

FIGHTING HATE SPEECH, BALANCING FREEDOMS: A REGULATORY CHALLENGE

—MAYA MIRCHANDANI*

The right to free expression is the heart of a democracy. It is necessary to ensure minimal government regulation in the communication of ideas, especially when it comes to diverse opinions and dissent. Freedom of expression entails the individual's right to participate in decisions that affect him/her and the public's right to unrestricted access to information that enables an informed decision to assist democratic decision-making and self-expression. Thus far, in instances of legal challenges to free speech and the need for civil discourse, the Supreme Court of India has been careful to avoid judicial overreach by clearly dictating its intention to function under the purview of existing laws, and a legal interpretation of the ambit of 'reasonable restrictions' as outlined in the Constitution of India. There is reluctance in creating exceptions since greater value has been accorded to expression.

The explosion of social media, the absence of filters in social communication, the credibility crisis of mainstream media and the spread of fake news and propaganda have all combined to create an environment in which the stifling of dissent is done not necessarily by the state, but by the public itself via hate speech, personal abuse and threats of violence. Hate speech relies on stereotypes which can carry on from historic ideas or symbolism usually directed towards minority groups (races, caste, gender, etc.). The prevalence of hate/extreme speech and the inadequate response to it by both the public and the state pose a challenge of definitions, of the need to for clarity on what reasonable restrictions mean and on ensuring that the response to it be principled and not whimsical in order to suppress what we don't like to hear. This paper aims to discuss the need for online self-regulatory mechanisms rather than Government involvement as well as adequate definitions without placing open-ended limitations on Speech. Online platforms must ensure sensitization and spread awareness by

* The author is an award-winning Indian broadcast journalist with vast reporting experience on Indian Foreign Policy, with a focus on South Asia. She has extensively covered the Kashmir conflict, the civil war in Sri Lanka as well as internal identity politics and conflicts in India. She now teaches Media Studies at India's Premier Liberal Arts Institution, Ashoka University; and is a Senior Fellow at the Observer Research Foundation, where she heads research on the intersections between hate speech, populist politics and violent extremism.

running counter narratives against hate, non-proliferation of misinformation is another issue that must be tackled extensively. The need to engage with hate groups now is even more to counter extremism by making them question their stance.

"There is more than one kind of freedom.....freedom to and freedom from".

—Margaret Atwood, *The Handmaid's Tale*

I. INTRODUCTION

Wearing a white doctor's coat, a stethoscope around her collar, a pink headscarf and a broad smile, Hadiya Asokan's **posed for a photograph posted online by her husband Shafin Jahan, in which he** announced to the world in February 2018, that his wife had finally completed her medical degree and was officially now, a doctor. Ordinarily, his proud photograph first published on Facebook (the post settings have since made it unavailable for public viewing) would have seemed routine, after all, the social media platform is full of individual, personal successes that its over 2 billion global users share with their social networks- made up of friends, family, colleagues, even strangers- every day. But Hadiya's story was nationally significant, her success nationally celebrated.¹

Twenty-four year old Hadiya (nee Akhila) found herself in the eye of a hurricane when she married Shafin in 2016. Her parents reported their daughter missing with the police, and approached the Kerala High Court with a habeas corpus petition to find her. They claimed that she had been forcefully converted from Hinduism to Islam in an effort to recruit her for the Islamic State and take her to Syria. Hadiya claimed she had converted and married Shafin on her own, but the Kerala High Court annulled her marriage a year later, on grounds of a report submitted by the National Investigation Agency (NIA) saying that Hadiya was a victim of indoctrination. It also handed her back to her parents, saying they had custody of an unmarried daughter according to 'Indian tradition.' Shafin Jahan, her husband, approached the Supreme Court of India to intervene and uphold her rights and freedoms as an adult. The nation's top court ultimately upheld Hadiya's freedoms and said the validity of an adult woman's marriage can only be challenged if she herself complained. In the two years between her father's complaint and the Supreme Court judgement, Hadiya's story became a fresh catalyst for growing online vitriol and hate speech driven by communal hatred and Islamophobia, as well as by sexist, discriminatory and abusive language towards women and choice in general. It was a textbook example of what the Hindu

¹ "Braving All Odds, Hadiya is Now a Doctor. " *The Hindu*, 28-2-2019. <https://www.thehindu.com/news/national/kerala/braving-all-odds-hadiya-is-now-a-doctor/article26397923.ece>.

Right in India had begun to term ‘love jihad’ - or the marriage of Hindu women to Muslim men and their religious conversion in the process.

Perhaps no other issue has become as much of a lightning rod for Hindutva groups as inter-faith marriages, especially those between a Muslim and a Hindu, and resultantly the battleground upon which much online hate is poured. The use of the term ‘*love-jihad*’ has become common, indicating a new normal in social discourse. In February 2018, Facebook had to pull down posts that enumerated over 100 interfaith couples, named the women who “had become victims of *love-jihad*”, and called for “Hindu lions” to hunt down those Muslim men³⁴. Later in September, members of the ruling Bharatiya Janata Party’s affiliate, the Vishwa Hindu Parishad attacked a couple in Meerut and handed them over to the local police. A video⁵, showing the girl- a Hindu- being beaten up inside a car by a woman police constable surfaced. As she is beaten, a male officer taunts her with innuendo laden abuse, saying “*You prefer Muslims when so many Hindus are around*”⁶.

Hadiya’s case was a textbook example of the intersections between social and political violence, online abuse and rule of law in a liberal democracy. By repeatedly accusing Muslims of ‘stealing their women’ and turning India into a Muslim majority nation by making them produce ‘Muslim’ children, Hindu right wing groups have stoked fear and enmity between communities and, until recently, with mainstream media’s unwillingness to provide platforms to this vitriol, successfully used social media to amplify their message at great speed.⁴ In several cases, Muslim men in the real world face consequences of being portrayed as bogeymen in such interfaith relationships; and, in many cases, are subjected to extreme violence as a result. The fear of “Love Jihad” preys on a manufactured sense of offendedness that preys on inevitably patriarchal socio-cultural anxieties around the supposed assault on family, religious and caste honor that interfaith relationships represent⁷.

If the constant whipping up of communal frenzy over this is problematic, the abuse against those who challenge the narrative and speak for individual

² A. Vaidyanathan, “Hadiya Adult, Can’t Question Her Marriage, Says Supreme Court” NDTV, 23-1-2018. <<https://www.ndtv.com/india-news/hadiya-love-jihad-case-supreme-court-says-nia-cant-probe-her-marital-status-1803501>>.

³ “FB Page Lists Interfaith Couples, Asks Hindus to ‘Hunt’ the Men.” The Quint, 5-2-2018. <<https://www.thequint.com/news/politics/hindutva-fb-page-calls-for-attack-on-interfaith-married-couples>>.

⁴ Karnika Kohli, “Hindutva Page Lists Over 100 Inter-Faith Couples, Asks Hindus to Hunt Them Down.” *The Wire*, 5-2-2018. <<https://thewire.in/communalism/hindutva-vrata-facebook>>.

