions for changes that would promote a more unified, uniform, and constitutionally sound framework for hate speech prohibition in India.

1.4 Research Methodology

As for the methods utilized in this study, the approach is predominantly doctrinal legal research with the emphasis on a comprehensive critical analysis of the legal literature and authoritative materials related to the problem at hand.

The investigation has been based on the meticulous analysis of statutory laws, case laws, and constitutional law provisions relevant to the issue of regulating hate speech in India.

The Constitution of India, the BNS, 2023, BNSS, 2023, and IT Act, 2000, together with all supplementary legislation, are primary sources. Besides, judgments by the Supreme Court of India and other high courts have been considered as they reveal the development of hate speech laws in the past and their present-day application.

Besides primary materials, secondary sources, for instance, monographs, articles from academic journals, and reports like that of the Law Commission.

1.5 Scope and Limitations

The scope of the present paper pertains exclusively to hate speech regulation in India through a discussion of digital governance, criminal liability, and constitutional law. Attention is paid to the complex relationship between legislation and fundamental rights as well as challenges arising due to technological changes in speech dissemination.

Despite the introduction of a comparative approach into the paper, the examination of foreign legal systems will be conducted solely as an illustration of the issues at stake and not as a broad comparison of the global regulation of hate speech.

It should be emphasized that the main approach to the examination of the issues under consideration remains strictly legal and doctrinal. This study refrains from conducting a sociological or empirical evaluation of hate speech, which precludes the application of statistical methods to assess the prevalence of hate speech or measure its impact. Finally, owing to the rapid development of technology, some recent advances in this field may lie beyond the scope of the current discussion.

II. CONSTITUTIONAL FRAMEWORK OF FREE SPEECH IN INDIA

Article 19(1)(a), which ensures freedom of speech and expression in the constitution, exists within a context that includes a history of nationalist struggle and colonisation. Understanding the present legal framework against hate speech in India, including the inherent tension in the exercise of the right to freedom due to the conflicting nature of such actions, will only be possible if one is aware of the development of the said provision within this context. Freedom of speech and expression was greatly curtailed during the era of colonisation through laws enacted by colonial rulers in their effort to silence the opposition to British rule. There were many laws enacted which served to curtail the use of political speech in order to suppress any nationalist efforts and specifically any acts related to sedition or any form of restrictions to the press. The colonial government had developed a tradition of controlling speech as an act that might harm their control rather than an intrinsic right.

2.1 Historical Evolution of Freedom of Speech in India

In *Romesh Thappar v. State of Madras*¹, the Supreme Court invalidated a governmental ban on a journal holding that freedom of speech and expression lies at the foundation of all democratic organizations, and that restrictions on speech must be narrowly construed and justified within the constitutional framework. Similarly in *Brij Bhushan v. State of Delhi*², where the Court struck down pre-censorship measures imposed on a newspaper, and prior restraint on speech is presumed unconstitutional, all these early decisions showed a commitment by the judiciary to defend free expression against interference by the state.

However, this led to concerns within the political elite about the State's ability to enforce public order in a newly independent and socially pluralistic country, leading to the Constitution (First Amendment) Act, 1951, which significantly broadened the scope of Article 19(2) by introducing more grounds of restriction such as public order and incitement to an offence thus marking a critical turning point in the constitutional history of free speech in India that tilted the balance towards greater regulatory powers. Judicial interpretation of free speech took an incremental approach with the goal of accommodating individual liberty and the State's interest in order and stability. In *Kedar Nath Singh v. State of Bihar*³, the Supreme Court upheld the constitutionality of sedition laws, but restricted their application to acts involving incitement to violence/public disorder, thus, creating a critical doctrinal safeguard based on the presence of a palpable threat.

Over time, the courts attempted to distinguish between simple promotion of ideas and incitement to illegal conduct by using this approach in subsequent instances.

The evolution of free speech jurisprudence in India thus reflects a continuous balancing of competing concerns, a firm constitutional protection of expression as a democratic value on the one hand and the concern about the disruptive potential of speech to public order and harmony in a deeply sensitive society.

This historical trajectory proves relevant in the context of hate speech given that the broadening of the restrictions under Article 19(2) in addition to the emphasis of the public order by the judiciary has

resulted in the establishment of a framework upon which harmful expression can be regulated without an identifiable threshold, thereby rendering the legal position a matter of contextual interpretation with no certainty in its application.

2.2 Meaning and scope of Article 19(1)(a)

Article 19(1)(a) gives to all citizens freedom of speech and expression. From the wording of the provision, it may seem a meagre right but its expanse and contours have been generously extended by the Supreme Court to include a large number of activities of communication thus converting it into one of the liveliest and fundamental rights in our constitution. Essentially the freedom of speech and expression means that you can state your views in any medium – be it verbal, written, printed, visual or even symbolic. The judicial interpretation has given it a further meaning. It means the protection is given to more than just speaking or writing. It includes the circulation of information as well as the receiving of information. The right of speech and expression encompasses both, because what is important about free speech is the right of the speaker but also the right of the listener to be heard.

Over the course of time, numerous peripheral rights have been incorporated within Article 19(1)(a). The Supreme Court has extended Article 19(1)(a) to cover the liberty of the press, which is not explicitly guaranteed by the Constitution, as part of freedom of expression. It was confirmed by the Supreme Court in the case of *Indian Express Newspapers v Union of India*⁴ held that freedom of the press is an integral part of Article 19(1)(a), and that any state action—whether direct or indirect—which has the effect of restricting the circulation of newspapers would fall within the ambit of this right, though such restrictions may still be justified under Article 19(2).

The right to receive information has also been recognized as a crucial component of the right of free speech. The information and transparency cases have become the venue for this realization where the court tied the right to know with the right of the people to enjoy democratic rights. In this way, the court brought another element under Article 19(1)(a) by making people aware of the fact that democracy depends on an informed people.

¹ *Romesh Thappar v. State of Madras*, AIR 1950 SC 124.

² *Brij Bhushan v. State of Delhi*, AIR 1950 SC 129.

³ *Kedar Nath Singh v. State of Bihar*, AIR 1962 SC 955.

⁴ *Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India*, (1985) 1 SCC 641.

A third significant dimension of the right is that it extends to both symbolic and non-verbal forms of expression. The Court has stated that "expression covers all manifestations of human thought and ideas." This includes gestures, art, etc., and ensures the right evolves as the mediums through which thoughts are expressed develop and include newer forms, such as those that exist in the online space.

Notwithstanding its broad scope, Article 19(1)(a) has to be viewed as one that has its limitations. It is a right that exists within a constitutional framework which allows it to be restricted under Article 19(2). The scope of the right in practice therefore varies greatly depending on the limits applied to Article 19(1)(a). This has necessitated a balance between expression and harm, in relation to the adverse social consequences which speech might generate.

This broad scope in the context of hate speech generates a fundamental tension. Firstly, its breadth implies that minority, hateful, unpopular and even dissenting views may be protected. Secondly, it therefore means that speech aimed at minorities, which may indeed incite hatred, also falls within the purview of protected expression, raising questions regarding the boundary between protected expression and harm, which will form the focus of the constitutional limitations.

2.3 Judicial Expansion of Free Speech Rights

Judicial interpretation has considerably broadened the scope of Article 19(1)(a), with the Supreme Court opting for an interpretation that is expansive and purposive regarding freedom of speech. The Court has not limited the scope of the right to the literal meaning; instead it has constantly interpreted the right in terms of its democratic purpose and therefore expanded the protection to a whole range of activities, indispensable for the free functioning of expression.

One of the most fundamental aspects of this judicial expansion has been the recognition of free speech as not merely individual expression but also as an aspect of the communication ecology in which these ideas are formulated, disseminated and received. This principle was reflected in the Supreme Court's decision in *Secretary, Ministry of Information and Broadcasting v Cricket Association of Bengal*⁵ where the Court observed that right to freedom of speech

comprises the right to communicate information via the electronic media. By acknowledging the importance of access to communication, the Supreme Court went beyond a restrictive concept of speech as mere utterance to a concept of participatory right of expression in a democracy.

