ISSN (O): 2584-2196

INTERNATIONAL JOURNAL OF LEGAL AFFAIRS AND EXPLORATION

Volume 2| Issue 2

2024

NAVIGATING THE INFORMATION LANDSCAPE: INDIA'S REGULATORY RESPONSE TO DISINFORMATION ON SOCIAL MEDIA PLATFORMS

Aarathi Namboodiri and R Ranga Priya

Final Year BA LL.B Students of SASTRA Deemed University

ABSTRACT

The proliferation of online disinformation and misinformation on social media platforms, such as Twitter and Facebook, has emerged as a formidable challenge to India's information ecosystem. This article seeks to investigate the regulatory strategies employed in India to counter the dissemination of false information on these platforms, with a specific focus on the impact of the removal of Section 66A from the Information Technology Act of 2000. In light of growing concerns surrounding the influence of media and the government's attempts to address this issue, the need for effective policies and legal frameworks becomes paramount. Through a comparative analysis, this study assesses the efficacy of existing policies, legal structures, and enforcement mechanisms in mitigating the adverse effects of disinformation and misinformation on Twitter and Facebook. The research aims to shed light on the challenges faced by Indian authorities in curbing the dissemination of false information and the broader implications for media influence and government intervention. By scrutinizing the outcomes of regulatory measures and their effectiveness, this study ultimately strives to contribute to the ongoing discourse on safeguarding the information ecosystem in India, where the government's past efforts have faced notable challenges in addressing the ever-evolving issue of online disinformation and misinformation.

KEY WORDS: *Misinformation, Disinformation, Social media platforms, Fake news, Facebook, Twitter.*

---

1. INTRODUCTION

In recent years, social media platforms have had a significant impact on communication systems. These platforms offer a platform for users to communicate and share information. During these communications, information that is not true or false is being disseminated rapidly. Fake news is rapidly spread through various social media platforms and reaches millions of users, who use them without checking for its accuracy. Misinformation and disinformation is an emerging area of

concern with respect to social media. Misinformation and disinformation are related concepts, both involving the spread of false or inaccurate information. However, they differ in the intent behind the dissemination of information.

Misinformation refers to the spread or sharing of false or inaccurate information without a deliberate intent to deceive. Misinformation can be unintentional and may result from a lack of knowledge, misunderstanding, or a genuine mistake. It can be spread through social media, traditional media, word of mouth, or other communication channels. For instance, in the context of social media, misinformation can include sharing false news articles, inaccurate statistics, or outdated information without the knowledge that it is false.

Disinformation, on the other hand, involves the intentional creation and dissemination of false information with the purpose of misleading and deceiving the audience. Disinformation is crafted and spread with the specific intent to manipulate public opinion, achieve a particular goal, or harm individuals or groups. It is often designed to look credible, making it more challenging for the audience to discern its falseness. Disinformation on social media may involve the deliberate creation of fake news stories, the spread of manipulated images or videos, or the use of bots and trolls to amplify false narratives.

The paper examines the main problems with India's current legal framework for policing hate speech and fake news online and makes recommendations based on what other democracies have done. It makes the case that a co-regulatory model must be adopted in order to change the Indian strategy. The ability of the platform to limit online harms can be strengthened by governmental support in the form of an ethics code and criminal penalties. To improve the existing content moderation procedures, strictly independent duties are advised.

2. RESEARCH SIGNIFICANCE

This study assumes great significance as it investigates the effectiveness of current policies, legal frameworks, and enforcement mechanisms in combating the detrimental impacts of disinformation on Twitter and Facebook. Given the increasing influence of these platforms on public discourse and the government's ongoing challenges in addressing misinformation, this research provides essential insights that can inform more effective regulatory strategies, thereby safeguarding the information ecosystem in an age characterized by the rampant spread of false information.