⁵ “Meerut Police Beat Up Woman on Suspicion of Affair with Muslim Man.” *India Today*, 26-9-2018. <<https://www.indiatoday.in/india/story/woman-abused-slapped-by-police-in-meerut-for-choosing-muslim-partner-1348803-2018-09-25>>.

⁶ *Ibid*.

⁷ “Facebook India: Towards the Tipping Point of Violence Caste and Religious Hate Speech.” Equality Labs, 2019. <<https://www.equalitylabs.org/facebook-india-report>>.

rights and liberties, is often far worse. Dissenting voices are regularly accused of being ‘anti-Hindu’ or ‘Muslim/terrorist sympathisers.’⁸ If the critics are women, the attacks inevitably take on a gendered lens. In spite of efforts to debunk both the term and the conspiracy theories behind “Love Jihad” it has proven persistent—abetted in part by a large disinformation echo-chamber on social media platforms that constantly stoke these sentiments. The continuous depiction and language around all instances around interfaith marriages challenge online platforms regularly as they try to navigate a balance between allowing users to express their opinions freely, and censuring them for what is now commonly called hate speech, or offering the freedom of the platform as an opportunity to publicize and promote discriminatory behavior. In India, anti-minority rhetoric, and campaigns to discredit political opponents, have emerged as two of the most significant challenges to hate or extreme speech online.

II. DEFINING & IDENTIFYING HATE SPEECH ONLINE

Categorising and defining “hate” is perhaps one of the most perplexing questions of our times, especially because the definition is uniquely tied to the impact of what has been said. In a socially networked world where comment is free and reactions are instant, lines between violent personal abuse and/or speech inciting violence against a community or group are becoming increasingly blurred. At times, even if the intent and language are not explicitly hateful, the implications can be. A study conducted by this author for the Observer Research Foundation defined “hate” as “expressions that advocate incitement to harm (particularly, discrimination, hostility or violence) based upon the targets being identified with a certain social or demographic group. It may include, but is not limited to, speech that advocates, threatens, or encourages violent acts.”⁹ The report also indicates that hate speech may be prone to manipulation at critical times—during election campaigns, for example, or used by those in power to curb legitimate dissent—where hate speech can take on the contours of what media theorist Cherian George terms “hate spin”¹⁰. This is where the nexus between disinformation and hate speech becomes an apparent threat and can be manipulated as propaganda to influence public sentiment. In fact, according to a report by the Council on Foreign Relations in the United States, analysts say trends in hate crimes around the world ‘echo changes in the political climate, and that social media can magnify discord.’¹¹ In extreme cases, invective or hate laden social media commentary has led to mob lynchings and even ethnic cleansing.

⁸ Annex. 1.

⁹ Maya Mirchandani, Dhananjay Sahai and Ojasvi Goel, “Encouraging Counter-Speech by Mapping the Contours of Hate Speech on Facebook in India,” Observer Research Foundation, 3. <https://www.orfonline.org/wp-content/uploads/2018/03/ORF_Report_Counter_Speech.pdf>.

¹⁰ Cherian George, *“Hate Spin: The Manufacture of Religious Offense and Its Threat to Democracy”*, Cambridge, MA: The MIT Press, 2016.

¹¹ “Hate Speech on Social Media: Global Comparisons.”, Council on Foreign Relations, 7-6-2019. <<https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons>>.

While there is no global consensus on what constitutes hate speech, Facebook - the platform that hosts a third of all humanity at 2 billion users worldwide- defines hate speech as a direct attack on people based on ‘protected characteristics’ of race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. Furthermore, it defines ‘attacks’ as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation¹². Refugees, immigrants of all kinds are also, ostensibly, offered some of these protections. With ownership of WhatsApp and Instagram, Facebook’s community standards apply broadly to all its products. The micro-blogging site, Twitter also spells out its ‘hateful content policy’ saying users ‘may not promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease.’ It also claims to stop users from displaying hateful images or symbols and expressing hate towards individuals or groups in display photos and online bios¹³. But Social media platform guidelines for acceptable behaviour have also been forced to become dynamic, as increasingly, their general and broad parameters are proving inadequate when it comes to tackling the prevalence, spread and impact of hate speech. In India, particularly, they are being forced to adapt to linguistic, religious or cultural sensitivities, as well as to watch against misuse by political propaganda. Facebook claims it has taught its artificial intelligence tools that detect hateful speech to new languages. The platform claims technology increased its capability to find and tackle hate speech before users report it by about 14 per cent during two quarters between 2018 and 2019¹⁴.

Most platforms have begun to prioritize the development of machine learning tools to catch violent, abusive or hateful content. YouTube — the video arm of Google where individual users are free to post video content on an open platform for wide consumption said in June 2019, that it would remove white supremacist videos from its site. It also claims to have made changes in its algorithms that ‘halved the views of videos deemed borderline content.’¹⁵ In fact, given government, media and civil society outcry against hateful content online platforms have had no choice but to respond by locking or suspending some accounts and removing some posts¹⁶. But users often complain that rather than taking down offensive posts, platform simply ‘shadow ban’ them- a term

¹² “Community Standards”, Facebook.<https://www.facebook.com/communitystandards/hate_speech>.

¹³ “Hateful Conduct Policy”, Twitter, <<https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>>.

¹⁴ Nieva, Richard, “Facebook’s New Rosetta AI System Helps Detect Hate Speech.” *CNET*.11-9-2018.<<https://www.cnet.com/news/facebook-s-new-rosetta-ai-system-helps-detect-hate-speech/>>.

¹⁵ “Our Ongoing Work to Tackle Hate,” Official YouTube Blog, <<https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html>>.

¹⁶ “Self-Regulation and ‘Hate Speech’ on Social Media Platforms,” ARTICLE 19, Rights, Equality and Citizenship (REC) Programme of the European Union, 2018. <<https://www.article19.org/resources/self-regulation-hate-speech-social-media-platforms/>>.

popularized by many supporters of US president Donald Trump¹⁷. This essentially means that platforms use their algorithms to make such posts less visible on line. In some cases, YouTube, which offers users a share of the advertising revenues each click on their video generates, was reported to have simply cut off the offending user's access to the share of revenues, and left posts untouched.

In fact, communications researcher and scholar Tarleton Gillespie argues that the idea of a 'neutral platform', or one that is able to or truly committed to a clean space online is a myth. Conflict drives the numbers of hits and interactions by attracting comment, criticism, even abuse. The more provocative, the better for traffic. Twitter, he says has allowed a culture of harassment, particularly against women to fester, largely unchecked. Explicit threats of rape and violence, doxing or releasing personal data of users on the platform were rampant, forcing Twitter to finally position the upholding of freedom of expression as the basis for more stringent protections against hate and abuse, thereby "allowing users to remain and opinions to flourish."¹⁸ Gillespie argues that social media platforms, especially in the United States emerged in a climate where they could, but were under no obligation to moderate content as they wished. In fact, early principles around the internet and its empowerment of the public expected that companies don't intervene at all. During the Arab Spring of 2011, Western public opinion marveled as the internet mobilized revolutions against authoritarian leaders across the Middle East. Companies prided themselves on having provided the platform for this mass mobilization. It is that legacy that has been hardest to give up, even though the levels of concern, the changing ethno-national or cultural tones around extreme speech have changed in the years. In fact in a seminal article in the journal *Foreign Affairs* in 2010, Google's Eric Schmidt and Jared Cohen prophesied the changes in the communications landscape as a result of mobile technologies and the rapidly growing internet, arguing that as states vie to control connection technologies to ensure economic and political power, the public could use it both as a force of good and ill¹⁹.