The judiciary has also given particular importance to the Press in relation to free expression. Freedom of Press has never been separately enumerated by the Constitution, though it has been consistently read into Article 19(1)(a). In *Bennett Coleman & Co. V Union of India*⁶ the government imposed limitations on newsprint which the court struck down by observing that limitation on the circulation of a newspaper is necessarily a limitation on freedom of speech and expression. The court therefore clearly acknowledged that indirect restrictions (i.e. Restricting access to means of communication) also limit freedom of speech.

Another crucial element added by judicial interpretation has been the explicit recognition of the right to receive information as an inherent facet of freedom of speech. Time and again the Supreme Court has stated that expression has value not just for the speaker, but for the audience too, as only in possession of varied viewpoints can they reach informed decisions. In a democracy, informed decisions are what make it functioning and therefore the expansion of the right has greatly aided in the development of right to information. The idea that a transparency is inherent in a freedom was established.

The judiciary also extended the ambits of Article 19(1)(a) to artistic, cinematic and commercial means of expression. In *S. Rangarajan v P. Jagjivan Ram*⁷, the Supreme Court held that freedom of expression cannot be fettered because there is an apprehension of demonstration or processions or threats of violence. The Court further stated it is the responsibility of the State to maintain law and order and not curtail freedom of speech for anticipatory disturbances. This is all the more significant for controversial or unpopular speech which is likely to provoke extreme responses.

This interpretation also went onto encompassing within Article 19(1)(a) the commercial form of speech in *Tata Press Ltd. V Mahanagar Telephone Nigam Ltd.*⁸, the Court finally declared that

⁵ *Secretary, Ministry of Information & Broadcasting v. Cricket Association of Bengal*, (1995) 2 SCC 161.

⁶ *Bennett Coleman & Co. v. Union of India*, (1972) 2 SCC 788.

⁷ *S. Rangarajan v. P. Jagjivan Ram*, (1989) 2 SCC 574.

⁸ *Tata Press Ltd. v. Mahanagar Telephone Nigam Ltd.*, (1995) 5 SCC 139.

advertising has the nature of commercial speech and was a form of expression guaranteed under the constitution. Commercial speech has a greater scope for regulation compared to the others but it was still regarded as expression within Article 19(1)(a). Modern methods of communication too have been held within the ambits of Article 19(1)(a). While striking down Section 66A of Information Technology Act in *Shreya Singhal v Union of India*⁹, the Supreme Court stated that the right of speech equally extended to online communication. As the internet became the most essential means for exchange of thoughts, it was obvious that constitutional protections had to extend to it but while declaring this the judgment also came up with a new distinction between advocacy and incitement which is very important in the sphere of regulation of harmful speech.

2.4 Reasonable Restrictions under Article 19(2)

Though the scope of free speech and expression guaranteed under Article 19(1)(a) is very wide and fluid, the restrictions imposed by Article 19(2) will control and limit the ambit within which it can be exercised. The restrictions under Article 19(2) are the constitutional tool in the hands of the state by virtue of which speech can be curtailed for the protection of some overriding interests in the society namely, public order, decency, morality and the security of the state and integrity of India. Any restriction on speech would thus have to meet the twin requirement of coming within one of the permitted categories under Article 19(2) and of being reasonable¹⁰.

The concept of "reasonable restrictions" under Article 19(2) has been central to judicial interpretation, requiring that restrictions must not be arbitrary or excessive. The Supreme Court has consistently held that the validity of a restriction must be assessed with reference to both its objective and its impact. In *Chintaman Rao v. State of Madhya Pradesh*¹¹, the Court emphasized that a restriction must have a rational nexus with the purpose it seeks to achieve and must not go beyond what is necessary for that purpose. This approach laid the foundation for the later development of the principle of proportionality in constitutional adjudication.

The grounds mentioned under Article 19(2) have changed in extent, after the Constitution (First

Amendment) Act, 1951 introduced the new heads "public order" and "incitement to an offense". It represented the desire of the State to maintain law and order, in an emergent India. It simultaneously made speech less protected and thus potentially opened the state to a wider regulation on speech matters.

Among the various heads mentioned under Article 19(2), the head public order stands special to the topic of hate speech, as the apex court has distinguished it from "law and order" and "security of the state" as any restriction can be justified only by actual and imminent threat of a disruption of public order and not by remote fear of future discontents. This notion of proximity thus acts as a fundamental safeguard.

The relationship between speech and public disorder was further clarified in *S. Rangarajan v. P. Jagjivan Ram*¹², where the Supreme Court held that the anticipated danger arising from speech must be proximate and not remote or speculative. The Court stressed that limitations cannot be put in place only because the expression might incite animosity or disagreement. It further noted that the state must uphold public order by guaranteeing the protection of free expression rather than by stifling it. In the context of hate speech, where worries about the "heckler's veto" frequently surface, this principle takes on special significance.

Another ground that falls under Article 19(2) and relates to the current context is decency and morality where restriction is based on public conscience and can be applied to restrict speech on grounds of obscenity and offense etc. Although not directly related to the topic of hate speech, the ground of decency and morality is however related to a social interest that might at times come in conflict with the protection afforded to expression.

The ground of "incitement to an offense" also has significant place in the discussion, for this is where, a demarcation can be made for distinguishing protected from unprotected speech which is directly related to the concept in *Shreya Singhal v Union of India* that speech can be punished only when it reaches the point of incitement not advocacy. Here the court has restricted the ambit of the permitted head by providing a nexus and relation between speech and unlawful action.

2.5 Public Order as a Ground of Restriction

⁹ *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.

¹⁰ Constitution of India, art. 19(2).

¹¹ *Chintaman Rao v. State of Madhya Pradesh*, AIR 1951 SC 118.

¹² *S. Rangarajan v. P. Jagjivan Ram*, (1989) 2 SCC 574.

Of the stated limitations on freedom of speech under Article 19(2), 'public order' holds a central role, particularly in a diverse society like ours and a history fraught with conflict. This is because many restrictions placed upon speech, such as hate speech, have been sustained under the ground of public order as such speech has the potential to disturb public tranquillity or cause disorder.

The Supreme Court has consistently emphasized that "public order" is distinct from "law and order" and "security of the State." In *Dr. Ram Manohar Lohia v. State of Bihar*¹³, the Court explained that while every breach of public order may affect law and order, not every breach of law and order amounts to a disturbance of public order. It created a graded distinction among them, stressing that "public order" represents a narrower and more serious form of disruption that affects society as a whole. This is particularly significant in the context of regulating speech since any restriction based on public order needs to have a definite connection with disrupting social harmony and not just local harmony.

It has become a fundamental condition to demonstrate the existence of a direct and immediate nexus between the speech or writing and the perceived disruption. Any restrictions that depend on vague possibilities of disturbance cannot be justified according to the Supreme Court's decisions. In *Superintendent, Central Prison v. Dr. Ram Manohar Lohia*¹⁴, the Court emphasized that there must be a close and reasonable connection between the speech and the public disorder sought to be prevented. This requirement ensures that restrictions are not imposed on the basis of uncertain or future contingencies, thereby preventing misuse of the public order ground to suppress lawful, controversial, or unpopular expression.

This is again complicated when the expression is targeted at a community or religion. Such speech could have the ability to inflame and instigate or create such social friction that can lead to disorders, even if violence is not immediately apparent. One is faced with the question whether "imminent disorder" can sufficiently meet such speech and in many cases there has been a liberal interpretation in favour of imminent possibility rather than actuality due to the sensitive social climate. However, the strict and

proximate nexus between speech and public disorder has been a constant feature.

Further, in *S. Rangarajan v. P. Jagjivan Ram*¹⁵ it was held that even an expression can't be restricted on the pretext that it may not be liked and would elicit adverse reactions as 'danger should not be conjectural and far-fetched'. The court highlighted the responsibility of the State to maintain law and order not suppress liberty as there may be a breach of it. The whole idea of the doctrine is that it guards the law against a 'heckler's veto'.