This study assumes great significance as it investigates a specific case on the impact of the removal of Section 66A from the Information Technology Act of 2000 and this leading to a butterfly effectiveness of current policies, legal frameworks, and enforcement mechanisms in combating the detrimental impacts of disinformation on Twitter and Facebook. Given the increasing influence of these platforms on public discourse and the government's ongoing challenges in addressing misinformation, this research provides essential insights that can inform more effective regulatory strategies, thereby safeguarding the information ecosystem in an age characterized by the rampant spread of false information.

3. LITERATURE REVIEW AND METHODOLOGY

The research draws on a comprehensive review of literature covering misinformation, social media, fake news, and regulatory frameworks. Methodologically, the study employs a comparative analysis to assess the efficacy of India's current regulatory measures.

The phrase "fake news" has been around since the US election of 2016.[1] But ever since then, its popularity has grown. Scholars have used the term "fake news" in a variety of ways. While some definitions reflect misinformation and disinformation, others reflect rumors. Unverified information that arises in unclear and dangerous situations, when information is scarce, and when people feel the need to comprehend or feel safe is referred to as rumor.[2]

Information intended or inadvertently created to mislead people is referred to as Misinformation (unintentionally false or misleading information). Information that is purposefully designed to deceive people is called Disinformation. “Fake news” can be both true and false.[3] According to data, more than 40% of the news shared on social media platforms was found to be fake. This fake news brings large societal risks of chaos and havoc.[4] In March 2023, an Indian survey on digital news consumption found that over 60% of participants said they occasionally came across potentially false information on the internet.[5]

[1] Quandt, T., Frischlich, L., Boberg, S., & Schatto-Eckrodt, T. (2019). Fake news. In The International Encyclopedia of Journalism Studies (pp. 1–6). American Cancer Society. https://doi.org/10.1002/9781118841570.iejs0128

[2] Meel, P., & Vishwakarma, D. K. (2019). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications, 112986. https://doi.org/10.1016/j.eswa.2019.112986

[3] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[4] Social Media Users — Global Demographics (2023)

[5] https://www.statista.com/statistics/1406289/india-frequency-of-seeing-fake-news-online/#:~:text=According%20to%20a%20digital%20news,encountered%20potentially%20fake%20news%20online.

4. MISINFORMATION AND DISINFORMATION

In the context of social media, both misinformation and disinformation can spread rapidly due to the nature of these platforms. The viral nature of content sharing, the lack of gatekeepers, and the algorithms that prioritize sensational or provocative content contribute to the challenge. Additionally, the anonymity and ease of creating and sharing content on social media make it a fertile ground for the intentional spread of false information. Social media platforms also play a crucial role in developing and enforcing policies to curb the spread of misinformation and disinformation on their platforms. Some scenarios include:

India has been plagued by fake news concerning cyclones, public health, and child abuse. When intertwined with religious or caste issues, the combination can be explosive and lead to violence. People have been killed when false rumors have spread through digital media about child abductions.[6]

Another example, a documentary titled "Plandemic" featured the views of scientist, Dr. Judy Mikovits, who argued that the SARS-CoV-2 virus was "manufactured" and part of a scheme by pharmaceutical corporations to profit from selling vaccine. The documentary managed to rack up about 8 million views in 8 days.[7]

On October 10, a social media post claiming Nobel laureate Amartya Sen had passed away went viral. The news was first shared on X (formerly Twitter), a handle named after Claudia Goldin, the 2023 winner of the Nobel Prize in Economics. "A terrible news. My dearest Professor Amartya Sen has died minutes ago. No words" the tweet stated. The news was then carried by a number of news organizations, including ETV Bharat, Press Trust of India, Times of India, Deccan Herald, Zee News, Free Press Journal, Deccan Chronicle, Firstpost, and Moneycontrol Hindi. These social media claims were shown to be untrue by Alt News[8].