Tech companies argue that ultimately the solutions can only lie with machine learning- while context and language may be culture, religion on region specific, artificial intelligence can be developed to recognize some of those sensitivities. YouTube claims recent changes to its AI cut down views of videos

¹⁷ Jason Wilson, "What is 'Shadow Banning', and Why did Trump Tweet About It?". *The Guardian*, 27-7-2018 <<https://www.theguardian.com/media/2018/jul/26/what-is-shadow-banning-conservatives-twitter-trump>>.

¹⁸ Tarleton Gillespie, "Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media", Yale University Press, 26-6-2018 <<https://yalebooks.yale.edu/book/9780300173130/custodians-internet>>.

¹⁹ Eric Schmidt and Jared Cohen, "The Digital Disruption." *Foreign Affairs Magazine*, 27-10-2010. <<https://www.foreignaffairs.com/articles/2010-10-16/digital-disruption>>.

accused of spreading misinformation by half- but those videos are still up²⁰. Twitter with just a fraction of Facebook's active global user at 350 million globally, learned the hard way, albeit quickly, that prioritizing machine learning over human analysis was open to questioning. In February 2019, India's Parliamentary committee on Information Technology summoned its CEO Jack Dorsey to appear before it and respond to charges of a 'liberal bias'²². The ruling BJP led committee based its summons on a complaint by party supporters – many of whom had themselves been accused of increasing toxicity on the platform since the 2014 general elections in India that saw the Hindu right party stride unassailable to power in India²³. The summons was seen as nothing short of ironic, given the massive support Prime Minister Narendra Modi and his party enjoy on the social media platform. The summons was seen as a leaf directly out of President Trump's playbook- just four months before the Indian parliament committee's summons, White House officials in the United States had distanced themselves from President Trump's demand that federal investigative agencies look into social media companies for 'online platform bias.'²⁴ If anyone wondered why the BJP decided to go after a platform where it finds so much of its own support, the answer was simply that Twitter punches well above its weight. Its significantly smaller user base consists of world leaders, celebrities, journalists who often use their tweets to bypass an entire ecosystem of official communication, press statements and interviews and mass media to tell their stories - so much so that tweets are now considered primary news sources for reporters across global media. In India, Gillespie's argument has takers on both sides of the political/ ideological spectrum. Those at the receiving end of the most vitriolic abuse and dangerous misinformation are often the ones most critical of the ruling BJP, yet, as the summons to Dorsey showed, the platform's attempts to regulate that has led to the exact and opposite charge by those it seeks to hold accountable.

III. THE NEXUS: HATE SPEECH AND DISINFORMATION

In 2014, the World Economic Forum called misinformation one of the ten greatest perils confronting society. Computing technologies spread opinion,

²⁰ "Fighting Fake News: YouTube to Show 'Information Panels' on News-Related Videos." *The Economic Times*, 7-3-2019. <<https://economictimes.indiatimes.com/magazines/panache/fighting-fake-news-youtube-to-show-information-panels-on-news-related-videos/articleshow/68302365.cms?from=mdr>>.

²¹ Neima Jahromi, "The Fight for the Future of YouTube." *The New Yorker*, 9-7-2019. <<https://www.newyorker.com/tech/annals-of-technology/the-fight-for-the-future-of-youtube>>.

²² Maya Mirchandani, "The Parliamentary Panel will Reinforce Bias on Twitter, Not Fix It", *The Wire*, 12-2-2019 <<https://thewire.in/government/parliamentary-panel-twitter-right-wing-bjp-jack-dorsey>>.

²³ M.K. Narayanan, "The Future of Parliamentary Democracy". *The Hindu*, 26-6-2019 <<https://www.thehindu.com/opinion/lead/the-future-of-parliamentary-democracy/article28138347.ece>>.

²⁴ Ben Brody and Jennifer Jacobs, "White House Drafts Antitrust Order to Investigate Google and Facebook's Business Practices", *Time*, 22-9-2018 <<https://time.com/5403843/white-house-order-investigate-google-facebook-practices/>>.

propaganda, unverified information that masquerades as news quicker than we could have imagined. It serves to amplify hate and disinformation and creates the scope for its rapid, automated spread through bots that are learning to mimic human behaviour, and imitate legitimate users. It has become next to impossible to reconcile the need for speed with the need to verify information- leading to the crucible of misinformation bubbling over.

Hate speech has moved from the realm of the obvious to that of signaling, where the proliferation of what is now called borderline content is becoming a real challenge. The combination of both- social media influencers with significant follower bases, as well as automated bots have led to a crisis that is unfolding faster than it can be combated. Misinformation, disinformation and mal-information fall into this space too – where information manipulation and computational propaganda fuel ideological, political, cultural or religious polarization online²⁵²⁶. This whirlpool of unverifiable, incendiary information preys on grievance, insecurity and fear – both economic and cultural. It is intended to fuel public mistrust in each other. Even if much of it may not necessarily be ‘fake’ or false news, it is a manipulation, distortion and de-contextualisation of news and information that is leading to ‘coordinated inauthentic behaviour’²⁷²⁸ and making the free exchange of ideas for healthy debate so crucial for democracy, next to impossible. Falsehoods, rumor, real news disaggregated and put back together with the aim of feeding fear and diverting public attention from accountability is often geared to generate mistrust in the mainstream media sow hate division amongst the people.

Excessive coverage of fake news or misinformation only fuels and consolidates fiction as fact. That is why, ‘the amount of oxygen’²⁹ supplied becomes critical as it poses a threat to democratic institutions and media. Simultaneously, it is becoming essential to combat the spread of misinformation and fake news by upholding and strengthening journalistic rule to stand against these ‘newest forms of journalistic manipulation’. With access to technology there is a massive increase in information from every corner of the world, making it difficult for journalist

²⁵ Lisa Reppell and Erica Shein, 2019, “Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions”. International Foundation for Electoral Systems, April 2019 <https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf>.

²⁶ “Outlook on the Global Agenda 2014”. World Economic Forum. <http://reports.weforum.org/outlook-14/top-ten-trends-category-page/10-the-rapid-spread-of-misinformation-online/?doing_wp_cron=1572152017.2426369190216064453125>.

²⁷ “Removing More Coordinated Inauthentic Behavior from Iran and Russia | Facebook Newsroom”. 2019. Newsroom.Fb.Com, 21-10-2019 <<https://newsroom.fb.com/news/2019/10/removing-more-coordinated-inauthentic-behavior-from-iran-and-russia/>>.

²⁸ “Facebook Removes Nearly 700 Pages Linked to Congress Due to ‘Inauthentic Behaviour’”. *The Economic Times*, 1-4-2019 <<https://economictimes.indiatimes.com/news/elections/lok-sabha/india/facebook-removes-687-pages-linked-to-congress-party-due-to-coordinated-inauthentic-behaviour/articleshow/68669174.cms?from=mdr>>.