However, even with regard to the doctrine of proximity and public order, there has been mixed application especially as far as hate speech is concerned. Since hate speech, directed at religious, cultural, ethnic groups often involves sensitivity on account of the social environment it is placed in and can quickly escalate tensions and instigate violence, authorities often rely on proactive restrictions. This proactive restriction may be justified but dilutes the proximate nexus between the speech and actual disorder.

Another angle to public order is that it also envisages a framework of social peace in India and such a perspective seems to underlie the notion of various speech offenses under the Indian penal code. Therefore, as stated, restrictions placed on speech due to concerns of public order would also include attempts to foster and maintain social harmony in the community. However, it is the question of interpretation that would remain critical as anything other than such a judicious restriction could pose a threat to free speech itself.

2.6 Doctrine of Proportionality

The mandate that the restrictions upon speech must be "reasonable" under Article 19(2) has, over the course of time, become substantively framed with the principle of proportionality. Proportionality acts as a standard for assessing State action that limits fundamental rights like the freedom of speech and expression. In addition to determining whether the State has a legitimate purpose to fulfil, it also necessitates that the means of achieving that purpose is proportionate to the right being interfered with.

The original formulation of reasonableness, as reflected in *Chintaman Rao v. State of Madhya Pradesh*¹⁶, required that State interference must not

¹³ *Dr. Ram Manohar Lohia v. State of Bihar*, AIR 1966 SC 740.

¹⁴ *Superintendent, Central Prison v. Dr. Ram Manohar Lohia*, AIR 1960 SC 633.

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

¹⁶ *Chintaman Rao v. State of Madhya Pradesh*, AIR 1951 SC 118.

be arbitrary or excessive and must bear a rational nexus with the objective sought to be achieved. Over time, this standard evolved into a more structured doctrine of proportionality. In *Modern Dental College & Research Centre v. State of Madhya Pradesh*, the Supreme Court articulated a four-pronged proportionality test requiring that (i) the measure must pursue a legitimate aim; (ii) it must be suitable to achieve that aim; (iii) it must be necessary, in that no less restrictive alternative is available; and (iv) there must be a proper balance between the extent of restriction and the importance of the objective.

This concept gained more clarity in the case of *Modern Dental College v State of Madhya Pradesh*¹⁷, where the Court accepted the proportionality test. The decision clarified that State action restricting fundamental rights must also satisfy the test of proportionality and thus is not solely dependent on the legitimacy of the end in question. It has thus moved away from a purely reverential attitude and introduced more rigorous judicial scrutiny over the State action.

In the context of freedom of speech and expression, proportionality is critical when the restrictions imposed fall under a permissible ground such as public order or incitement to offence. This principle demands that the regulatory measure must be confined and designed to avoid undue restraint on freedom of speech. Therefore, any legislation imposed restricting the ambit of Article 19(1)(a) must not go beyond the necessities of what is required for prevention of such harm.

In *Shreya Singhal v Union of India*, although not expressly formulated in terms of the proportionality test, the judgment demonstrates the underlying rationale of proportionality. Section 66A of the Information Technology Act was struck down for being vague and overly broad, which in essence implies that the said restriction was not necessary in a democratic society and that a less restrictive means was available to punish cybercrimes. It established the principle that the laws regulating speech must be specific and have narrow parameters.

Proportionality is of paramount importance in the sphere of regulating hate speech. Considering the comprehensive nature of Article 19(1)(a), it can be inferred that the limitation imposed upon such speech has to be specifically limited in scope and must only target those elements of speech which poses a

substantial danger or harm in a certain context and are proportional in their consequences. Overbroad regulations even with a laudable intent might result in the stifling of legitimate discourse, whilst narrow and carefully drafted limitations can more readily meet constitutional scrutiny.

While providing a necessary check upon government power, proportionality is not free of challenges when applied in cases such as hate speech. It often requires subjective considerations with regards to context, audience and impact of speech in question, which may not always be easily quantifiable. Such challenges introduce an element of subjectivity and the test of proportionality may not lead to similar outcomes in different cases, creating inconsistencies and ambiguities.

2.7 Conflict Between Free Speech and Dignity

Article 19(1)(a) protects freedom of speech and expression, and it is not considered in isolation. It is protected concurrently with other fundamental rights, the most obvious being the right to life and personal liberty under Article 21. Through a gradual process of judicial development, Article 21 was widened to include the protection of human dignity and this has created competing constitutional values that conflict with the right to free speech and expression. The control of hate speech exists at the confluence of the right to freedom of speech and the value of human dignity.

The doctrine of human dignity as a constitutional value emerged through judicial pronouncements on the intrinsic value of a human being. In *Francis Coralie Mullin v Administrator, Union Territory of Delhi*¹⁸, Supreme Court held that the right to life extends to human dignity, along with its contents. This broadens the scope of protection offered by the constitutional value, moving beyond just the bare fact of human existence to incorporate all necessary conditions for human dignity, self-respect and personhood.

When applied to the realm of speech, human dignity as a constitutional value becomes crucial when the speech targets specific individuals or group on basis of their identity-be it religious, caste or ethnic. This is so, as hate speech often involves a form of expression which seeks to demean, stigmatize and even dehumanize targets on the basis of who they are. It creates a constitutional tension between freedom of

¹⁷ *Modern Dental College & Research Centre v. State of Madhya Pradesh*, (2016) 7 SCC 353.

¹⁸ *Francis Coralie Mullin v. Administrator, Union Territory of Delhi*, (1981) 1 SCC 608.

speech and expression under Article 19(1)(a) which protects the right to speak and the dignity right under Article 21 that seeks to protect those who may be subject to it.

The debate and interpretation by the courts to deal with the collision between freedom of speech and dignity has not provided us with an established or universally accepted test and it seems that we have adopted a case-by-case balancing exercise approach. In *Subramanian Swamy v Union of India*¹⁹, Supreme Court upheld the constitutionality of defamation while highlighting that freedom of speech cannot be exercised in a way that damages another's reputation and dignity; the constitution value of dignity has been deemed as a relevant goal justifying restrictions on free speech.

This means that a shift away from a solely speech-centric approach to one that incorporates other rights to be balanced, has been observed. However the infusion of human dignity within speech regulation leads to a greater subjectivism in the sphere. The meaning of dignity's violation can vary, in terms of contexts, perception and societal understanding as compared to clearer harm-inflicting notions such as incitement to violence.

Adding complexity to the debate is the question of group-based harm to human dignity. Hate speech can inflict harm on an entire collective community, yet, constitution law and doctrines have generally developed from an individualist perspective, with greater ease of fitting concepts like liberty with individual liberty rather than collective freedom or dignity for a collective entity. This is something that existing legal theories appear to be lacking.

Simultaneously, expanding human dignity as a ground for restricting speech may lead to excessive censorship where even speech that might be critical of prevailing practices, sarcasm or humour is targeted simply on the ground that it offends dignity interests of certain sections of society. Hence, the challenge lies in striking the right balance so that the restriction imposed based on human dignity is not abused to stifle valid expression.

The clash between Articles 19 and 21 in speech and expression is but one aspect of the broader constitutional challenge of finding equilibrium between individual autonomy and the protection of others from the harm that speech can cause. In the absence of an explicit doctrinal framework to address

this problem, a degree of judicial discretion has permeated judicial outcomes on hate speech.

2.8 Constitutional Issues in Hate Speech Regulation
These interrelated issues arise out of the tension in Indian constitutional law that has arisen as a result of broad guarantees of free expression and broad grounds for restriction. The problems are not simply matters of interpretation, but of a deeper structural imbalance which exists due to a lack of a clearly defined standard for identifying and prohibiting hate speech.

Firstly, there is no constitutional or statutory definition of hate speech. The grounds under Article 19(2) explain the occasions on which speech can be restricted, but not the class of speech that falls under it; this is treated indirectly through concepts of public order, incitement and morality, leading to a difficult task of drawing the line at which speech can transition from merely protected speech to criminal act and again leaving significant scope for judicial interpretation.

