



# Content Regulation of Social Media: A Phenomenon of Violation of Freedom of Speech

**Name of the Author: Adv. Riddhi Chaudhary Bajaj**

**Name of the College: Kirit P. Mehta School of Law, NMIMS**

## **Abstract**

Article 19 of the Universal Declaration of Human Rights protects the freedom of speech, by its proclamation in the year 1948 at United Nations General Assembly wherein it recognised freedom of expression as a fundamental right. This right has been echoing in world-wide democratic values. Today's digitalisation is in fact an extension to these rights, wherein, it is witnessed that social media platforms are playing a crucial role in expressing self and putting forth one's opinions.

As the digital era shifts the world societies to a single global community there is an increase of free speech in the virtual space. Perhaps this increase in free speech has in turn resulted in an increase in content regulation practices which has led to a debate over the balance between moderation and censorship.

This paper critically examines global instances where content control measures on social media have led to the violation of freedom of speech – such as content takedowns and arbitrary suspensions, with key attention to corporate policies, law of the land and any governmental pressures that influence these regulatory practices, often undermining the UDHR's principles.

Further, this paper delves into the recent content moderation – community-driven systems and community notes, which aims to protect freedom of speech and user accountability. The analysis assesses how these collaborative models may either reinforce or undermine the foundational right to freedom of speech, while balancing the need to curb misinformation and harmful content.

This paper aims to highlight the potential conflicts and resolutions in maintaining the integrity of free expression in the digital age by studying restrictive content policies with the principles enshrined in Art 19. Furthermore, this paper is a call for harmonizing digital freedom with accountability, ensuring the digital sphere remains a space for diverse voices without compromising human rights.

## **Keywords**

Freedom of Speech, Freedom of Expression, Social Media, UDHR, Twitter, Meta and X

## **Introduction**

Freedom of Speech stands as a cornerstone of democratic societies, a principle enshrined in Article 19 of the Universal Declaration of Human Rights since its proclamation in 1948 by the UN General Assembly. This right has found new avenues for expression in this digital era, specifically through social media platforms which have become pivotal in enabling individuals to voice their opinions and share diverse perspectives.

This paper delves into the critical issues surrounding content control measures on social media, which have at times led to violations of freedom of speech. It examines global instances where practices such as arrest of an individual, content takedowns and arbitrary suspensions are influenced by corporate policies, national laws and governmental pressures, often contrary to the principles enshrined in the UDHR.

The study further explores innovative approach – such as community-driven content moderation systems and collaborative mechanism like community notes, which aim to uphold user accountability while protecting free speech. By analysis these models, the paper studies its potential to reinforce the foundational right to freedom of expression. Further, it emphasizes the need for a balanced approach that preserves the integrity of free expression in the digital age, ensuring that the digital sphere remains a space for diverse and inclusive dialogue without compromising human rights.

## **Global Recognition of Freedom of Speech**

In the evolution of Human Rights, The Universal Declaration of Human Rights (herein after to be referred as UDHR) is a historic achievement. During the General Assembly resolution 217 A held on 10<sup>th</sup> December, 1948 in Paris, UDHR, which was crafted by representatives of diverse cultural and legal backgrounds, ensuring inclusivity across all regions, was adopted by the United Nations General Assembly as these countries agreed on a comprehensive list of inalienable human rights. For the first time a global benchmark for justice, equality and dignity was upheld with the help of a comprehensive framework of fundamental rights established by UDHR which are to be universally upheld and influence the development of international human rights law.

This cornerstone of modern human rights law has been translated into over 500 languages, emphasizing its universal reach.<sup>1</sup> It has been witnessed that UDHR has in fact influenced the creation of more than seventy human rights treaties world over.

International and regional human rights standards are a result of these legally binding agreements which continue to shape them and ensures the protection of individual freedoms and promoting accountability among nations. The UDHR's enduring impact reaffirms the collective commitment to safeguard human dignity and justice for all, making it a foundational guide for global human rights advocacy and legislation.

Perhaps one of the most important rights which is the core of UDHR is Article 19 from which all of us derive our right to freedom of speech and expression.

<sup>1</sup> United Nations, Global Issues – Human Rights, last visited on 16<sup>th</sup> Jan, 2025 - <https://www.un.org/en/global-issues/human-rights>

*“Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.”<sup>2</sup>*

As we study further in this paper, we see that freedom of speech was never absolute from its inception.

### **Old School and New School of Freedom of Speech**

It is noteworthy to note how from the twentieth century to twenty-first century the regulation of speech has evolved significantly. Historically, a dualist model was followed in the twentieth century, where speech regulation involved two primary actors –

- a. Territorial governments and
- b. Individual speakers.