²⁹ Whitney Phillips, “The Oxygen of Amplification”. Data & Society, 22-5-2018 <<https://datasociety.net/output/oxygen-of-amplification/>>.

to cross-check, filter and remove, or ignore information that could be tangentially relevant to a story. Given the power of falsehood and misinformation to amplify rumor, ‘to amplify or not to amplify’³⁰ or ‘to report or not to report’,³¹ is often the key question driving (at least, independently run) newsrooms across India.

When in 2018, a spate of deaths by lynching that were the result of rumors about child kidnappers proliferating on WhatsApp³², the Indian public was finally forced to sit up and take a hard look at how it was becoming a part of the rumor factory. WhatsApp, currently with 400 million daily active users in India was forced to take steps to restrict user ability to forward messages and set up tools to indicate what was not original content. This action came after the Ministry of Information and Technology stepped in, threatening the platform with legal action³³, however the ministry itself was motivated only when the nature of such mob violence had no religious connotations, even though similar mob lynching deaths of Muslim or Dalit cattle rearers and traders have indicated a clear caste or religious bias. In several such cases, disinformation campaigns have used hate speech to drive existing social wedges deeper, with sometimes fatal impact.

IV. FREEDOM OF SPEECH V. HATE SPEECH

Hate has been and can be defined in several ways. A UNESCO report published in 2015 defined it as speech at “the intersection of multiple tensions. It is the expression of conflicts between different groups within and across societies.”³⁴ In a hyper-globalized world with ostensibly ‘borderless’ nation-states, the internet is acting as a facilitator for communication that transcends physical boundaries, and allow a free flow and exchange of ideas and information. However, as open and accommodative as the internet tries to be, the tightrope between the right to free speech and extreme, discriminatory speech is often hard to balance. It is this gap that companies, civil society and governments seek to regulate. In its report on the promotion and protection of the right to freedom of opinion and expression, the United Nations Human Rights Council recognized that the internet is “a key means by which individuals can exercise their right to freedom of opinion and expression³⁵,” but also highlighted that existing international human rights law has put in place several standards restricting this right, based

³⁰ *Ibid.*

³¹ *Ibid.*

³² Uma Sudhir and Anindita Sanyal, “Techie Beaten to Death by Mob in Karnataka After WhatsApp Rumours.” NDTV, 15-7-2018. <<https://www.ndtv.com/karnataka-news/techie-killed-3-injured-by-rumour-driven-mob-in-karnatakas-bidar-1883444>>.

³³ Kiran Rathee, “Fake News Controversy: WhatsApp Plans India Team as Govt Cracks Whip.” *Business Standard*, 5-8-2018. <https://www.business-standard.com/article/current-affairs/fake-news-controversy-whatsapp-plans-india-team-as-govt-cracks-whip-118080500007_1.html>.

³⁴ *Countering Online Hate Speech*, UNESCO, 2015 <<https://unesdoc.unesco.org/ark:/48223/pf0000233231>>.

³⁵ Art.19, *The Universal Declaration of Human Rights*, 1948 <<https://www.un.org/en/universal-declaration-human-rights/>>.

precisely on implications and intent. In India, the Supreme Court mandated TK Viswanathan Committee stated that “hate speech must be viewed through the lens of the right to equality, and relates to speech not merely offensive or hurtful to specific individuals, but also inciting discrimination or violence on the basis of inclusion of individuals within certain groups. It is important to note that it is the consequence of speech that is the determinative factor in interpreting hate speech, more so than even perhaps the content of the speech³⁶³⁷”. The closest understanding of hate speech by the judiciary is described in the 267th Report of the Law Commission of India, in which hate speech is defined as “an incitement to hatred primarily against a group of persons defined in terms of race, ethnicity, gender, sexual orientation, religious belief and the like³⁸”. It expanded the ambit of the definition to any word written or spoken, signs, visible representations within the hearing or sight of a person with the intention to cause fear or alarm, or incitement to violence,” in keeping with the Constitution.³⁹ The Law Commission report also took the attempt to define hate speech a step further - arguing that the status of both the author and the victim of the speech, and its potential to incite violence need to be factored as key criteria in the identification of hate speech. Therefore, context is important. In its study on counter speech on Facebook, the Observer Research Foundation pointed out that defining hate speech can be made harder by social and cultural implications - in fact it can be uniquely tied to the impact of the speech itself in situations where the vocabulary may not be explicitly hateful, but intent and implications are⁴⁰. It defined hate speech in the Indian context as expressions that advocated incitement to harm, discrimination, hostility and violence, based on targets being identified with a certain social or demographic group protected under the Indian Constitution. So, derogatory language in the context of caste or religious minorities which are punishable under the Indian Penal Code, are also considered grave transgressions on social media, and cases that seek intervention through the legal and criminal justice system appear frequently in the courts. Appeals have been made to platforms

³⁶ “Report No. 267, Hate Speech”, Law Commission of India, March 2017 (Hereinafter: Hate Speech, Law Commission, 2017) <<http://lawcommissionofindia.nic.in/reports/Report267.pdf>>.

³⁷ Amber Sinha, “New Recommendations to Regulate Online Hate Speech Could Pose More Problems Than Solutions”, *The Wire*, 14-10-2017. <<https://thewire.in/law/new-recommendations-regulate-online-hate-speech-problems>>.

³⁸ “Report No. 267, Hate Speech”, Law Commission of India, March 2017.

³⁹ Art.19(2) in the Constitution of India, 1950:

“(2) Nothing in sub clause (a) of clause (1) shall affect the operation of any existing law, or prevent the State from making any law, in so far as such law imposes reasonable restrictions on the exercise of the right conferred by the said sub-clause in the interests of the sovereignty and integrity of India, the security of the State, friendly relations with foreign States, public order, decency or morality or in relation to contempt of court, defamation or incitement to an offence” (Art. 19, Constitution of India, 1950).

⁴⁰ Maya Mirchandani, Dhananjay Sahai and Ojasvi Goel, “Encouraging Counter-Speech by Mapping the Contours of Hate Speech on Facebook in India,” Observer Research Foundation, 2018, <https://www.orfonline.org/wp-content/uploads/2018/03/ORF_Report_Counter_Speech.pdf>.

to recognize and address these transgressions through their own guidelines and community standards as well.

In India, hate speech specifically is not a restriction to speech under Article 19(2) of the Constitution but is usually read under its purview as “incitement to an offence”, “defamation” or “security state” etc.⁴¹ Unlike the USA and the UK where a “clear and present danger” with imminent and direct cause must be established, Indian law provides a wider ambit. The Indian Penal Code under Sections 153 A, 153 B, 295 A, 298, 505 (1) and 505 (2) aim to curb hate speech based on religion, ethnicity, culture or race. Under these sections, any form of speech that “promotes disharmony, enmity, hatred or ill-will” or “offends” or “insults” is liable to be punished.

This dilemma over the best way to regulate online platforms is compounded in several ways. As platforms become more proactive and act, warn or suspend abusive posts and accounts, users, especially trolls who engage in abuse online now often anticipate such responses and are careful with language. Worse, thin skinned political leaders today regularly trade accusations of hate speech in cases of criticism by political opponents^{42,43,44}. Charges of criminal defamation and sedition are regularly used against critics, and in well-choreographed offensives,^{45,46,47,48,49,50} online hate is targeted through social media at the same time,

⁴¹ Art. 19, Constitution of India, 1950.