Addressing misinformation and disinformation on social media requires a combination of user education, fact-checking initiatives, platform policies, and technological solutions. Efforts to promote media literacy, critical thinking, and responsible sharing of information are essential to mitigate the impact of false information in the digital age. The misuse of these platforms can

[6] Vidhi Doshi, "India's Millions of New Internet Users are Falling for Fake News – Sometimes with Deadly Consequences," Washington Post, October 1, 2017.
[7] Daniel Funke, Fact-checking 'Plandemic': A documentary full of false conspiracy theories about the coronavirus, *PolitiFact.com,* May 7, 2020.
[8] 'News' of Amartya Sen's passing false, confirms Nobel Laureate's family, alt News, October 11, 2023.

have economic, psychological, and political impacts, both online and offline[9] and can lead to discrimination and even violence[10].

5. EXISITNG POLICIES AND ACTS IN INDIA

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021[11]:

Social media platforms with more than 5 million users are required to enable traceability of the originator of information. This provision aims to hold individuals accountable for the spread of misinformation and disinformation. Intermediaries, including social media platforms, are required to establish a grievance redressal mechanism to provide users with a way to report and resolve complaints regarding the dissemination of misinformation on these platforms. The rules empower the government to block content that is in the interest of sovereignty and integrity of India, defense of India, security of the state, and public order to curb the spread of misinformation that may pose a threat to national security or public order. The government has the power to block public access to information during emergencies to prevent the dissemination of information that could potentially harm public safety or contribute to the spread of misinformation during emergencies.

- ***Information Technology (IT) Act, 2000[12]:***

It is the primary legislation in India that addresses various aspects of cyber law. It plays a pivotal role in regulating online content. It provides a legal framework to combat various cybercrimes, including disinformation and misinformation. Under this Act, social media platforms like Twitter and Facebook are expected to adhere to certain guidelines and take necessary steps to curb the spread of false information.

- ***Self-Regulatory Measures:***

Twitter and Facebook have implemented various self-regulatory measures to curb misinformation and disinformation on their platforms. Please note that the strategies and effectiveness of these measures may have evolved since then, and it's recommended to check the latest information for

[9] Annie Gowen and Manas Sharma, "Rising Hate in India", *The Washington Post*, October 31, 2020.
[10] UN News, Hate speech exacerbating societal, racial tensions with 'deadly consequences around the world', say UN experts, September 23, 2019.
[11] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021
[12] The Information Technology Act, 2000

the most up-to-date details. Here's a comparison of the self-regulatory measures adopted by

Twitter's Approach

- ***Warning Labels:*** Labels tweets that contain misleading or disputed information, providing users with additional context or directing them to reliable sources.
- ***Reducing Visibility***: Adjusts algorithms to reduce the visibility of tweets that violate its policies, making it less likely for users to encounter such content.
- ***Birdwatch:*** Introduced a pilot program called Birdwatch, allowing users to identify and provide context to potentially misleading information. It aims to leverage the collective wisdom of the community to address misinformation.
- ***Algorithmic Changes:*** Continuously updates its algorithms to identify and limit the spread of misinformation. The platform also explores ways to address echo chambers and filter bubbles.
- ***Deactivation of Fake Accounts***: Takes measures to deactivate fake accounts and bots that are often associated with the amplification of false information.
- ***Public Interest Notices***: May apply Public Interest Notices to certain tweets that violate its rules but are deemed to be in the public interest to remain accessible. This helps balance the need for information with the prevention of harm.

Facebook's Approach

- ***Fact-Checking Partnerships***: Collaborates with third-party fact-checkers to assess the accuracy of content. If content is identified as false, its distribution is reduced, and warning labels are added.
- ***Reduced Reach for False Content***: Adjusts its algorithm to reduce the reach of content identified as false by fact-checkers. This helps limit the spread of misinformation.
- ***Promoting Authoritative Sources***: Prioritizes content from authoritative sources in users' news feeds, aiming to elevate reliable information and reduce the visibility of false content.
- ***Information Labels***: Adds labels to certain posts that provide context or direct users to reliable sources. This is particularly common for content related to COVID-19 and elections.
- ***Content Removal***: Removes content that violates its policies, including misinformation that may incite violence or pose immediate harm.
- ***Election Integrity Efforts***: Implements specific measures during elections, such as restricting political ads and providing information hubs to educate users about the electoral process.