However, at present it has shifted from dualist model to a pluralist model involving multiple key players, wherein it can be said that one key player being the nation-states and supranational entities like the European Union; the second key player comprising of privately owned internet infrastructure companies such as social media platforms, search engines, broadband providers and electronic payment systems and the third key player includes a wide range of speakers, including legacy media, civil society organizations, hackers and trolls.

Traditional (old school) and Modern (new school) forms of speech regulation, along with private policies, all influence speaker’s ability to communicate freely. It is pertinent to note that both nation-states and civil society organizations exert pressure on companies to moderate content, thus complicating the free speech ecosystem.

The pluralist model give rise to three significant challenges –

1. Social Media companies are increasingly pressured to regulate free speech through modern regulatory methods.
2. Social Media companies have created internal policies and enforcement mechanisms that often operate arbitrarily.
3. The end users face increased vulnerability to digital surveillance and manipulation such as targeted campaigns, privacy breaches and other potential exploitation.

Further, freedom of speech has heavily become dependent on third party player – Social Media Platforms in the twenty-first century. These digital platforms through which people communicate today, is controlled and managed by these Social Media companies who shape the digital public space. Therefore, these private entities play a crucial role in governing online speech, raising complex issues around regulation and control.

<sup>2</sup> Article 19 of Universal Declaration of Human Rights last visited on 16<sup>th</sup> Jan, 2025 - <https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2019,media%20and%20regardless%20of%20frontiers.>

A very recent example of controversies of speech regulation rather violation is the European Union's "Right to be Forgotten" by which it mandates the controller (social media and digital Platforms) to erase the personal data concerning data subject without undue delay and the controller will be in the obligation to do the same.<sup>3</sup>

Germany's NetzDG law is also a violation of free speech as it requires social media platforms to remove various forms of speech.<sup>4</sup>

The above cases show the shift from the dualist or dyadic speech regulation system of the twentieth century to a more complex pluralist model. The dualist approach primarily involved two key players as stated above. Governments regulated speech by overseeing various forms of mass media, including publishing houses, newspapers, radio stations, television networks and film studios. This traditional model focused on state-driven enforcement of speech norms and censorship laws.

While the modern pluralist model involves three primary players – nation-states, private social media/digital companies and speakers. Even though the nation-states continue to enforce speech laws but they now do so indirectly by pressuring private companies to moderate content.

Thus it can be concluded that there is an increasing reliance on private companies (social media/digital companies) for the regulation of speech, thereby raising many critical concerns such as – the risk of "collateral censorship"<sup>5</sup>, where platforms remove more content than necessary out of fear of penalties; rise in corporate speech governance and the influence of social media and digital platforms in shaping the discourse of public, as these companies gather vast amounts of personal data to tailor content and advertisements.

This complex relation and interaction amongst nation-states, private corporations and speakers reveal the challenges of maintaining free expression guaranteed under Art. 19 of the UDHR in digital landscape dominated by both state control and corporate interests. Balancing the protection of speech with security, privacy and accountability remains a central challenge in modern speech regulation.

### **Freedom of Speech on Social Media a Myth**

As discussed above that Freedom of Speech is in fact a fundamental right of everyone, however, certain reasonable restrictions are implemented on the same in various countries and these restrictions varies nation to nation.

<sup>3</sup> Article 17 Right to erasure of General Data Protection Regulation, last visited on 16<sup>th</sup> Jan, 2025 - <https://gdpr.eu/article-17-right-to-be-forgotten/>

<sup>4</sup> Diana Lee, *Germany's NetzDG and the Threat to Online Free Speech*, Media Freedom and Information Access Clinic, Yale Law School, dated 10<sup>th</sup> October, 2017, last visited 16<sup>th</sup> Jan, 2025 - <https://law.yale.edu/mfia/case-disclosed/germanys-netzdg-and-threat-online-free-speech>

<sup>5</sup> Jack M. Balkin, *SYMPOSIUM: A First Amendment for All? Free Expression in an Age of Inequality*, Columbia Law Review, NOVEMBER 2018, Vol. 118, No. 7, pp. 2011-2056, last visited on 16<sup>th</sup> Jan, 2025 - <https://www.jstor.org/stable/10.2307/26524953>

In India, Article 19 (2) of the Indian Constitution<sup>6</sup> states the reasonable restrictions on Freedom of Speech yet this restriction was extended by Section 66 A of the Information Technology Act<sup>7</sup> which was observed in the case of *Shreya Singhal v. Union of India*<sup>8</sup>, wherein, two women were arrested for allegedly offensive and objectionable comments on Facebook in relation to the shutting down of Mumbai after the death of a political leader. It is pertinent to note that these women had in fact only practiced their fundamental right by expressing their objection to the closure of the city of Mumbai due to the death of a politician leader which in no way was objectionable or in violation of someone else rights nor did it fall under the restrictions stated under Art. 19 (2). The Supreme Court of India held that the Section 66 A of the Information Technology Act, 2000 in fact is unconstitutionally vague and its scope were beyond the permissible restrictions under Article 19 (2) of the Indian Constitution which is why the Supreme Court struck down this Section.