⁴² Iain Marlow and Archana Chaudhary, “Senior Indian Politicians Censured for Hate Speech as Election Begins.” *Bloomberg*, 15-4-2019. <<https://www.bloomberg.com/news/articles/2019-04-15/senior-politicians-censured-for-hate-speech-as-india-polls-begin>>.

⁴³ Nimisha Jaiswal, Sreenivasan Jain and Manas Pratap Singh, “Under Modi Government, VIP Hate Speech Skyrockets - By 500%.” *NDTV*, 19-4-2018. <<https://www.ndtv.com/india-news/under-narendra-modi-government-vip-hate-speech-skyrockets-by-500-1838925>>.

⁴⁴ Manoj K, “Hate-Speech Accused 3 Times More Successful in Elections.” *The Wire*, 28-3-2018. <<https://thewire.in/communalism/hate-speech-accused-3-times-more-successful-in-elections>>.

⁴⁵ “Why India Activist Arrests have Kicked Up a Storm.” *BBC News*, 31-8-2018. <<https://www.bbc.com/news/world-asia-india-45294286>>.

⁴⁶ Ramachandra Guha, Twitter Post. 28-8-2018, 1.10 p.m., <https://twitter.com/ram_guha/status/1034344925112881152?lang=en>.

⁴⁷ Ramachandra Guha, Twitter Post. 28-8-2018, 5.13 p.m. <https://twitter.com/Ram_Guha/status/1034406178548137985?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1034406178548137985&ref_url=https%3A%2F%2Fwww.ndtv.com%2Findia-news%2Fafter-raids-on-activists-historian-ramchandra-guha-said-they-would-have-arrested-gandhi-1907612>.

⁴⁸ Rahul Gandhi, Twitter Post. 28-8-2019. 9.13 p.m. <https://twitter.com/RahulGandhi/status/1034466583614238720?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1034466583614238720&ref_url=https%3A%2F%2Fthewire.in%2Frights%2Fpolice-take-sudha-bharadwaj-into-custody-raid-homes-of-lawyers-activists-across-cities>.

⁴⁹ Shefali Vaidya, Twitter Post. 4-9-2019, 9:21 PM <<https://twitter.com/shefavidya/status/103700524787540377>>.

⁵⁰ Kavita Krishnan, Twitter Post. 10-11-2018, 4.13 p.m. <https://twitter.com/kavita_krishnan/status/1061207652967968768>.

creating an atmosphere of hostility towards critics⁵¹⁵²⁵³. Social media users often cite examples of hateful speech, accusing those who disagree with their views of “trolling,” and spewing verbal abuse and threats against them⁵⁴⁵⁵⁵⁶.

The power of signalling on social media is strong. In such cases, especially when the vocabulary itself is within acceptable boundaries of speech outlined by platform community standards, it is exactly this analysis of the impact of the speech, rather than the vocabulary itself that needs to be understood and identified as hateful- often making broader definitions of hate speech open to interpretation, and problematic. In October 2019, a Bihar court order directing the police to file an FIR against 49 eminent citizens who wrote an open letter to Prime Minister Modi expressing concerns over anti-minority mob lynchings, was overturned. The fact that a court thought it fit to advise the police to file charges under Section 124 of the Indian Penal Code of sedition, hurting religious feelings and insulting with intent to provoke a breach of peace in the first place points to this very abuse of legitimate concerns over hate speech and free speech by arms of the state. Sedition charges have a broad scope, and the intent of the FIR, rather than to perhaps prosecute, was to be a deterrent to future criticism, and to impose a sense of self-censorship. For example, multiple Facebook posts online have been taken down for being offensive in nature which were also directed at politicians. These users were often arrested under Section 66 (A) of the IT Act, which aimed to punish “offensive, false or threatening information” through electronic devices. Eventually the said section was taken down as it was deemed “arbitrarily, excessively and disproportionately invades the right of free speech and upsets the balance between such rights and the reasonable restrictions that may be imposed on such right.”⁵⁷ The defendant in this case received violent threats and thrashing of a relative’s clinic for her Facebook post directed at criticizing a bandh in Bombay. The complexities of multiple legal statutes, outdated laws, need for minimal governmental control, dynamic nature of social media and jurisdictional problems

⁵¹ Osama Manzar, “Hate Speech and the Role of Social Media.” Livemint, 9-2-2018, <<https://www.livemint.com/Opinion/ZAHBp4YDLpiBcCnlluwFON/Hate-speech-and-the-role-of-social-media.html>>.

⁵² Raphael Cohen-Almagor, “Balancing Freedom of Expression and Social Responsibility on the Internet.” Springer, 2017. <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6086244/>>.

⁵³ Imran Awan and Irene Zempi, “We Fear for Our Lives: Offline and Online Experiences of Anti-Muslim Hostility.” TellMAMA (Measuring Anti-Muslim Attacks), October 2015. < <https://www.tellmamauk.org/wp-content/uploads/resources/We%20Fear%20For%20Our%20Lives.pdf>>

⁵⁴ Swati Chaturvedi, *I Am a Troll: Inside the Secret World of the BJP’s Digital Army*. Juggernaut, 2016.

⁵⁵ Olga Jubany and Malin Roiha, “Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-Country Analysis .” Universitat de Barcelona and European Union, n.d.<http://www.unicri.it/special_topics/hate_crimes/Backgrounds_Experiences_and_Responses_to_Online_Hate_Speech_A_Comparative_Cross-Country_Analysis.pdf>.

⁵⁶ Mark Buchanan, “Hate Speech & Trolling! How Social Media is Pitching Us against One Another.” *Business Standard*, 5-9-2018. <https://www.business-standard.com/article/international/hate-speech-trolling-how-social-media-is-pitching-us-against-one-another-118090500833_1.html>.

⁵⁷ S. 66-A, Information Technology Act, 2000.

to identify perpetrators has made platform-based self-regulation effectively the best-case scenario.

In the vast grey zone between free speech and hate speech, the focus is squarely on the role of online platforms and their responsibility as providers of news and information. Media critics now argue that social media, even though it doesn't produce original content, has become a major source of unverifiable, often hate filled information, without subjecting it to the checks and balances of a regular newsroom. Their response has been consistent- that they are merely platforms and not content creators. In fact, they argue that the media as an influencer needs to revive its commitment to responsibility in the current climate. Determinations on the role of the media as an influencer and incriminatory, often defamatory language used against individuals are also imperative while addressing concerns around the spread of hate on social media.⁵⁸

During the onslaught of abuse against Jawaharlal Nehru University students leaders Kanhaiya Kumar and Umar Khalid, charged with sedition in 2016 for having led protests against the death sentence⁵⁹, former Jammu and Kashmir chief minister Omar Abdullah condemned the vitriol baying for Khalid's blood as a hate campaign using both social and mainstream media⁶⁰. The upsurge of violence and condemnation could be noted after a news debate on the popular TV channel *Times Now* where 'the news anchor called them "more dangerous to this country than Maoist terrorists" and anti-national'^{61,62}. Simultaneously, in an open letter to Arnab Goswami, civil rights activists 'stated their intention to boycott *Times Now*' as they felt the programme incited 'hate speech and demonised human rights activists'⁶³. A letter by civil rights activists to the network called out the channel's use of terms like 'Naxal', 'terrorist' and 'terrorist sympathiser', as a "gross misuse of the media's immense power"⁶⁴

⁵⁸ Devika Agarwal, "Hate Speech in India: Media's Rabble-Rousing Doesn't Help Cause, Proves Counter-Productive to Free Speech." Firstpost, 14-9-2018. (Hereinafter: Devika, Hate Speech in India) <<https://www.firstpost.com/india/hate-speech-in-india-medias-rabble-rousing-doesnt-help-cause-proves-counter-productive-to-free-speech-5182231.html>>.

