MODERATING SOCIAL MEDIA IN INDIA: USER-GENERATED CONTENT IN AN ERA OF VIRAL FALSE NEWS, DISINFORMATION AND HATE SPEECH

January 2021 | *Issue No. 006*

ESYA
centre

ABOUT THE AUTHOR

Megha Bahree is a Fellow at the Esya Centre.

ABOUT THE ESYA CENTRE

The Esya Centre is a New Delhi based technology policy think tank. The Centre's mission is to generate empirical research and inform thought leadership to catalyse new policy constructs for the future.
It aims to build institutional capacities for generating ideas that will connect the triad of people, innovation, and value to help reimagine the public policy discourse in India. More details can be found at www.esyacentre.org.

Design Illustrations by Taniya O'Connor, Design by Drishti Khokhar.

TABLE OF CONTENTS

THE INDIAN CONTEXT

Social media globally, and in India, is widely afflicted by two main problems: hate speech and false news. The reason this is a pressing problem is because eventually both these elements end up hurting democracy. At Esya Centre we have taken an in-depth look at the social media ecosystem in India to identify the problems and come up with potential solutions.

While India has multiple bills in the works that aim at tackling some of these issues, the ambit of those bills is very broad and encompasses multiple themes. We've kept our paper focused on social media and how to moderate user-generated content.

For this paper, which is not sponsored by any social media company, we spoke with a range of companies, lawyers, researchers, and academicians in the ecosystem. We also studied developments in the U.S. and Europe and adapted from there suggestions that we think will help improve the ecosystem in India without killing business. However, this research does not reflect anyone else's opinions.

AS PER INDIAN LAW, SOCIAL MEDIA PLATFORMS ARE CONSIDERED INTERMEDIARIES.

As per Indian law, social media platforms are considered intermediaries. Section 2 (w) of the IT Act, 2000 defines intermediaries as any person who on behalf of another receives, stores or transmits any electronic record or provides any service with respect to that record and includes telecom service providers, network service providers, internet service providers, web-hosting service providers, search engines, online payment sites, online auction sites, online market places and cyber cafes.

The gamut of who qualifies as an intermediary is clearly very vast and we have kept our focus on social media companies.

At the outset we'll say that while new challenges like disinformation, revenge porn and deep fakes among other issues are spreading on, and via, social media, we don't believe that the platforms should be stripped of their safe harbor protections. It's important to strike a balance between platform and end-user interest and that equilibrium is maintained under section 79[1] of the Indian IT Act 2000 which not only lays out the protections to platforms but also has caveats to that protection to ensure that they don't shirk their responsibilities.

We think the focus should be on improving and strengthening their systems and processes rather than making them vulnerable to attacks from everywhere. Moreover, to chase the millions of pieces of content that get uploaded daily is also practically impossible. At the same time, we believe it's important to protect freedom of expression—it's the cornerstone of a democracy. This is a delicate balance. First, the easy bit.

THE PROBLEM

ILLEGAL CONTENT, HATE SPEECH AND FALSE NEWS

There shouldn't be any debate about removing obviously illegal content. Terrorist recruitment, child sexual abuse material, violent content are all illegal content. One of the more horrific instances of this in recent times took place in March 2019 when a gunman killed 51 people in two mosques in Christchurch in New Zealand and livestreamed his actions on Facebook. In the wake of that the social media company modified its rules to at least temporarily block users who break its rules[2] from broadcasting live video.

Then there are other problematic content which are dangerous in a more insidious manner. Colloquially referred to as fake news, this category also includes misinformation and disinformation – the difference between the two is that the latter is driven by an intent to deceive, by an agenda.

A recent example of disinformation is how social media in Europe was littered with very strongly opposing views in the lead up to the referendum in 2016 for the U.K. to leave the European Union—the so-called Brexit. A 2017 research paper[3] from academics at Swansea University and the University of California, Berkeley found that information automated software agents or 'bots' were used to spread either 'leave' or 'remain' social media stories during and after the Brexit referendum which drove the two sides of the debate further apart.