This second problem of undefined hate speech relates to the broad grounds of restriction provided in Article 19(2). Vague notions like 'public order', 'decency' and 'morality' allow for varied interpretation according to the needs of time and place, making consistent judicial decisions extremely hard, especially when the severity of the expression involved can vary greatly.

The inconsistent application of doctrinal tests is inescapably linked with this question. Although it is important to differentiate incitement from advocacy, this criterion has not been equally applied; some have relied solely on the incitement test, whereas others argue that social peace and disturbance constitute valid grounds for restriction.

Another challenging area of constitutional interpretation is the relationship between free expression and dignity. It has already been demonstrated how Article 21's guarantee of dignity protects certain aspects of speech, but it is challenging to establish precise guidelines that determine when speech violates dignity, causing ambiguity and the potential for subjectivity in the evaluation of such harm.

Another issue with the proportionality theory is that, despite providing a structural mechanism to balance the right against the restriction, the criteria of need

¹⁹ *Subramanian Swamy v. Union of India*, (2016) 7 SCC 221.

and less restrictive measures appear to rely on a mainly subjective evaluation of the probable impact of speech.

III. CRIMINAL REGULATION OF HATE SPEECH IN INDIA

The law in India does not define criminal hate speech in a unitary and single statutory provision. Instead, various offenses like instigating enmity, insulting religion and threatening national integrity address different types of harmful expression as criminal hate speech in a discrete manner. This lack of single statutory definition leads to the identification of criminal hate speech as a sum of provisions not a definition.

Criminal hate speech deals with a collective form of harm. Unlike defamation and obscenity that concerns individual injury or collective standards of morality, hate speech affects specified sections or groups of society. It prohibits and penalizes expression intended or likely to sow animosity or ill-will between religious groups or any other sections of society in a religiously diverse and complex social setup like India. This feature distinguishing it from speech offenses that protect individuals or collective morals.

3.1 Concept and Nature of Criminal Hate Speech

The concept of crime in hate speech offenses also concerns intention and context. Whether a statement is made with the intent to sow hatred, or the statement would indeed generate hatred in the audience are relevant factors while identifying the speech as criminal hate speech and such evaluation necessarily brings into question the speaker, the audience and the social background of the society at the time of uttering such speech. These elements again require interpretative efforts from the court and thus further enhance ambiguity.

Moreover, criminal hate speech represents an approach of a preventive rather than curative nature; thereby, speech may be penalized as an action that, with a fair degree of certainty, could result in crime or social disorder. This characteristic of criminal law appears more relevant in a society where historical and social issues influence perceptions. However, such approach also increases scope for overreach and can penalize speech that has not as of yet resulted in any real harm.

Unlike constitutional law's attempt to protect expression until it crosses the stated boundary,

criminal law on the other hand has a far lower threshold for intervention, where such speech may cause disorder in the social fabric. The tension is thus generated where a given statement might attract different legal treatment under the constitutional framework than it may under the criminal law framework.

In a nutshell, absence of a well-defined concept of criminal hate speech therefore creates a double dilemma; of flexibility for the law to meet various circumstances and ambiguity for the people to determine where lies the border of criminality, and in the Indian case it defines the character of criminal hate speech as an interpretative notion rather than a strictly defined category.

3.2 Overview of the Bharatiya Nyaya Sanhita, 2023
The Bharatiya Nyaya Sanhita, 2023 (BNS), which supersedes the Indian Penal Code, 1860 while preserving the fundamental framework of speech-related offences, maintains the statutory restriction of hate speech in India. Notably, laws like Section 196 BNS (which corresponds to Section 153A IPC) make it illegal to encourage animosity between various groups on the basis of religion, race, place of birth, domicile, language, or community, as well as to take actions that are detrimental to the preservation of harmony. In a similar vein, intentional and malicious acts meant to offend religious sentiments are punishable under Section 299 BNS (equivalent to Section 295A IPC). These rules continue to be the primary legal means of dealing with hate speech, despite being inherently ambiguous in relying on terms such as "enmity", "hatred" and "outrage". The fundamental problems of interpretive ambiguity and discretionary enforcement, therefore, continue to be embedded in the statutory framework in spite of the IPC having been substituted for the BNS.

Offences relating to the concept of hate speech under the BNS mainly pertain to the acts that are prejudicial to national unity, the acts that inflame hatred, and the acts that wound religious sensibilities. In a country that is multilingual, multireligious, and multiregional, the objective of these rules is presumably to prohibit hate speeches and ensure that there are no attempts made at fostering enmity. Through making such utterances illegal, it would seem that the legislative intent was to maintain harmony in society.

This statutory framework's focus on the form and substance of an utterance is a defining characteristic as the statute criminalizes an action if it carries the

risk of causing disturbance or disorder. Consequently, this makes any attempt to foster enmity a punishable offence even if it is conveyed through sign, graphics or any other medium of expression.

Another pertinent aspect of the statutory framework is the usage of wide and undefined words like 'hatred', 'enmity' and 'ill-will'. These terms lend elasticity and flexibility in the penalization of diverse kinds of harmful expression but at the same time they introduce vagueness and leave significant discretion in the hands of the law enforcement agencies as well as in the interpretative function of courts. The lack of specific definition of these terms implies a dependence on ad hoc decision making process.

3.3 Key Penal Provisions: A Detailed Section-wise Analysis

Criminal regulation of hate speech in the Bharatiya Nyaya Sanhita, 2023 operates through a multiplicity of provisions designed to target different types of expressive conduct that can be detrimental to the social fabric. The BNS does not provide for a single omnibus offence of hate speech. Instead it utilizes a network of provisions to target specific harms caused by such expression. This section deals with the major relevant provisions of hate speech, exploring their scope and requirements, along with the interpretational concerns raised.

3.3.1 Promotion of Enmity Between Groups

One of the most prominent of the hate speech related provisions criminalises attempts and acts done with the intent to promote enmity between different groups on the ground of religion, race, place of birth, residence, language, caste or community. The provision further makes criminal speech which promotes, or attempts to promote feelings of enmity, hatred or ill-will between groups, or that is prejudicial to the maintenance of harmony between them.

The breadth of the said provision lies in its emphasis on tendency rather than demanding actual violence or disturbance of the public peace. Words that have the potential to divide the society and foster hostility fall within its purview, the Law intervening at an early stage, in situations where inter-communal tensions can spiral quickly.

However the use of terms like hatred and ill-will introduces uncertainty about the nature of words to be considered criminal. These terms don't possess clearly defined meanings and will always be context-dependant, a position in which judiciary has played a significant role in delineating the boundaries of the

offence of promoting enmity by looking at the intentionality behind such expressions.

There also exists mens rea (criminal intent) in this section. Generally, the onus would be on the prosecution to prove the accused had intention to promote enmity or at least knowledge of the likely impact of his expression. This forms a protection against any incidental and unintentional consequences of an expression being criminalised, although its practical application has varied.

3.3.2 Acts Prejudicial to Maintenance of Harmony

Overlapping with the provision for promoting enmity, the provisions in the BNS also deals with the crime of statements that are prejudicial to the maintenance of harmony. It targets actions which do not directly incites hatred, yet can potentially harm harmonious relationships between different communities.

The focus of the provision here is broader: the impact on social stability. Words causing fear or suspicion between groups could fall under this provision, even though the same may not directly express malice or hatred. This aspect makes it quite crucial in cases where subliminal messages or words carrying implications of hostility are conveyed through indirect language.

However, a consequence of its vagueness is its extensive ambit. Such provisions are also difficult to distinguish, as the potential harm can be interpreted as 'public mischief'. This increases the risk of its use against speech that is critical or unorthodox but does not actively incite hatred and cause actual public disruption. Overlap also arises in distinguishing with the provision for promoting enmity.

3.3.3 Outraging Religious Feelings

Thirdly, The BNS criminalises acts that insult the religion or religious beliefs. It provides that deliberately and maliciously insulting any class of citizens or their religious beliefs shall be punishable. Here, deliberately and maliciously are the keywords. Judicial interpretation suggests that such a provision is aimed to punish words of offense rather than words challenging the religion (which may be considered as critique or reform), particularly as religion holds great significance for a part of the society.