⁵⁹ *The Indian Express*, "Who is Umar Khalid", 13-8-2018 <<https://indianexpress.com/article/who-is/who-is-umar-khalid-jnu-kanhaiya-kumar-5304561/>>.

⁶⁰ "Attack on Umar Khalid Fallout of Motivated Hate Campaign Against Him: Omar Abdullah." News18, 13-8-2017<<https://www.news18.com/news/politics/attack-on-umar-khalid-fallout-of-motivated-hate-campaign-against-him-omar-abdullah-1843227.html>>.

⁶¹ Mahapatra, Basudev, "Living in a Hate Culture: Hatred Conquers India's Civic Space." Indian Cultural Forum, 8-4-2019. <<https://indianculturalforum.in/2019/04/08/living-in-a-hate-culture-hatred-conquers-indias-civic-space/>>.

⁶² Devyani Onial, "JNU Row: TV Debates Where Anchors Played Judge and Jury." *The Indian Express*, 23-2-2016. <<https://indianexpress.com/article/india/india-news-india/jnu-row-umar-khalid-kanhaiya-kumar-anti-nationalism-tv-debates/>>.

⁶³ "Full Text: Arnab Goswami Violates Norms of Professionalism and Fairness, Say Activists in Open Letter." Scroll.in.,26-2-2015.<<https://scroll.in/article/709880/full-text-arnab-goswami-violates-norms-of-professionalism-and-fairness-say-activists-in-open-letter>>.

⁶⁴ Devika Agarwal, "Hate Speech in India: Media's Rabble-Rousing Doesn't Help Cause, Proves Counter-Productive to Free Speech." Firstpost, 14-9-2018. [Hereinafter: Devika, Hate Speech in

V. LEGAL FRAMEWORKS: CENSORSHIP OR REGULATION?

In India, the concept of freedom of speech is a relative, and not an absolute concept. However, every country has a different definition and interpretation of freedom of speech and expression. Also, most legal frameworks attempt to distinguish between social and legal concepts of freedom of speech, and therefore hate speech in order to tackle it. In the United States of America, free speech is interpreted as 'protected as an absolute and fundamental right under the First Amendment'. This extends to even protecting free expression like that of 'xenophobia, racism and religious discrimination' until and unless it furthers 'direct violence against an individual and/or a group'. On the other hand, European nations place more emphasis on 'civility and respect'. For instance, in Germany's Basic Law of 1949 "protection against insult" is a right and is also guaranteed in its first article.

It is precisely the understanding of varied contexts that leads to legal confusion over a standard regulatory norm for hate speech. On a spectrum between self-regulation and established, top down governmental regulation, platforms favour self-regulation, political leaderships who can potentially use regulation to stifle dissent lean towards law enforcement or government led regulatory mechanisms, but civil society which populates social media platforms, and uses its support of absolute free speech for both benign and malign objectives, fall somewhere in between- seeking a multi-stakeholder approach with the possibility of an autonomous body to monitor these spaces- these could be global, or national.

An array of laws currently regulates media in India. Television is regulated by Cable Television (Networks and Regulations) Act, 1995, that empowers the local Police Commissioner and local magistrate to take action against the telecast of obscene material, programmes that are likely to incite disorder or violence, or against material likely to cause class hatred, and is defamatory. Newspapers fall under the purview of the Press Council Act, 1978 which empowers an industry body to regulate newsprint content. Under this statute, the courts have the power to punish violators for contempt. The Censor board, which was constituted under the Cinematographic Act, 1952, regulates films and the grounds for censorship are similar to the grounds under Article 19 (2) of the Constitution of India. Offences relating to obscenity, defamation, disruption of public order or national security are punishable under both the Indian Penal Code and the Information Technology Act, 2000. The IPC suggests punitive action against speech that promotes enmity based on religion, race, etc.⁶⁵ It also penalizes speech that deliberately hurts religious sentiments⁶⁶. Sedition⁶⁷ and statements contributing to

India] <<https://www.firstpost.com/india/hate-speech-in-india-medias-rabble-rousing-doesnt-help-cause-proves-counter-productive-to-free-speech-5182231.html>>.

⁶⁵ S.153-A, Penal Code, 1860.

⁶⁶ S.298, Penal Code, 1860.

⁶⁷ S.124-A, Penal Code, 1860.

Public mischief⁶⁸ are also punishable under IPC. Representation of The People Act⁶⁹, CRPC⁷⁰, Religious Institutions (Prevention of Misuse)⁷¹ and the Scheduled Caste and Scheduled Tribe (Prevention of Atrocities) Act⁷² all aim to curtail Hate Speech directed at specific groups. The Supreme Court⁷³ also sought analysis from the Law Commission of India of the linkages between hate and extreme speech and election integrity. The commission in response argued that the highest standard of restriction must be employed against free speech⁷⁴ under Article 19 (2) the said “speech” must have a direct and proximate connection with “threat to public order”. The report also recognized the reluctance of the lawmakers to create exceptions further by defining Hate speech as it might curtail freedom of speech - both a bulwark and a cornerstone of any democracy⁷⁵.

It is essential to uphold the power of a free internet in a democracy as it plays an imperative role in decentralizing discourse by empowering individuals to criticize policy and hold power to account, freely. Efforts to curb hate extreme speech must therefore necessarily be employed within a defined framework in conjunction with existing (print and broadcast) industry practices.

Self-regulatory bodies like Indian Broadcasting Federation (IBF) and News Broadcasters Association (NBA) have often played an active role in regulating sensitive content. In *Destruction of Public and Private Properties v. State of A.P.*⁷⁶, the Supreme Court accepted the F.S. Nariman Committee recommendations on regulation of media. These recommendations included principles of self-regulation for both broadcasters and the industry- based on impartiality, neutrality, objective and responsible reporting of sensitive issues especially crime, violence, agitations and protests. The recommendations urged respect for privacy, national security and sensitivity around reportage on women and children in particular. A jury of peers is expected to adjudicate on violations of industry standard. But technology has overtaken such a model. Speed and scope- that lead to the immediate scalability of content is challenging all existing norms of engagement. Under the circumstances, the extension of an autonomous industry body to regulate contents and/or censure violations to a similar code are the only viable way forward, if the balance is to be maintained. This naturally means it is important for platforms to feel equally involved, with as much at stake as anyone else. It is only in the widest accessibility of news and information also, that a balance between free speech and censorship can and will be maintained.