MISINFORMATION IS WHEN INCORRECT INFORMATION IS UNKNOWINGLY SHARED BY USERS

Misinformation is when incorrect information is unknowingly (as in, the user doesn't know it's false news) shared by users.

In this paper we have deliberately replaced the term "fake news" with false news. With President Donald Trump labeling any and all news stories that he disagreed with as "fake news", many researchers, and media organisations, are staying away from that term and we are following that practice.

Then there's a third category, that of hate speech. This tends to go viral on social media platforms. The question that researchers in the U.S. and the EU are grappling with is how to moderate such content.

FALSE NEWS AND HATE SPEECH TEND TO GO VIRAL ON SOCIAL MEDIA AND DRAW IN MORE USERS

At this point it is critical to understand the silent role of false news and hate speech in driving the business of platforms. This will be a crucial point to keep in mind while evaluating how responsive and responsible platforms are in tackling such content.

Simply put, false news and hate speech tend to go viral on social media[4] and draw in more users. (Mark Zuckerberg himself admitted in a 2018 memo[5] that "people will engage disproportionately with more sensationalist and provocative content.")

This is not to say this is the only kind of content that goes viral, of course, or that platforms are actively encouraging it. Not at all. It is a combination of user interests (what were the posts she liked/commented upon/shared) which is amplified by the platform's algorithm[6] which will show more of that kind of content, encouraging the creation of a little bubble. For instance, in the final three months leading up to the 2016 U.S. elections, the top-performing fake election news stories on Facebook generated more engagement[7] than the top stories from major news outlets.

Platforms, for their part, encourage such user engagement as it gives them access to information about their users-demographics, what are their likes/dislikes (based on the content they share/like/click on), information they can offer to advertisers[8] for them to target users more precisely.

In other words, viral content, including conspiracy theories[9] that can actually harm people—such as a May 2020 video on the novel coronavirus titled Plandemic, which, among many false statements, said that wearing a mask would activate the virus—helps drive business for social media companies.

ONE OF THE TASKS FOR POLICY MAKERS IS TO LOOK AT HOW CONTENT IS TARGETED AND AMPLIFIED

Hence, one of the tasks for policy makers is to look at how this content is targeted and amplified or how it can be made more transparent.

One way to go is to follow the path of the EU's privacy law, the General Data Protection Regulation (GDPR), which places restrictions on what data companies may acquire, and once collected, how those companies may use the data. If a social media platform is using the data it collected for a purpose different from what it stated to its users, that is a clear problem. That's what happened in the case of the Cambridge Analytica scandal[10], for instance, when it got data improperly from Facebook and used that to build voter profiles.

Another approach could be to bring in transparency in the content/advertisements themselves. Meaning, make it clear if something is a paid ad, who has paid for it (this is especially relevant when it comes to political ads), how much was paid for it (again, relevant for political ads because of limits on election spends). Similarly, if a piece of content is actually paid content, such as when an influencer may be pushing a product, it should be made clear that it's an ad and not a regular user's post. It's important for users to be aware of what is paid content, especially when it comes to marketing of food and drink or health-related products for consumption.

THERE IS NO ROOM FOR HATE SPEECH IN A PLURALISTIC DEMOCRACY LIKE INDIA

Another potential debate one can get caught up in when it comes to moderating controversial content is that of hate speech vs. freedom of expression. While we firmly believe in the latter, and we think it's important to hear views other than your own, there is no room for hate speech in a pluralistic democracy like India.

The good part here is that India has several laws that draw the line beyond which speech constitutes hate speech and which clearly state that promoting communal disharmony or feelings of hatred between different religious, racial, language or regional groups or castes or communities is a criminal offence.