However the interpretation has often proven a balancing act between protecting the sensibilities of the religious followers on one hand, and avoiding restriction on critical and dissenting speech on other, which may be misconstrued as intended to cause insult and provocation, thereby raising concerns of

misuse for discouraging critical discussion. The presence of mens rea element is vital in such circumstances.

3.3.4 Statements Conducing to Public Mischief

Lastly, there are provisions concerning public mischief which include the publication of false news capable of inducing fear or panic. This section can be invoked where speech can incite fear or create conflicts within communities without being considered hate speech per se. In such cases, the danger with the speech is its ability to induce fear and create conflicts within societies, leading to hatred for certain groups.

The problem with this clause is its broadness, since it can cover any form of speech that creates public mischief, regardless of intent. The difficulty with establishing "the intention to cause public mischief" becomes more pronounced considering the modern world of instant information exchange through the Internet.

3.3.5 Observations on the Statutory Framework

Some of the key characteristics of the framework include tendencies towards general offenses, a focus on the negative impact of the speech, and the inclusion of mens rea and context. Nevertheless, one definite flaw in the BNS regulations concerning hate speech is that these regulations are both vague and overlapping, making them seem like a fragmented approach rather than something that can be considered a code. It should be noted that the court will have to construe these provisions in relation to the Constitution.

3.4 Role of Mens Rea in Hate Speech Offences

As per the Bharatiya Nyaya Sanhita, 2023, the existence of a guilty mind (mens rea) is just as significant for criminal liability in hate speech as the nature of the utterance itself. An essential protection within the criminal justice process that ensures one is not entirely liable for any unforeseen results of the expression is the requirement of mens rea.

Terms such as "intent," "knowledge," and "probability" could be incorporated in the substance of the sections addressing the promotion of enmity and other acts prejudicial to harmonious relations. This is because in addition to establishing the fact that the utterance is likely to instigate animosity, the prosecution must prove that the utterer intends to instigate animosity or knows that incitement is probable. The implication of these terms is that

expressions can be understood differently in a pluralistic environment. Therefore, only persons who know that their utterances may lead to enmity and animosity can be made liable.

Mens rea is an integral part of differentiating between intended actions and carelessness in this case. Incitements and offensive statements resulting from carelessness cannot be considered as angry statements designed to instigate enmity.

Judicial pronouncements have consistently reaffirmed the requirement of mens rea in such offences. The Indian courts have taken a clear stand on the aspect that mere utterance (even with potential for disturbance) will not bring the accused within the ambit of the law unless accompanied by a requisite guilty mind. However, the proof of intention or guilty mind has always been an inherently difficult task and requires close analysis of the attendant circumstances including language, audience, and occasion of utterance.

With regards to the offence of outraging religious feelings, the mens rea requirement is expressed even more overtly, mandating deliberate and malicious intent thereby making it a higher burden of proof for the prosecution. Such a high degree of mens rea in this particular offense acts as a protection against the framing of charges for genuine debate on religious matters, while simultaneously taking into account the potential for misuse to cause hurt. The interpretation of such "malicious intent" also largely depends on the response that the utterance may elicit, since the listener's subjective response cannot be the sole criterion for determining whether the utterance is genuinely "malicious" and intended to outrage.

Nevertheless, the assessment of the mens rea component in regard to hate speech proves to be much more complex due to its susceptibility to contextual factors, such as the political atmosphere, events happening in society at large, and even on the Internet, where political undertones can imbue statements that might otherwise be innocent with significant meaning.

Because it blends an objective component (what an average person would anticipate) with a subjective component (the accused's assessment of likelihood), the element of likelihood also presents an intriguing conundrum. Therefore, if an individual's conduct were reasonably anticipated to result in the banned outcome, they may still be held accountable even in the absence of actual intent. Because of the intricate relationship between likelihood and intent, a speaker may be guilty of hate speech even in the absence of

overt provocation if a reasonable person would have anticipated the injury.

The concept of mens rea is now called into question by the emergence of digital technology, since statements can now be easily reposted and recontextualised, frequently going much beyond the speaker's original meaning or expectations. However, when the statements are magnified and altered by others, it also becomes extremely difficult to determine the subjective mind of the original speaker. Therefore, it may be said that mens rea serves as both a defence against absolute responsibility and a tool that adds subjectivity and complexity to the definition of hate speech offences.

3.5 Judicial Approach to Criminal Hate Speech

In dealing with criminal hate speech in India, the courts seem to have embarked on a venture of reconciling the expansive reach of the statutory provisions with constitutional limitations on free speech. In applying the criminal law to cases of hate speech, courts have, so far, generally sought to balance it with the framework under Article 19(1)(a) and Article 19(2), by using a combination of textual, contextual, and constitutional analysis.

In cases involving offences relating to promotion of enmity between different groups, courts place significant emphasis on both intent and the tendency of the speech. In *Bilal Ahmed Kaloo v. State of Andhra Pradesh*²⁰, the Supreme Court held that mere criticism of another religion does not amount to an offence unless it is accompanied by an intention to promote enmity or hatred between groups. The Court clarified that the existence of mens rea is a necessary element of such offences, and that liability cannot be imposed solely on the basis of the content or effect of speech.

While stressing the importance of intent and impact, the courts have also acknowledged the context in which the speech is made, as the meaning of words is dependent on the circumstance in which they are delivered. The speaker and audience identity, as well as the social and political context are some of the relevant factors that the courts seem to be inclined to consider in all matters. In fact, the focus on the context becomes particularly pertinent in cases where it is contended that the speech is likely to result in disturbing communal harmony, as similar speech,

which in other contexts might pass off, could easily lead to communal disturbances.

Another key ingredient the courts consider, in conjunction with constitutional dictates, is proximate connection between speech and harm. The Supreme Court in *Manzar Sayeed Khan v. State of Maharashtra*²¹ held that the effect of allegedly offensive words must be judged from the standpoint of reasonable, strong-minded individuals, rather than those who are weak, hypersensitive, or easily provoked. By adopting this objective standard, the Court ensured that criminal liability is not imposed on the basis of exaggerated or overly sensitive reactions. This approach prevents the misuse of penal provisions by requiring that the impact of speech be evaluated in a rational and balanced manner.

While dealing with speech relating to religion, emphasis has been laid on a deliberate and malicious intent and the need to differentiate it from mere satire, ridicule or mere condemnation of a religion. The judiciary has sought to maintain this distinction through careful parsing of the speech. However, there has not been much consistency in the interpretation as far as intent and context are concerned, which ultimately boils down to judicial interpretation of words and context.

Yet another aspect that seems to be kept in mind by the courts is avoidance of overlapping prosecution. In cases where multiple sections of the Indian Penal Code can be applied, courts have sometimes also questioned if the ingredients of all the offences have been established. Such an examination is a pragmatic attempt by the courts to avoid multiplicity of prosecutions based on the fragmented structure of criminal law provisions dealing with hate speech.

However, the above aspects do not form a clear and uniform doctrine for dealing with cases involving hate speech and the varying judicial response in such cases suggests that much has been left to individual interpretations. In some cases, courts have been quite lenient and not restricted even speech which has the potential of causing disturbances in public order and the element of proximate connection and likelihood of a probable incitement have been required. Whereas in certain other cases a rather restrictive approach has been adopted where, apart from a clear intent and proximity of effect, a high probability of likely harm also needed to be established. In the end, it must be acknowledged that these factors, such as those

²⁰ *Bilal Ahmed Kaloo v. State of Andhra Pradesh*, (1997) 7 SCC 431.

²¹ *Manzar Sayeed Khan v. State of Maharashtra*, (2007) 5 SCC 1.

derived from constitutional law like proximity and incitement, often do not get integrated into the interpretation of criminal law smoothly, hence leading to varying outcomes in similar fact-situations.

3.6 Landmark Case Law Analysis

The judicial approach to criminal hate speech provisions in India can be examined through a series of cases that highlight both the scope of statutory offenses and limitations imposed by constitutional principles. Analysis of relevant cases demonstrates the courts' considerations of intent, context and impact, as well as the difficulties presented by broad statutory provisions.