- ***User Reporting***: Encourages users to report false content, which helps the platform identify and take action against misinformation.

Both Twitter and Facebook continuously refine their approaches and introduce new measures to combat misinformation and disinformation. The effectiveness of these measures is a subject of ongoing evaluation and discussion within the broader context of online information ecosystems. Users also play a crucial role in these efforts by being discerning consumers of information and reporting content that appears to be false or misleading.

6. EFFECTIVENESS OF ENFORCEMENT MECHANISMS

Here are some common points raised by critics and challenges associated with curbing online misinformation and disinformation on Twitter and Facebook:

- ***Scale and Speed:*** The sheer volume and rapid dissemination of information on these platforms make it challenging to identify and address misinformation in real-time. Automated systems and algorithms may struggle to keep pace with the speed at which false information can spread.
- ***Algorithmic Bias:*** Critics argue that algorithms designed to identify and mitigate misinformation may have biases, potentially leading to the unequal treatment of content or the inadvertent suppression of legitimate information.
- ***Human Moderation Challenges:*** Human moderators play a crucial role in content moderation, but the scale of user-generated content on these platforms makes it difficult to review every piece of information effectively. Moderators may face challenges in distinguishing between misinformation, satire, and genuine expression of opinions.
- ***False Positives and Negatives:*** Automated systems and human moderators may make errors in identifying and handling misinformation. False positives (removing accurate information) and false negatives (allowing misinformation to stay) are ongoing concerns.
- ***Evolution of Tactics***: Those spreading misinformation often adapt their tactics to circumvent detection mechanisms, making it a constant challenge for platforms to stay ahead.
- ***Lack of Accountability:*** Critics argue that the lack of transparency in content moderation decisions and the perceived inconsistency in enforcing policies undermine the platforms' accountability.

And most importantly, following the unconstitutionality of Section 66A of the IT Act[13], no provision under the IT Act currently aims to curtail either online or offline hate speech or to control social media platforms.

Section 66A talks about sending offensive messages through communication service, etc. This section criminalized sending offensive messages through communication services. The main objective is to penalize individuals for transmitting information that is false and could cause annoyance, inconvenience, danger, obstruction, insult, injury, criminal intimidation, enmity, hatred, or ill will. Section 66A of the Information & Technology Act, 2000 was overturned by the Supreme Court in Shreya Singhal and Ors. v. Union of India[14], in what is being called one of the most significant rulings supporting the right to free speech in recent memory. Both the general public and legal experts applaud the decision, which determined that the Cyber law section was illegal since it restricted the freedom of speech of Indian residents and was open-ended and ambiguous. But still a question exists that is there any sections to replace Section 66A and what all are the negative impacts of the removal of the said section?

Shreya Singhal v. Union of India (2015)[15]

It was a landmark case in India that dealt with the constitutional validity of Section 66A of the Information Technology Act, 2000. Section 66A criminalized the sending of offensive messages through communication services and was widely criticized for its potential misuse and violation of free speech rights. The case resulted in the Supreme Court of India striking down Section 66A as unconstitutional. Here are the key facts and judgments of the case:

Facts: Arrests under Section 66A: The case originated from instances where individuals were arrested for posting allegedly offensive content on social media platforms under Section 66A.

Petition by Shreya Singhal: Shreya Singhal, a law student, filed a petition challenging the constitutional validity of Section 66A, arguing that it violated the right to freedom of speech and expression guaranteed by Article 19(1) (a) of the Indian Constitution.

Judgments: Supreme Court's Decision (March 24, 2015): The Supreme Court, in a landmark judgment, declared Section 66A unconstitutional and struck it down.

The court held that Section 66A was overly broad, vague, and had a chilling effect on free speech. It violated the constitutional principles of free speech and was not saved by any of the reasonable

[13] *supra 12*
[14] Shreya Singhal v Union of India
[15] *supra 14*

restrictions provided in Article 19(2) of the Constitution.