Here are some of the sections of the Indian Penal Code that make such behaviour a punishable offence:

295 (A)[11] (deliberate and malicious acts intended to outrage religious feelings of any class by insulting its religion or religious beliefs)

Section 153[12] (wantonly giving provocation with the intent to cause riot)

153(A)[13] (promoting enmity between different groups on grounds of religion, race, place of birth, residence, language, etc., and doing acts prejudicial to maintenance of harmony)

Section 298[14] (uttering words etc. with deliberate intent to wound the religious feelings of any person)

Section 505 (1) and (2)[15] (make any statement or rumour with intent to incite, or which is likely to incite, any class or community of persons to commit any offence against any other class or community)

WE DON'T WANT BUSINESSES DETERMINING THE STANDARD OF FREE SPEECH

At the same time, we don't want businesses determining the standard of free speech that we have and it's best to follow the law of the land on this. India has some clear rules on this thanks to a 2015 ruling by the Supreme Court of India in the Shreya Singhal v. Union of India[16] case in what is considered a watershed moment for online freedom of speech in the country. In that case the court ruled that restriction on online speech, under section 66(A) of the IT Act 2000, was unconstitutional as it violated the freedom of speech as guaranteed under the Indian Constitution as it was vague and over-broad and could have "a chilling effect on free speech"[17]. It also said that online intermediaries would only be obligated to take down content on receiving an order from a court or government authority.

For everything else social media platforms have their own terms of service (TOS). That apart, we also recommend they work with multiple stakeholders, including civil society and technologists, to draw up guidelines on how to moderate content. (It's important for civil society organisations to disclose their source of funds and any potential sources of conflict to ensure it doesn't become an agenda-driven situation.) Both the TOS and the suggestions via stakeholders should be within the Indian law. Meaning Article 19(1)(a)[18] of the Indian Constitution gives all citizens the right to freedom of speech and expression and Article 19(2)[19] outlines the reasonable restrictions to free speech. So any TOS, as well as other content moderation guidelines, need to be within these laws because the moment they start to moderate/censor beyond that, it becomes a threat to freedom of expression and a violation of the law.

PLATFORMS SHOULD NOT REMOVE CONTENT UNDER PRESSURE FROM ANY CORNER

Similarly, platforms should not remove content under pressure from any corner, including, and especially, government authorities and political parties, unless it is specifically in violation of the law or their TOS.

While it's not possible to track each piece of content, and that's not the goal here either, what's important to check is how platforms deal with such incidents, what is the room (and process) for redressal for any affected party, how consistent are they in their approach and how transparent are they in their disclosures.

WHAT PLATFORMS CAN DO

If we focus on the activities of platforms, that will help us focus on where the content is being created rather than chasing billions of pieces of content across the internet. In a nutshell, transparency and regulation should be the two pillars for the functioning of platforms.

TRANSPARENCY

This is crucial as it helps establish trust between the platform as well as the government and the user, the public. To establish that, platforms should regularly—ideally once a quarter or at the very least once in six months—disclose the number of takedown requests they got, under which article/offence of the law, by which entity (government/law enforcement/individual etc.) and how did they respond to those demands.

Transparency also helps as it offers data to make evidence-based policy. At the same time it's also important to be transparent about why some troubling content was not removed as that helps establish trust in the platform and the ecosystem.

This may sound like a tall order but it's not. Some of these companies already do at least some of this in the U.S.—not because it's required by law, but rather because it's required by the business environment. There's pressure from users, civil society organisations, government and employees to be transparent and the companies at least attempt to appear to do that in their home market.

For instance, in 2018 New America's Open Technology Institute, as part of a coalition of organisations, advocates, and academic experts who support the right to free expression online, released the Santa Clara Principles on Transparency and Accountability Around Online Content Moderation[20]. These lay out the minimum standards tech platforms must meet in order to provide adequate transparency and accountability on their efforts to take down user-generated content or suspend accounts that violate their rules, provide meaningful due process to impacted speakers, and ensure that the enforcement of their content guidelines is fair.

In 2019 on the first anniversary of the principles it evaluated[21] the work of the three biggest platforms—Facebook, YouTube and Twitter—on the basis of these principles. It found that although the three platforms had made greater progress in implementing the recommendations related to their "notice" and "appeals" efforts, they fell "woefully short" when it comes to meeting the standards set forth for the "numbers" category, the report said. That granular information would have been useful for researchers to understand and evaluate the scope and scale of the content moderation efforts of the social media firms, how much were they enforcing their TOS and how that was impacting their user speech.