In an early, significant case, *Ramji Lal Modi v State of Uttar Pradesh*²², the Supreme Court upheld the constitutional validity of Section 295A IPC, which penalizes deliberate and malicious acts intended to outrage the religious feelings of any class of citizens. According to the Court, the clause constituted a legitimate restriction under Article 19(2) since it only addressed severe kinds of insult that were committed with malevolent and intentional purpose. The Court limited the scope of the offence while maintaining its validity by highlighting the requirement of mens rea and making a distinction between conduct meant to offend or insult and genuine criticism of religion.

The Supreme Court considered a conviction for inciting animosity amongst various communities in *Bilal Ahmed Kaloo v. State of Andhra Pradesh*. The distribution of allegedly offensive content was at issue in this case. The Court determined that a necessary component of the offence is the presence of mens rea, which takes the shape of a purpose to incite animosity or hatred amongst groups. The accused was found not guilty by the court because such an intention was not proven. The ruling emphasised that publishing content alone—without the necessary intent or propensity to incite hatred—is not enough to result in criminal culpability.

The importance of the objective test in such cases was emphasized by the Supreme Court in *Manzar Sayeed Khan v State of Maharashtra*. The case involved allegations that a publication insulted religious sentiments and promoted enmity between communities. The Court held that the effect of such expression must be judged from the perspective of a reasonable, strong-minded individual, rather than

those who are hypersensitive or prone to taking undue offence. It emphasized that speech must be evaluated in its proper context and in a dispassionate manner, so as to prevent disproportionate or exaggerated interpretations.

In more recent times, the Supreme Court in *Amish Devgan v Union of India*, looked at the limits of speech on the controversial allegations made against a particular religious figure. The Supreme Court case had its genesis in criminal proceedings against the author of the objectionable publication, but touched upon broader concerns about hate speech and its capacity for harm to social harmony. While it acknowledged the harms that speech can cause to religious communities and the larger society, the court's observations still revealed the difficulty of striking a balance between free speech and protection from harm, with a lack of concrete doctrinal standards.

The above case decisions reflect some common themes of judicial approach to criminal hate speech in India. First, it has consistently been stressed that the provision should be applied strictly and that mens rea, in the form of intention to cause outrage, hostility or enmity, should be established. Second, the courts have underscored the role of context in judging whether the speech or publication in question may harm a community or group and have also tended to favor the interpretation from the perspective of a reasonable, non-extra-sensitive citizen. The third observation that is borne out of this analysis of case law is that the various laws enacted in this regard, while having the capacity to prohibit and penalize speech for the purposes it is intended for, have not been approached and applied in a coherent, principled and well-defined manner leading to uncertainty in interpretation.

3.8 Misuse and Selective Application of Law

In India, the criminal prosecution for hate speech is frequently beset with concerns of misuse and selective implementation. Although the legislative framework aims to prevent speech that would cause social unrest, the vague and expansive wording allows these provisions to be applied beyond their intended meaning and scope. This frequently results in situations where the application of the statute is more influenced by context and external factors than by the severity and nature of the expression made.

²² *Ramji Lal Modi v. State of Uttar Pradesh*, AIR 1957 SC 620.

The problem lies in the usage of criminal law beyond its legal application of regulation. In cases of expression promoting enmity or outraging religion, sections such as Section 196 and Section 197 of the Bharatiya Nyaya Sanhita, 2023, along with Section 299 are sometimes used in relation to criticism, satire and dissension, rather than malicious incitement. The lack of objective tests for determining the extent of criminality renders these actions susceptible to the arbitrary application of law on the basis of perceptions of offense, rather than demonstrable harm, expanding the application of criminal law beyond what is necessarily protected under Article 19(1)(a).

The selective application of these laws, and of laws that prohibit hate speech more broadly, are another layer to this problem. Action on such expression often stems from external circumstances such as the visibility of the speaker, the sensitivity of the issue, and media interest in the matter, rather than any principled application of the law, allowing those in positions of authority, and thus those in more visible circumstances, to receive disproportionately more criminal scrutiny for similar or perhaps less offensive speech made by others. It leads to an application of law that is not uniform for everyone, violating the principle of equality before law. The subjective application of these laws often depends on the given context and situation rather than a universal application of the legal provisions laid down.

In some instances judicial observations have taken into account the potential for misuse of the laws, as was seen in *Pravasi Bhalai Sangathan v Union of India*²³, where the Supreme Court considered the problems with the legal regulation of hate speech and warned of the dangers of misuse, thereby indicating a gap between how these provisions were meant to be applied and how they were actually applied.

Yet another form of misuse involves the filing of parallel prosecution in more than one forum for the same act. This is becoming particularly prominent now with the proliferation of the Internet and social media, where it is possible to disseminate content simultaneously across several platforms and jurisdictions. An individual can be made to undergo several trials simultaneously based on a single instance of expression, creating inconvenience and harassment through law itself.

Chilling effects on free speech have emerged due to this misuse and selective application, as it may deter citizens from expressing themselves on matters of public interest, especially if such speech may be perceived as provocative by others, regardless of whether it actually constitutes hate speech and despite evidence of their intention to act otherwise, for fear of protracted litigation. It makes clear what might be deemed criminal speech unclear to the average citizen who might be unaware of the exact implications of various utterances, and thus refrains from engaging in lawful expression.

IV. DIGITAL TECHNOLOGY AND HATE SPEECH

Digital technologies have drastically altered the character, volume and reach of speech, producing a new regulatory context for hate speech. The nature of speech was traditionally bound by geography and the means of its dissemination; the digital space has effectively removed these constraints, allowing for instantaneous and global dissemination.

The sheer pace and volume of conversation on the internet is one of its main characteristics. Speech may reach thousands or perhaps millions of people almost instantly thanks to social media networks, instant messaging services, and online forums. Due to the ease of forwarding, sharing, and reposting, speech that could have had a national or local reach if it had been written in a newspaper can quickly gain international recognition.

4.1 Rise of Online Hate Speech

The Internet, especially social media sites and blogging platforms, allows individuals, in the absence of the mediation of a newspaper, publisher or television network, to reach an enormous audience and to publish whatever they wish on the internet. The reduction of the barrier to entry into speech has created new possibilities for the production and circulation of harmful speech by many people for whom traditional barriers may have operated as sufficient regulation.

Another dimension of the issue is the effect of anonymity in on-line spaces. Since individuals who post controversial and perhaps actionable speech often do so under an assumed identity which cannot easily be traced to the actual person who published it,

²³ *Pravasi Bhalai Sangathan v. Union of India*, (2014) 11 SCC 477.

hate speech flourishes with a relative absence of social or legal accountability. The ability to use one's true identity, even in spite of the potentially large or offensive nature of the speech used, creates a form of speech which is altogether different than the forms previously subject to legal regulation.

This effect is amplified by algorithmic amplification of speech. Since platform owners have no incentive for speech to have a certain value of merit, and the value they are concerned with is whether and how much someone will interact with a post or story. Algorithmic dissemination allows a certain type of speech, i.e., outrage, that may be offensive to certain communities, to take on enormous significance and velocity when it would otherwise not; in these ways hate speech can become exponentially more prominent with little real action required.

Further increasing the intensity of hate speech on-line is the fragmentation of the audience. It is easy for one to surround himself with like-minded people on-line so that certain opinions, even potentially racist ones, are echoed and repeated back until the person believes that what he says represents mainstream discourse and popular opinion, but in actuality only represents the echo chamber he surrounds himself in on-line. This dynamic can occur slowly over long periods of time and as speech becomes more extreme, the line becomes incredibly fine when deciding what is actionable.

Lastly, hate speech on-line utilizes new forms of communication. Hate speech in the on-line realm is often composed in memes, or in words, sounds or images that only specific audiences, or the creator of the hate speech and others like him or her, can understand, due to a particular context or frame of reference. This is the opposite to traditional forms of communication, the primary concern of which was its literal meaning.