Key Observations: The court emphasized the importance of protecting the freedom of speech, even in the online space. It noted that the language used in Section 66A was vague and lacked clarity, leading to its arbitrary and discriminatory enforcement.

Problems Regarding the Removal of Section 66A: The repeal of S. 66A does not, however, grant unrestricted freedom of speech because similar provisions of the Indian Penal Code (IPC) still apply to social media. These provisions include S. 295A, which for intentionally insulting religion or religious beliefs, S. 153A, which for encouraging animosity between groups on the basis of religion, race, or other factors, S. 499, S. 505 for statements conducive to public mischief, S. 509 for insulting a woman's modesty, S. 509 for criminal intimidation, S. 499 for defamation, and S. 124-A for sedition. According to S.66A, the police had to determine for themselves the seriousness of the content uploaded before they could make an arrest. The most significant drawback of its repeal will be for those who were really the receivers of those emails or communications, since they would no longer be able to get timely counteractive remedy due to the more onerous process involved in handling similar offenses under the IPC. It will now be the difficult task for complainants to approach the Court first, without whose approval no action can be taken. That is, some critics argued that the removal of Section 66A left a gap in addressing online harassment and abuse, as the provision was initially intended to prevent the sending of offensive messages. The removal of Section 66A highlighted the need for updated legislation that could address online offenses without compromising free speech rights. There have been calls for a comprehensive and well-defined legal framework to handle cybercrimes effectively. Law enforcement agencies faced challenges in dealing with certain types of online offenses due to the absence of a specific provision like Section 66A. The absence of a clear legal framework might lead to difficulties in prosecuting certain cybercrimes.

"In Section 66A, there was immediate relief to a lot of people because it was struck down. Also, the ruling with respect to intermediary rules has been immensely beneficial for free speech. However, there are points of concern given that the Supreme Court itself is applying the ruling inconsistently and not expanding or following the course set by it," said lawyer Apar Gupta, who also represented the non-profit People's Union for Civil Liberties, a petitioner in the batch of cases with the Shreya Singhal case. As an example, Gupta referred to a recent interim order[16] by the

[16] PTI, Supreme Court asks three search engines to set up body to delete info on sex determination, *The Indian Express,* February 16, 2017.

apex court directing intermediaries to have an in-house mechanism for screening of search results. “It seems to conflict with the Shreya Singhal case," he said. The manner in which courts have handled issues of free speech and access to information does not augur well. Gupta said, “At the very least it shows inconsistency by the court and otherwise it shows that Shreya Singhal may be a case in isolation. It undermines the significance of the case."[17]

Policymakers' growing pessimism about internet companies' ability to self-regulate has sparked worries about the unexpected but serious ramifications of the way a platform is now understood. A government order may require an intermediary to remove objectionable content under Section 79 of the IT Act without subjecting the intermediary to publishing liability. Other than complying with state-determined policy, platforms also moderate harmful and unethical use through their own policies[18]. Many argue that social media’s role has evolved, as these platforms perform more than an intermediary role as content host[19].

7. SUGGESTIONS TO CURB THE MISINFORMATION AND DISINFORMATION

Investments in Fake News Detection Technology: Investing in advanced technologies and tools dedicated to identifying and combatting fake news. This may involve developing machine learning algorithms, natural language processing (NLP), and other data analytics techniques.

Identification of Fake News Through Stringent Algorithms: Implementing sophisticated algorithms that analyze content for signs of misinformation. These algorithms can be designed to identify patterns, inconsistencies, or other characteristics associated with false information. Using Crowdsourcing as a Means to Prevent and Control the Spread of Misinformation and Disinformation: Involving the collective intelligence of a crowd to assess and verify information. Crowdsourcing can be employed for fact-checking, reviewing content, and flagging potentially false information. Platforms like Facebook and Twitter have utilized community reporting systems.

High Degree Accuracy Tests to Detect the Accuracy of Information Before It Is Shared on Social Media: Implementing stringent accuracy tests to evaluate the credibility of information before it is allowed to be disseminated on social media platforms. This could involve partnerships with fact-checking organizations or the use of advanced verification processes.