In recent history some of these companies have also deployed outside organisations to review complaints and then fully disclosed their findings, even when they were not complementary, such as a civil rights audit[22] that Facebook commissioned an independent attorney to carry out.

Similarly, this year during the ongoing COVID-19 pandemic social media companies have reacted fairly swiftly to disinformation with things like posting info boxes with links to trusted organisations, removing apps that were spreading COVID-19 disinformation, and even deleting misleading tweets from major political figures.

In a first-ever, Facebook in Myanmar—where it has a very troubled history[23]—in the run-up to elections in November, limited users from sharing old pictures without context (users get a warning message that the pictures they want to share are violent and more than a year old), a common trick to spread misinformation, and worked with local partners to verify the pages of political parties and fact check information there. It also expanded its community standards[24] that it uses to police content to include rumours that could have impacted the voting process. These practices should be extended to other markets during their elections as well, including for the state elections in India.

"PICTURES ARE A HUGE PART OF THE PROBLEM OF MISINFORMATION"

Pictures are a huge part of the problem[25] of misinformation and Facebook's actions in Myanmar show that it is possible to tackle that. Another idea to be explored is to put a date (at least the year), and if possible, location in a photo, similar to way a lot of news organisations now tag old stories to say which year they are from. That way an old photo of a mob, for instance, cannot be used in an ongoing volatile situation and that will be a massive step towards curbing that problem.

Companies can explore a combination of technology and human intervention to tackle this. For instance, technologies like artificial intelligence (AI) can be used to tag photos which can then be flagged to fact checkers/moderators who work with these social media companies to identify if the photos are, indeed, problematic or being misused. Using just AI to automate the removal of problematic content may lead to an over the top approach and would come with its own problems because of the biases built into that technology and also its limitations.

For instance, in 2019 a stream of news articles focused on how in a remote corner of Tamil Nadu a few TikTok users[26] of two different castes were making videos attacking each other, fueling hate speech. While the company said at the time that it had moderators for 15 Indian languages, for a country like India that's not enough. Language is one issue. The other, tougher one, is the nuances of caste, class, culture, symbols associated with each of these. For instance, in the Tamil Nadu case one video used a visual of slippers to attack a member of the other, Dalit caste. Neither AI nor a human moderator who is not well versed in the nuances of that caste, class and language can catch these situations and it shows that companies need to make vast investments to tackle these problems in India if they are serious about doing so.

It's important to remember that the culture of platforms comes from the platforms themselves. They need to decide what they want to stand for and build the terms of service around that. If they make an example of a few cases, it will help set the tone for users on what's acceptable and what isn't. For instance, after years of allowing President Trump a free reign on Twitter, the company last year introduced measures like labels, warnings, and retweet restrictions before finally banning him from the platform[27], setting the example that even high-profile and powerful politicians will be held accountable for their tweets.

Then there's Discord, a social media video-and-voice chat app that was launched in 2015 and has 15 percent of its employees as part of its trust and safety team[28]. It also allows moderators within groups—regular users and not just employees—to report bad behaviour and to add bots to scan for offending language. Like any social media company, the success of this approach would depend on how stringently the company applies it and it's an idea worth exploring.

Some other measures that platforms can adopt include pointing to sources that are reliable and trusted (as some platforms have done during the Covid-19 pandemic because people want reliable news). In cases where platforms are hosting content but not creating it and don't want to get in to the business of curating it, they can appoint third party fact checkers for this work.

ANOTHER OBVIOUS SOLUTION: MAKE RELIABLE NEWS MEDIA MORE VISIBLE

Another obvious solution: make reliable news media more visible. How do you define reliable news media? They should meet journalistic standards of factchecking and citing multiple sources (preferably on the record). Platforms can also work with their stakeholders to identify which news media should be included.