4.2 Role of Social Media Platforms

The pervasiveness of online hate speech is closely related to the structure and mechanism of social media websites. Platforms are dynamic actors that control the visibility and dissemination of online material rather than merely being passive channels of communication; in fact, it is primarily because of these structural features that they are crucial to the regulation and spread of such content.

Social media platforms are conducive to the creation, transmission and interaction with content at a scale unimaginable with previous communication media. They are "publishers by implication rather than

choice, their technological architectures rather than editorial judgments shaping the exposure of content. Providing a technically mediated communication structure, platforms do not engage in prior editorial control over user-generated content, allowing vast amounts of speech, including hateful speech, to enter public discussion unfiltered and in huge volume." This ability to upload large quantities of speech to a public forum without prior editorial review is fundamental to their ability to become so influential in shaping public debate, but also to the potential for harmful speech to proliferate: such speech has always been able to exist within a space that makes itself available to all and is structurally conducive to speech rather than to carefully calibrated or curated discourse.

Platforms, however, do not treat all speech in their systems with the same impartiality. The design of the algorithms used for content curation means that some speech is necessarily given higher levels of prominence than others. Typically, this occurs with speech which produces high levels of engagement, meaning that a high degree of emotional content will be amplified: speech which is polarising, controversial, or hate speech will therefore be disseminated to more users and remain on users' screens for a longer amount of time compared with mundane speech which has less capacity to provoke strong emotions from users. The amplification of hate speech is not necessarily the product of conscious intent but rather an 'emergent property' of their design.

Although platforms are structurally inclined towards the dissemination of content, they have also developed content moderation policies. These policies involve both automated processes (algorithms designed to detect and filter out harmful content) and human moderators who review potentially harmful content, attempting to remove violations and uphold platform rules. The effective operation of these policies has been a difficult task, due to the enormous scale of content, the diversity of languages and cultural context, and automated systems' inability to easily differentiate the nuances between harmful and legitimate speech; as a result, "hate speech. "Crucial in the context of content moderation is the tension between the need to curtail harmful speech and the principle of freedom of expression. While platforms undertake a quasi-regulatory function in moderating their sites, ensuring that their moderation policies uphold the right to freedom of expression is a paramount

concern; over-removal can prevent legitimately expression, whilst under-removal fails to contain the harms. As yet, no universal, transparent, and enforceable code of practice exists, leaving users and those affected by hate speech to face uncertainty as to what kind of speech will be considered punishable and what kind of speech will be protected by either international conventions on speech or internal platform rules.

The international reach of platforms complicates all of the foregoing considerations. In the present communication space, speech is 'global in scope', and even where speech posted in one country is, under that country's law, only able to be viewed by people who reside within that country (e.g. Due to age verification rules, which can themselves be problematic), other users across the world will nonetheless still be able to access that speech due to the nature of international internet communication. Platforms must therefore simultaneously operate under the different regimes of different states: "Each state's laws will dictate which kinds of speech should be constrained, and states are more and more relying on platform intermediary companies to perform these duties for them." This has implications for the idea of 'global internet freedom of expression'; this notion of free speech is not necessarily freedom from the laws of different countries, but rather freedom from those laws, even if there is an expectation that countries themselves should try to align their laws with one another in relation to online hate speech, but even here there is no easy consensus on how to do that.

Overall, the role of the platform is complex and ambiguous: a quasi-public utility in relation to speech, yet operating as a private entity whose structural dynamics have, by necessity, had a massive influence over the way speech operates and disseminates online; thus, its rules and regulations are intertwined with issues of law and technology and the balance between them.

4.3 Algorithmic Amplification and Echo Chambers

Hate speech cannot be discussed solely as a result of user activity but should be addressed within its architecture as an algorithm. The platforms' algorithms are designed in a way to direct visibility, relevance and circulation; these features contribute to the form of expressions being prominent in public space.

Algorithms work by enhancing engagement of contents by pushing popular contents which has most engagements. This is mainly composed of content

which evokes higher emotional response, such as outrage or hate, fear etc. Thus provocative speech gets more visibility and less thought given whether the contents impact public positively. In the case of hate speech it also causes those contents to be more viral, that are targetting or reinforces division toward certain groups.

The amplification effect of algorithmic systems is exacerbated by the formation of echo chambers, where only information conforming to pre-established beliefs is seen. With personalized algorithms, platforms filter the data with prior activities and interests, leading to homogeneous space. Here, contents which could be argued/challenged elsewhere gets reinforced and the divisive, hate, prejudiced thoughts flow freely.

Through echo chambers, harmful expression get normalized, where users get used to the expression and may see the harm less as it is frequently shown. Over time this can strengthen ideas that lead to exclusive or discriminatory social thought, even without direct incitement.

Amplification is also observed in terms of speed and network propagation. Once a certain piece of content becomes visible, it spreads through various channels, possibly with differing levels of distortion of the context or the content itself. The very process of diffusion may not provide any information that would indicate whether there was some intentionality behind the original expression.

However, the issue becomes more complicated when recommendation algorithms guide users towards more and more radical content according to their browsing history and activity. It should be emphasized that this approach does not involve deliberate attempts to promote malicious content. Rather, this could become an unintended consequence of optimization processes aimed at maximizing user engagement regardless of other considerations. The resulting interaction of algorithms and users forms a positive feedback loop where users react to and engage with whatever is being offered to them, and the stronger their reaction, the more likely it is that such content will be shared with others.

The implication is not limited to isolated examples of individual hate speech but can be observed to change the underlying structure of public discourse; when algorithms determine what and how certain voices are amplified, they consequently alter conditions for democratic debate. Without measures to mitigate the amplification of hate speech and the marginalization

of excluded voices, existing algorithms can support its pervasiveness.

4.4 Intermediary Liability under the Information Technology Act, 2000

A fundamental determinant of effective regulation of hate speech on digital platforms is the manner in which the legal regime regulates intermediaries—the entities through which online content is created, hosted, and distributed. In India, the existing legal regime that governs the liability of intermediaries is laid down by the Information Technology Act, 2000 ("IT Act").

The IT Act provides for a conditional exemption to intermediaries from liability. Often referred to as a "safe harbour", this exemption provides immunity to the platform from liability for any third-party content uploaded onto their platform if the platform can meet certain statutory conditions laid down in the IT Act. The logic behind providing for a "safe harbor" regime is that it allows digital platforms to flourish without requiring them to monitor each and every piece of user generated content while holding them liable if they are aware of infringing content.

However, this exemption is not absolute. Intermediaries have a duty to observe "due diligence" which essentially means that when they are aware of infringing content they must, within a stipulated time, disable or remove such content. Inability by the intermediary to meet its due diligence requirements will result in the intermediary losing the "safe harbor" status, making the intermediary responsible for the user-generated content on the intermediary's site. This "conditional immunity" model aims at ensuring that the intermediary functions with some level of independence but still remains answerable for what it does and fails to do regarding illegal content.

The role played by judicial interpretation cannot be underestimated when it comes to defining intermediary liability and situations where an intermediary will be considered to have "knowledge". In the landmark case of *Shreya Singhal v Union of India* (2015) 5 SCC 1, the Supreme Court while striking down section 66A of the IT Act observed that "Knowledge on the part of an intermediary would then have to be based on either a court order or a notification from the government authorities". It was further clarified by the court that a "private complaint should not, by itself, be grounds for treating an intermediary as one who has knowledge of unlawful content", and such a position

would otherwise compel intermediaries to arbitrarily remove content and thus severely impact free speech. While the *Shreya Singhal* judgment narrowed the scope of circumstances under which intermediaries are obliged to remove unlawful content, it has also presented practical issues with regard to tackling harmful speech as it relies on official notices to prompt removal, which can lead to the proliferation of hateful and dangerous content on digital platforms. Given the scale and speed with which hate speech can be transmitted online, intermediary liability holds significant relevance. Since the intermediary controls the channel through which content flows on the platform, they are best positioned to stop the proliferation of unlawful content on their platform. However, the successful use of this intermediary power hinges on having an adequately robust regulatory regime.