⁶⁸ S.505, Penal Code, 1860.

⁶⁹ The Representation of the People Act, 1951 (Act No. 43 of 1951)

⁷⁰ The Code of Criminal Procedure, 1973, (Act No. 2 of 1974).

⁷¹ Religious Institutions (Prevention of Misuse) Act, 1988, (Act No. 41 of 1988).

⁷² Scheduled Castes and Scheduled Tribes (Prevention of Atrocities) Act, 1989, (Act No. 41 of 1989).

⁷³ *Pravasi Bhalai Sangathan v. Union of India*, (2014) 11 SCC 477 : AIR 2014 SC 1591.

⁷⁴ Art. 19(1)(a), *Constitution of India*, 1950.

⁷⁵ Report No. 267, “Hate Speech”, Law Commission of India, March 2017.

⁷⁶ *Destruction of Public and Private Properties v. State of A.P.*, (2009) 5 SCC 212 : AIR 2009 SC 2266.

VI. INTERMEDIARY LIABILITY

As media professionals and civil society demand a multi-stakeholder approach to regulation, the fight on this front as it stands between platforms and government is now honing in on one particular battle- of Intermediary liability. Multiple government led institutions have come up with different policies to identify what this would look like, and regulate online content on tech platforms.⁷⁷ In fact, just recently, in October 2019, the Union Ministry of Electronics and Information Technology proposed amendments to the Intermediary Rules, 2011 under Section 79 of the IT Act.⁷⁸ These rules provide a mechanism for removing “unlawful” content on an Intermediary’s platform.⁷⁹ They also extend “safe harbour protection” - making it understood that the intermediary is merely a platform provider and does not play any role in content creation or modification of data and information.⁸⁰ The aim of new amendments⁸² is to ensure platforms don’t shirk their own responsibility. Monthly warnings, providing traceability of users, keeping data for 180 days or more, “identifying and removing or disabling public access to unlawful information or content” and establishment of local offices registered in India are all part of the proposed regulations on intermediary liability, but in a vitiated political environment, it is unclear if they will come through. As a current proposal for Social Media regulation is underway,⁸³ there are doubts over what will be deemed “unlawful” content, given the opacity of definitions.⁸⁴ The idea of leaving interpretation to the whim and fancy of government is anathema to lawyers, activists and the media. The expectations of

⁷⁷ Intermediaries facilitate exchange of information between people without moderation of their data, these include social media firms and e-commerce marketplaces.

⁷⁸ S. 79(2)(c) of the Information Technology Act, 2000, “the intermediary observes due diligence while discharging his duties under this Act and also observes such other guidelines as the Central Government may prescribe in this behalf”.

⁷⁹ S. 79(3)(b) of the Information Technology Act, 2000, “(b) upon receiving actual knowledge, or on being notified by the appropriate Government or its agency that any information, data or communication link residing in or connected to a computer resource controlled by the intermediary is being used to commit the unlawful act, the intermediary fails to expeditiously remove or disable access to that material on that resource without vitiating the evidence in any manner”.

⁸⁰ S. 79 of the Information Technology Act, 2000, “an intermediary shall not be liable for any third-party information, data, or communication link made available or hosted by him”.

⁸¹ In *Shreya Singhal v. Union of India*, (2015) 5 SCC 1, (S. 505, IPC) the Supreme court held that any content removal from an intermediary platform could only be permitted by a court or government agency [The Representation of The People Act, 1951 (Act No. 43)]. In another judgement, the Supreme Court has clarified that a restriction on speech, in order to be reasonable, must be narrowly tailored so as to restrict only what is absolutely necessary [The Code of Criminal Procedure, 1973, (Act No. 2)].

⁸² The Information Technology [Intermediaries Guidelines (Amendment) Rules], 2018 : <https://www.prsindia.org/sites/default/files/bill_files/Draft_Intermediary_Amendment_2018.pdf>.

⁸³ “Social Media Regulations to be Ready by January 2020: Centre to Supreme Court.” [⁸⁴ R.9, The Information Technology \[Intermediaries Guidelines \(Amendment\) Rules\], 2018, “The Intermediary shall deploy technology based automated tools or appropriate mechanisms, with appropriate controls, for proactively identifying and removing or disabling public access to](https://www.livemint.com/Livemint, 21-10-2019. <https://www.livemint.com/news/india/social-media-regulations-to-be-ready-by-january-2020-centre-to-supreme-court-11571664731405.html>.”</p>
</div>
<div data-bbox=)

divulging individual user data and decryption of platforms like WhatsApp that promise privacy and end to end encryption in order to trace originators of messages threatens both privacy and free speech.⁸⁵ Most platforms are compliant with the new The Global Data Protection Regime (GDPR) that is designed to protect individual user information, create awareness and stop targeted messaging, and such proposed regulations are likely to run afoul of the GDPR, thereby opening up- potentially- a whole new area of conflict which diplomacy will have to tackle.

VII. CONCLUSION

The shooting at a mosque in Christchurch, New Zealand by white supremacist Brendan Tarrant in March 2019 was a wakeup call for everyone. The power of technology to spread hate via livestream was reminiscent to what we saw in India two years prior - when Mohammed Afrazul, a Bengali Muslim migrant worker in Rajasthan was hacked to death with a meat cleaver by a right wing Hindutva worker named Shambhulal Regar. His body was then burnt at the scene of the murder. Regar had the entire incident videotaped, and put it up on YouTube along with a sermon against 'love jihad.' The video went instantly viral and managed to raise money for his wife and family through advertising before it was caught and removed.⁸⁶ Regar was celebrated by fellow members of the Vishwa Hindu Parishad.⁸⁸ The police said that Regar's actions were individual, that he was a 'lone wolf hate killer'⁸⁹ but his language in subsequent inter-

unlawful information or content" <https://www.prsindia.org/sites/default/files/bill_files/Draft_Intermediary_Amendment_2018.pdf>.

⁸⁵ "Proposed Changes in IT Rules will Lead to Over-Censorship, Undermine Encryption, Warns Mozilla." *The Economic Times* 3-1-2019. <<https://economictimes.indiatimes.com/tech/internet/proposed-changes-in-it-rules-will-lead-to-over-censorship-undermine-encryption-warns-mozilla/articleshow/67366097.cms>>. ("Mozilla policy adviser AmbaKak: The proposal "calls into play numerous fundamental rights and freedoms guaranteed by the Indian constitution," and "Whittling down intermediary liability protections and undermining end-to-end encryption are blunt and disproportionate tools that fail to strike the right balance.")

⁸⁶ "Video: Man Burnt Alive in Rajasthan's Rajsam and after Being Axed for Alleged Love Jihad." *The Financial Express*, 7-12-2017. <<https://www.financialexpress.com/india-news/for-alleged-love-jihad-man-hacked-by-axe-burnt-alive-in-rajasthan-caught-on-camera/964351/>>.

⁸⁷ "Ram Navami Tableau Honours Shambulal Regar, Man Who Killed Muslim Labourer on Video." *News18*, 27-3-2018. <<https://www.news18.com/news/india/ram-navami-tableau-honours-shambulal-regar-who-killed-muslim-labourer-on-video-1701467.html>>.