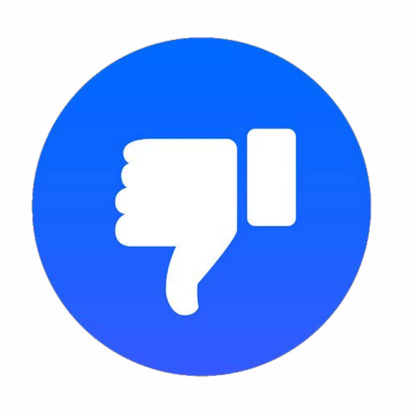
As part of building a healthy ecosystem, it's also important for platforms to disclose their policies and the kinds of problems they are seeing. While the company policy should be widely available for all users to see, the range and nature of problematic content they are experiencing on their platforms, like child porn, should also be disclosed as that's an important step toward tackling them. For instance, after the killings in New Zealand, countries and technology companies came together to form the Christchurch Call to Action[29] to stop the use of the internet for disseminating violent extremist content. Platforms have made other efforts to remove terror related content[30] and have seen some progress there.

Similarly, another important issue that platforms need to tackle and disclose is related to the bias in their algorithms. In the U.S. two Democrat senators introduced the Algorithmic Accountability Act of 2019[31] which would require companies to assess their automatic decision systems for risks to "privacy and security of personal information" and risks of "inaccurate, unfair, biased, or discriminatory decisions." They must also "reasonably address" the results of their assessments, the proposed bill says. In other words, the bill proposes that companies audit their machine-learning powered systems for bias. The move stems from news that Facebook has been serving some discriminatory advertising[32], specifically, ads for the housing market. As the Indian advertising market grows for Facebook (and other players), such audits should be carried out here as well to ensure biases are weeded out of the system here.

Platforms can make their disclosures in their quarterly (or half-yearly) report and offer detailed information on them to a group of vetted researchers who in turn work on designing solutions and the industry should come together to work on broader problems such as child porn, among other issues. This is all the more important keeping in mind that technology, and how it's used and abused, adapts and grows to manipulate existing processes very quickly and hence needs to be constantly watched and updated.

RECOGNIZING AND DEALING WITH BAD ACTORS

Bad actors are those whose main purpose is to spread disinformation, often with the aim of sowing/fanning discontent in society. Their identity is often unknown (they can be state-sponsored as well), and they use a mix of techniques including bot armies to spread their agenda. For instance, Russia has been accused of meddling in the 2016 U.S. elections[33] as well as in the Brexit vote[34]. The idea here, again, is not to focus on the content but to understand if it is, indeed, a bad actor at play.



> **SOME OF THE BIGGER PLATFORMS HAVE STARTED TO USE AI TO HIGHLIGHT THESE BAD ACTORS AND ALERT THEIR USERS TO MESSAGES FROM SUSPICIOUS ACCOUNTS**

Some of the bigger platforms have started to use AI to highlight these bad actors and alert their users to messages from suspicious accounts. For instance, Facebook has on more than one occasion removed[35] hundreds of pages, accounts and groups operated by fake accounts on its platforms. If platforms can coordinate these efforts during a sensitive time, say in the lead up to elections or during a pandemic or in case of clashes between different communities, it will go a long way in curbing the reach of potential bad actors.

REDRESSAL

As part of disclosing its policies, social media platforms should also inform users if their account is being suspended or content being removed, the reason behind the move, including what part of their community standards were violated by that content and offer them a process for redressal of their complaints.

Companies can appoint an ombudsman as well, one who would operate sort of like an independent reviewer year round, and who can be the point of contact for users to voice their complaints. There should be an appeal mechanism where a user can escalate her complaint to a higher authority, perhaps to the ombudsman, one time.

THE ROLE OF GOVERNMENT

REGULATION

Is there a role for a regulator in all of this? Potentially, yes. The main purpose of a regulator should be one of oversight and to check on the effectiveness of the platforms, audit their performance as well as be a route of redressal for users. A regulator can be in addition to an ombudsman, as the latter is an internal appointment of each social media platform (one who it empowers to conduct independent audits and flag issues of concern to the top management), and who can work with industry across the board.