[17] Section 66A of IT Act; What has changed two years on? Shreeja Sen, https://www.livemint.com/Politics/qjqRU6mS4NZWUvNZZbT5eP/Section-66A-of-IT-Act-2000-What-has-changed-two-years-on.html, *Livemint*, March, 2017.
[18] Megha Mandavia, Facebook says will remove content to mitigate adverse legal or regulatory impact, *The Economic Times*, September 2, 2020.
[19] Shuchi Bansal, Content regulation lapses casts doubts on Facebook's biz model, *Livemint*, August 24, 2020.

Automatic Tools to Identify Sites Giving Out Fake News: Developing tools that automatically identify and flag websites or sources known for spreading fake news. This could involve blacklisting certain domains or using web crawling algorithms to assess the credibility of information sources.

Use of Software at the Time of Generation of Message to Detect Its Accuracy: Implementing real-time accuracy checks during the creation or generation of messages. This could involve integrating software solutions that analyze the content before it is posted, aiming to prevent the dissemination of false information from the outset.

Rather than deleting a protective clause in its entirety, the judiciary should have attempted to provide appropriate interpretation and context for the unclear elements included in it, placing them under the jurisdiction of Article 19 (2). The Bench may have thought about defining precise criteria for what would be considered an offense under S.66A and setting thresholds that would have reduced the likelihood that the provision would be applied arbitrarily while still permitting the Constitution's reasonable limitations to be implemented. As was pointed out by the Additional Solicitor General in this case while justifying the necessity to retain such a provision "If the medicine is bitter then we can have sugar after it instead of throwing the medicine. People have to take the medicine as it is for their benefit"[20]. The Bench should have kept in mind the purpose and intent of S.66A as well as the ongoing need to keep an eye on the detrimental effects of unfettered free speech while simultaneously making sure that baseless and ludicrous accusations from vested interests are kept to a minimum. The Court ought to have used this case as an opportunity to address the question of which categories and types of freedom of expression should be reasonably restricted, given the speed and scope with which information goes viral in the modern world.

India may gain insight about social media platform regulation from other nations. Various nations, including Germany, the US, the UK, Singapore, and Russia, have implemented particular accountability guidelines that are linked to the online platforms. In order to prevent arbitrary intervention, these nations have established clear limits on the reach of government authority and have even designated impartial regulators or judges to carry out moderation duties.

Encourage digital education programs and workshops to help young people develop the skills they need.

In conclusion, regulating online disinformation is a complex task that requires continuous efforts from both social media platforms and governments. While Twitter and Facebook have made

[20] Amit Choudhary, Sec 66A draconian, but is needed: Govt, *Times of India*, February 05, 2015.

strides in tackling disinformation, there is still room for improvement. Striking a balance between freedom of expression and preventing the harmful effects of disinformation remains a challenge. It is crucial for policymakers, platforms, and users to work together to create a safer and more informed online environment.

8. CONCLUSION

The spread of harmful and hateful misinformation and disinformation content on the internet has gotten so bad and this will to the scapegoating and "disproportionate effects" that harm marginalized communities, including their denial of access to healthcare, increased victimization, and harsh security measures. The growing influence of social media platforms has not been accompanied by the development of appropriate limitations and regulations. Without any prior warning, awareness, or ability to resist, users are exposed to harmful content. The main draws for a social media platform are still its enormous user base, enhanced connectivity and sharing, and revenue. Every law is susceptible to abuse. Because practically every other legal provision would then be susceptible to removal using the same logic, the mere possibility of abuse should not have been the justification for eliminating an essential provision in its entirety. Given that the ability to express oneself freely while upholding one's reputation and dignity is a fundamental human right, this case may have provided the Court with a suitable opportunity to outline a reasonable line that separates the protection of rights from their restriction. It is hoped that the analysis and suggestions provided in this paper will aid in the investigation of potential regulatory measures for India to combat false information on social media.