⁸⁸ Vishwa Hindu Parishad, or VHP as a conservative Hindu nationalist organisation that abides by the ideologies of Hindutva and is often characterised as "militant" for initiating anti-social activities like the Ram Janmabhoomi movement that resulted in demolition of the Babri Masjid. The group was founded by M. S. Golwalkar and S. S. Apte in 1964. It operates on ideology which is "to organise, consolidate the Hindu society and to serve, protect the Hindu Dharma." (Maya Mirchandani, "Digital Hatred, Real Violence: Majoritarian Radicalisation and Social Media in India." ORF, 29-8-2019. <https://www.orfonline.org/research/43665-digital-hatred-real-violence-majoritarian-radicalisation-and-social-media-in-india/#_edn9>.)

⁸⁹ Harsh Mander, John Dayal and Kavita Shrivastava, "Rajasthan Hate Murder: The Other Tragedy in Afrazul's Killing is a Famine of Compassion, Outrage." *Scroll.in*. 18-12-2017. <<https://scroll.in/article/861826/rajasthan-hate-murder-the-other-tragedy-in-afrazul-khans-murder-is-a-famine-of-compassion-outrage>>.

views and his manifest use of social media to disseminate his hate filled video and message was alarming. Tarrant was Regar in another form, on another continent- both killers united by extreme ideology, hate, violence, and their savvy with social media. The live streaming of the shooter's video, taken with a gopro camera attached to his cap showed the world the raw, brute power and violence of extreme ideologies fuelled by online propaganda and hate. Just as Tarrant's network online proved that support for White supremacy has flourished in both open and closed spaces on social media for years, protected by the right to absolute free speech, so has militant, extreme right wing Hindutva ideology found a platform. And irrespective of both law enforcement and civil society concerns, these platforms seem unable to tackle the speed and scale of their spread, are being manipulated by identity politics and a manufactured sense of offense and offendedness.

The constant validation of extreme speech by political, religious, cultural groups may generate the numbers they seek, but there is a cost, as they also serve to embolden bad actors who disregard community standards knowing that both the right to free speech and ever changing technology will protect them from censure. Autonomous bodies are often critiqued as having no real teeth.⁹⁰ Corporate media- increasingly eyeing the generation of high traffic on their websites- are susceptible to external, political pressures.⁹¹ Machine learning tools and Artificial Intelligence are increasingly picking up language, re-engineering sharper, quicker image detection technology and monitoring content, but the volume is often too much for even the machines to handle, and local contexts next to impossible for them to fathom. Both proportionality and rationality must be driving factors while attempting to regulate any form of media. And, in fact, it is also up to us, the users to ask whether we desire more government regulation, or whether we advocate for social media agencies to tighten their policies? Or, are we hoping for public-private collaboration in this realm?

Responsible broadcasting and institutional arrangements as a result of consultations between social media platforms, media industry bodies, civil society and law enforcement are an ideal regulatory framework, and a possible way forward for like a Code of Conduct or an independent regulatory body for self-regulation can be appointed. This must be achieved without creating an ambiguous statutory structure that could leave avenues for potential legislative and state control. Australian and Germany have some of the strictest laws and regulation over social media wherein they have had imposed fines and imprisonment for 'inaction on extremist hate speech' within tight time frames. Simultaneously, the European

⁹⁰ Krishnadas Rajagopal, "SC to Examine Powers of EC to Curb Hate Speech." *The Hindu*, 15-4-2019. <<https://www.thehindu.com/news/national/sc-to-test-eci-claim-of-toothlessness-against-hate-speech/article26841877.ece>>. ("Election Commission, an independent body, was pronounced "toothless and powerless against hate speeches" by the Chief Justice Gogoi while listening to a case where senior political leaders were accused of Hate Speech").

⁹¹ "Digital Globalization: The New Era of Global Flows." McKinsey Global Institute, 2016.<https://www.mckinsey.com/-/media/McKinsey/Business_Functions/McKinsey_Digital/Our_Insights/Digital_globalization_The_new_era_of_global_flows/MGI-Digital-globalization-Full-report.ashx>.

Union has also established a code of conduct to ensure non-proliferation of hate speech under the framework of a ‘digital single market.’⁹² The United Kingdom recently published a paper titled ‘Online Harms’, that establishes what it calls the ‘duty of care’⁹³. Most other countries have been struggling between threats to tech companies and platforms and law enforcement responses against users that aren’t anonymous, but given the transnational nature of technology, a coherent global regulatory standard would ensure a uniformity of governance and regulation, while keeping specific local contexts in mind. With new platforms like TikTok⁹⁴ both the means of expression online and social responsibilities are constantly evolving.

The last decade has seen an exponential rate of growth in the technology sector. Smartphones today are virtually anatomical appendages to each one of us. Affordable gadgets, cheap data and the growth of languages on the internet have led to the emergence of a media discourse that now is inextricably linked to the ease of digital technology. Technology has erased rural-urban divides and transcended boundaries and barriers, make development and governance achievable targets in a country as large and diverse, with areas as remote as India is. But it has also had an equal and opposite reaction. As technology teaches us to be borderless, it provides a vehicle to rally around identity, grievance and fear- all exploited by savvy politicians out to garner votes⁹⁵.

However, as this paper argues, over-criminalizing speech may have troubling outcomes. It is the crossroads between technology, profit, freedom, politics, identity, power and insecurity that any effort to regulate hate speech on social media will have to traverse. The need is to seek redress and action, not censorship. But the means towards this end have to be independent of government control. Arriving at a regulatory mechanism is unlikely to be easy- a fact that is evident in the ongoing global debate, as different national governments cite their own legal frameworks to set precedent and seek control. But as hard as it may be, any regulatory framework that evolves will necessarily need to ensure that it not only protects the right to free speech in a democracy, but equally if not more so- creates safeguards and curbs against the impact and the process of online, social media amplification of hate speech that can lead to offline, real world violence.

⁹² “Code of Practice on Disinformation.” Digital Single Market - European Commission, 17-6-2019. <<https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>>.

⁹³ “Online Harms White Paper.” Department for Digital, Culture, Media & Sport, Government of UK, 12-2-2020. <<https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper>>.

⁹⁴ Tiktok is an app which facilitates creation and sharing of small videos. The app was launched for an international audience excluding China in 2017. It is funded by the Chinese firm, Bytedata.

⁹⁵ Amulya Gopalakrishnan, “A Climate of Fear: It Helps Autocracy – but Those Who Disagree Should Persist Past Our Cowardice.” *Times of India Blog*, 4-9-2019. <<https://timesofindia.indiatimes.com/blogs/toi-edit-page/a-climate-of-fear-it-helps-autocracy-but-those-who-disagree-should-persist-past-our-cowardice/>>.

Annexure: The Facebook post below directed at Hindu girls says, “This is not love dear sister, they pretend to love you and convert you and use you for terrorist activities. The girl called Deepa Cherian became a terrorist and went to jail through this kind of love jihad.” It ends with an advertisement for a “Hindu Helpline.”⁹⁶



⁹⁶ “Facebook India: Towards the Tipping Point of Violence Caste and Religious Hate Speech.” Equality Labs, 2019. <<https://www.equalitylabs.org/facebook-india-report>>.