Alternatively, instead of waiting for the government to appoint a regulator for the sector as that can be a long drawn-out process, it might be more practical to focus regulation in the form of an industry organisation, similar to NASSCOM for the IT sector. The industry body can focus on research for the things the platforms flag, it can coordinate among ombudsmen of all platforms and can be the representative for the industry to take up issues with the government. The industry body can also make standards to regulate itself, like the video game industry[36] in the U.S. and Europe did—it developed its own rating system to deal with violence in videogames.

> **REGULATION NEEDS TO BE TIERED—THE BIGGEST FIRMS SHOULD HAVE DIFFERENTIATED REQUIREMENTS IN TERMS OF TRANSPARENCY AND OVERSIGHT IN COMPARISON TO THE SMALLER PLAYERS**

While the entire sector should be regulated, regulation needs to be tiered—the biggest firms should have differentiated requirements in terms of transparency and oversight in comparison to the smaller players as the latter often don't have the means to make the same investments as the bigger firms.

In case the government does decide to appoint a regulator for this sector, it should be a new body, along the lines of the U.K.'s Ofcom perhaps, and not be merged in with the duties of an existing regulator. Reason being social media is a different beast from all other kinds of platforms and contents and its needs are very specific, and urgent, and an existing regulator who already has his hands full may not be able to do justice to the requirements of this situation.

Also, it's important to keep in mind that if trust has to be established between companies and the regulator, but also with the users, it's of utmost importance that the regulator be independent. If social media platforms are to be held accountable and transparent, so should the regulator governing them. Regulators must be accountable for their decisions and both the decisions and the process to arrive at them must be evidence-based.

India can also look at Europe for some ideas on how to tackle these thorny regulatory issues. The E.U. recently introduced the Digital Services Act[37] which would require digital platforms to take responsibility for removing illegal content, from hate speech to counterfeit goods. The DSA also insists on some "safeguards" for users whose content has been erroneously removed. It also calls for more transparency on the platforms' online advertising and on the algorithms used to recommend content to users.

Similarly, Germany introduced in 2017 the Network Enforcement Act or the NetzDG law[38] under which online platforms can be fined up to 50 million euros for systemic failure to delete illegal content. Supporters of the legislation see it as a necessary step to curb online extremism and hatred while its opponents view it as a move toward draconian censorship.

India, too, can consider a penalty for systemic failures by platforms with a few caveats—the platforms should be given sufficient time to take down the illegal content in question and should be penalized only if they are violating the country's laws as noted above or their own TOS. Moreover, penalties should be administered only in case of a systemic failure—that can be defined by the platform in consultation with its stakeholders or can be similar to what it lays out for users (in some cases that's three strikes and you're out). Penalties for reasons beyond that would be a case of overreach and in danger of curbing freedom of expression.

DO EXISTING LAWS NEED TO BE MODIFIED/STRENGTHENED?

Social media platforms must be allowed to perform their jobs. Pressure tactics and threats (whether from government, political parties or powerful companies) to remove pieces of content do not help that process. Rather, it sets a bad precedent as platforms are not equipped to make such decisions and if they give into pressure tactics once, they set the precedence to do so regularly. As we have stated above, platforms should follow the law of the land and their own TOS when it comes to making decisions on what content to remove or keep. They just need to be transparent about the process and ensure there's a process for redressal.

That said, there is room to improve the current laws. The current IT Act came up in the year 2000 and technology has changed vastly since then. The law, too, needs to keep up with the changes. For instance, under the Act, intermediary includes even cyber cafés and it doesn't make sense to equate a neighborhood cybercafé with large platforms like Facebook and YouTube and to put the same burden of investments on them.

> **THERE SHOULD BE LEGISLATIVE CHANGES TO ENSURE TRANSPARENCY IN ADVERTISING AS WELL AS THAT'S THE CORE BUSINESS MODEL OF THESE PLATFORMS**

Similarly, we believe there should be legislative changes to ensure transparency in advertising as well as that's the core business model of these platforms as we discussed above. And there should be some legal definition of disinformation and false news with the focus placed on the intent else it can be easily abused to clamp down on dissenting views.

Some of these moves have started to take place in parts of the world as noted above.