Moreover, further developments through secondary legislation in the form of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 have clarified intermediary due diligence and created additional requirements which will certainly compel intermediaries to be more active in the management of user-generated content on their platform. For instance, the additional due diligence requirements placed on the Indian intermediaries by the Rules include making the content of their privacy policy and terms of service public and informing users of intermediary laws and regulations.

The crux of intermediary liability is in finding the appropriate balance between platform responsibility and individual rights. If intermediaries are overly liable, they may resort to too many preventive measures that could potentially result in the removal of lawful content, leading to a regime of censorship that curtails freedom of speech. On the other hand, if there is no regulation, harmful content will proliferate, thus defeating the purpose of hate speech law.

Also, the international dimension of internet companies makes it difficult for the regulation of intermediaries since intermediary liability will not be confined merely to servers housed in the nation but will have to encompass the content seen by Indian users, even if these contents are stored on servers outside the country.

The importance of intermediary liability in regulating hate speech in cyberspace is perhaps one of the toughest challenges that exist. The regulatory scheme as created under the IT Act tries to apportion liability

in a manner that involves the three major stakeholders – the intermediaries, users, and the state.

4.5 IT Rules, 2021 and Due Diligence Obligations

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("IT Rules, 2021") represent a significant extension of the regulatory framework for digital intermediaries in India. Drafted under the Information Technology Act, 2000, the rules impose stringent due diligence obligations upon the intermediaries and redefine the boundary between platform immunity and platform liability. In the domain of hate speech, these duties are central to the extent to which the harmful content can be contained quickly and effectively.

One of the defining features of the 2021 rules is the due diligence obligations required from intermediaries in terms of surveillance and management of user-generated content. Among other requirements, the rules mandate the obligation upon intermediaries to inform their users not to host, display, upload or share content which is illegal and specifically, content that is in contravention of inter-group enmity or threats to public order. By incorporating these standards in platform terms, the rules attempt to bring about private moderation mechanisms within the legal framework.

The rules also establish a grievance redressal mechanism that requires intermediaries to appoint a grievance officer and create a structure for receiving and processing grievances. The users now have a formal platform where they can file a grievance concerning content they believe is illegal and against the terms of use of the platform including hate speech. Intermediaries are bound to acknowledge receipt of the grievance within a specified period of time and also to take necessary action, institutionalising platform-level content regulation.

Further, certain categories of intermediaries especially significant social media intermediaries, have been subjected to additional obligations which require them to appoint compliance officers, provide periodic compliance reports and also have systems in place to identify the originator of the information, in specific circumstances. Traceability has been a highly contentious aspect especially for communication platforms which use end-to-end encryption due to obvious concerns of privacy and free expression.

The IT rules, 2021 require a time-bound take-down or disabling of content upon receipt of actual

knowledge through a court order or a notification by the government. This clause attempts to address the issues related to delayed responses which are pertinent in the case of online hate speech given the speed at which such content can be disseminated and harm caused, even before official intervention can take place. The enforceability of such clauses however, depends upon clarity of such orders and intermediary capacity.

Another point that the 2021 rules address is the requirement of monitoring and tracing by an intermediary especially in relation to some forms of content. Though the rules do not make intermediaries universally obliged to monitor content, they demand the use of reasonable measures by the intermediaries including automated tools for identification and removal of unlawful content. This brings an element of pro-activity into content regulation but also raises the potential for over-filtering or suppression of legitimate expression.

Thus, the IT rules 2021 broaden the scope of obligations required from the intermediaries thus positioning them as a central mechanism for regulation of speech. This shifts the burden of liability more towards intermediaries as against the state for the content posted on the platforms.

V. CONCLUSION

The current paper conducts a multi-layered analysis of the legal regulation of hate speech in India, taking into account aspects related to the constitution, criminal laws, digital regulation, theoretical framework, comparative aspects, and enforcement procedures. The paper comes to the conclusion that, despite India's extensive legal protections against speech that incites hatred and disturbs public order, this regulatory framework is severely limited due to conceptually hazy boundaries, a fragmented statutory framework, and uneven implementation.

Constitutionally, the constitutional protection granted to freedom of speech by virtue of Art 19(1)(a) coupled with the limitations permitted under Art 19(2) makes the protection conditional. Judiciary's attempts to refine the limits of this protection by drawing a line between advocacy and incitement, and by demanding "close proximate relationship" between speech and harmful act, could not be translated into statutory wording nor in constant practical application. The investigation into criminal law showed that hate speech isn't tackled by a singular, integrated approach but rather by various

and intersecting sections and provisions, leading to lack of definitive conceptual boundary and thresholds and hence making a loophole for either overreach or under-enforcement.

The issue is further complicated in the case of the digital space, due to the rapid pace and massive scale of communication, which defies traditional regulatory methods. Comparative study showed that India fits in somewhere in the middle between a liberal or speech protective system and illiberal system of speech control; however, this middle position does not seem to provide India a rigid, structured or coherent method but on the contrary raises a query about lack of definition clarity and proportionality. The study of implementation further showed the systemic issues including arbitrary application, procedural lag and outside pressures which led to the inefficiency of the scheme of speech control in general and restriction on free expression in particular.

A reactive approach to the problem of dealing with hate speech in India needs to be transformed into a systematic and structured approach based on well-defined standards. Rather than extending the list of limitations, an effort should be made to establish a specific set of guidelines that would regulate the existing limitations framework such that it only prohibits the statements causing harm.

More effective procedural safeguards, more balanced criminal laws, precise definitions and adequate control over online platforms are essential for implementing such a framework. Institutional efforts and legislation aimed at establishing conducive circumstances for promoting hate speech also need to be considered.

Given the changes in social environments and technology, the constitutional challenge related to achieving a proper balance between the two conflicting rights – freedom of expression and prohibition of harm – needs to be constantly addressed. A proper mechanism ensuring the stability of such a balance can be established as a result.

Finally, the findings made during this study show that the problem of regulating hate speech is systemic rather than merely legal in nature. To ensure speech protection, as well as its non-harming character, a coherent and consistent legal framework needs to be developed.

REFERENCES

- [1] Romesh Thappar v. State of Madras, AIR 1950 SC 124.
- [2] Brij Bhushan v. State of Delhi, AIR 1950 SC 129.
- [3] Kedar Nath Singh v. State of Bihar, AIR 1962 SC 955.
- [4] Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India, (1985) 1 SCC 641.
- [5] Secretary, Ministry of Information & Broadcasting v. Cricket Association of Bengal, (1995) 2 SCC 161.
- [6] Bennett Coleman & Co. v. Union of India, (1972) 2 SCC 788.
- [7] S. Rangarajan v. P. Jagjivan Ram, (1989) 2 SCC 574.
- [8] Tata Press Ltd. v. Mahanagar Telephone Nigam Ltd., (1995) 5 SCC 139.
- [9] Shreya Singhal v. Union of India, (2015) 5 SCC 1.
- [10] Constitution of India, art. 19(2).
- [11] Chintaman Rao v. State of Madhya Pradesh, AIR 1951 SC 118.
- [12] Dr. Ram Manohar Lohia v. State of Bihar, AIR 1966 SC 740.
- [13] Superintendent, Central Prison v. Dr. Ram Manohar Lohia, AIR 1960 SC 633.
- [14] Modern Dental College & Research Centre v. State of Madhya Pradesh, (2016) 7 SCC 353.
- [15] Francis Coralie Mullin v. Administrator, Union Territory of Delhi, (1981) 1 SCC 608.
- [16] Subramanian Swamy v. Union of India, (2016) 7 SCC 221.
- [17] Bilal Ahmed Kaloo v. State of Andhra Pradesh, (1997) 7 SCC 431.
- [18] Manzar Sayeed Khan v. State of Maharashtra, (2007) 5 SCC 1.
- [19] Ramji Lal Modi v. State of Uttar Pradesh, AIR 1957 SC 620.
- [20] Pravasi Bhalai Sangathan v. Union of India, (2014) 11 SCC 477.