At the end we'd like to say that it's time for India to have a more sophisticated and nuanced approach in dealing with social media platforms as they are deeply enmeshed in our lives and have far reaching impact. At the same time, we believe it should be a combination of self-regulation and co-regulation with the end goal of empowering the user.

ENDNOTES

[1] https://cis-india.org/internet-governance/resources/section-79-information-technology-act

[2] https://www.reuters.com/article/us-facebook-extremists-idUSKCN1SL07Q

[3] https://www-2018.swansea.ac.uk/press-office/news-archive/2017/researchsuggestsbotsgeneratedsocialmediastoriesduringeureferendum.php#accept

[4] https://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles

[5] https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/

[6] https://www.researchgate.net/publication/336553684_Fake_News_in_Social_Media_Bad_Algorithms_or_Biased_Users

[7] https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

[8] https://knightcolumbia.org/content/first-things-first-online-advertising-practices-and-their-effects-on-platform-speech

[9] https://www.theverge.com/2020/5/12/21254184/how-plandemic-went-viral-facebook-youtube

[10] https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

[11] https://indiankanoon.org/doc/1803184/

[12] https://indiankanoon.org/doc/857209/

[13] https://indiankanoon.org/doc/345634/

[14] https://indiankanoon.org/doc/1774593/

[15] https://indiankanoon.org/doc/926966/

[16] https://indconlawphil.wordpress.com/2015/03/26/the-striking-down-of-section-66a-how-indian-free-speech-jurisprudence-found-its-soul-again/

[17] https://indiankanoon.org/doc/110813550/

[18] https://indiankanoon.org/doc/1142233/

[19] https://indiankanoon.org/doc/493243/

[20] https://santaclaraprinciples.org/

[21] https://www.newamerica.org/oti/reports/assessing-youtube-facebook-and-twitters-content-takedown-policies/?utm_medium=email&utm_campaign=OTI%20-%20One-year%20anniversary%20Santa%20Clara%20Principles&utm_content=OTI%20-%20One-year%20anniversary%20Santa%20Clara%20Principles+CID_bb6b629515a02e7fccae9e1364f519ed&utm_source=Campaign%20Monitor%20Newsletters&utm_term=assessment

[22] https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf

[23] https://uk.reuters.com/article/us-myanmar-rohingya-facebook/u-n-investigators-cite-facebook-role-in-myanmar-crisis-idUKKCN1GO2PN

[24] https://restofworld.org/2020/pressing-pause-on-fake-news-in-myanmar/

[25] https://www.wired.com/2016/12/photos-fuel-spread-fake-news/

[26] https://www.wired.co.uk/article/tiktok-india-hate-speech-caste

[27] https://www.reuters.com/article/idUSKBN29J044

[28] https://www.forbes.com/sites/abrambrown/2020/06/30/discord-was-once-the-alt-rights-favorite-chat-app-now-its-gone-mainstream-and-scored-a-new-35-billion-valuation/?sh=59d99d05b6b2

[29] https://www.orfonline.org/research/one-year-since-the-christchurch-call-to-action-a-review/

[30] https://www.cambridge.org/core/journals/international-journal-of-law-in-context/article/regulating-terrorist-content-on-social-media-automation-and-the-rule-of-law/B54E339425753A66FECD1F592B9783A1

[31] https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/06/Algorithmic_Accountability_TWG_MacCarthy_Oct_2019.pdf

[32] https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

[33] https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump

[34] https://www.npr.org/2019/01/19/686830510/senate-finds-russian-bots-bucks-helped-push-brexit-vote-through

[35] https://www.forbes.com/sites/mnunez/2019/12/20/facebook-removes-hundreds-of-fake-pro-trump-accounts-using-ai-generated-profile-photos/?sh=1577f9956175

[36] https://www.hindustantimes.com/analysis/adopt-a-self-regulation-model-for-tech-industries/story-7S6RyZGpfETTRRhjvt8wFO.html

[37] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

[38] https://techcrunch.com/2017/10/02/germanys-social-media-hate-speech-law-is-now-in-effect/

ESYA centre

www.esyacentre.org