Based on the above decision of the Apex Court in India, it can be said that the Judiciary attempts to maintain the balance between Freedom of Speech and reasonable restrictions by not giving unreasonable powers to the state whereby a very large amount of protected and innocent speech could be curtailed.

Further, even in the recent case in Nigeria, we can witness the modern pluralist model with three key players – the state, social media companies and the speakers. The Social Media Platform Twitter decided to delete a tweet by President Muhammadu Buhari which led to the suspension of Twitter in Nigeria<sup>9</sup> – this exemplifies the complex interplay between the state, freedom of expression and the rights of private entities in the digital age.

This act of twitter also reflects how even the government in power at that time can in fact be restricted of their respective freedom of speech under the UDHR by deleting their expression in the digital world by stating that it was violating its policies. There by questioning the states sovereignty and transnational nature of digital platforms.

Governments – the states, as sovereign entities, possess the authority to regulate activities within their jurisdictions, including those occurring online. However, this authority is not absolute and must conform to constitutional guarantees and international human rights obligations. A pertinent question about the proportionality of the Nigerian governments action and whether it adhered to the principles of necessity and legality were asked.

Another aspect to be taken in to account was that the ban implemented by the Nigerian Government impacted the freedom of speech of its citizens who used the platform as a mode of expression. This in turn resulted in its citizens quickly acquiring VPNs to continue using the said platform, which enabled them to bypass the ban, while some changed their business model, making Facebook and WhatsApp hubs of their economic

<sup>6</sup> Article 19 (2) Reasonable restrictions of the Indian Constitution, last visited on 21<sup>st</sup> January 2025 - <https://indiankanoon.org/doc/1218090/>

<sup>7</sup> Section 66A Punishment for sending offensive messages through communication service, etc of the Information Technology Act, 2000, last visited on 21<sup>st</sup> January 2025 - [https://www.indiacode.nic.in/show-data?actid=AC\\_CEN\\_45\\_76\\_00001\\_200021\\_1517807324077&orderno=77](https://www.indiacode.nic.in/show-data?actid=AC_CEN_45_76_00001_200021_1517807324077&orderno=77)

<sup>8</sup> *Shreya Singhal Vs. Union of India*, AIR 2015 SUPREME COURT 1523 - <https://indiankanoon.org/doc/110813550/>

<sup>9</sup> *Nigeria says it suspends Twitter days after presidents post removed*, Reuters, dated 5<sup>th</sup> June 2025, last visited on 21<sup>st</sup> January 2025 - <https://www.reuters.com/technology/nigeria-indefinitely-suspends-twitter-operations-information-minister-2021-06-04/>



transactions, even as Minister of Information and Culture Lai Mohammed threatened to clamp down on those platforms as well.<sup>10</sup>

This incident reflects the governments struggle on regulating its own freedom of speech on social media platforms and digital platforms; while to show its control on these digital platforms by suspending them, violates the freedom of speech of its own citizen rather than drafting legislations in order to regulate these social media giants.

Another, instance where Donald Trump was banned by Meta. Donald Trump's relationship with social media platforms reveals significant shifts in platform governance, corporate policies and the intersection of politics with digital freedom of expression.

In the year 2021, following the attack on the U.S. Capitol<sup>11</sup>, Twitter permanently suspended Donald Trump's account, citing risks of further incitement to violence. Following this, Meta too implemented indefinite suspensions<sup>12</sup>. These acts of the social media giants were justified under their respective community standards – aiming to prevent harm and misinformation.

In 2022, Elon Musk acquired Twitter, a pivotal moment in platform governance as he advocated for maximalist free speech, promised a more open approach to content moderation and reinstated Trump's account<sup>13</sup> – signalling a shift in Twitter's policy to prioritize freedom of expression over stringent content controls.

Meta lifted restrictions on Donald Trump's accounts on Facebook and Instagram in July 2024<sup>14</sup> after Trump's political comeback and subsequent re-election as U.S. President in 2025. Further, Meta also announced plans to cease third-party fact-checking and loosen restrictions on certain content, citing the need to balance safety with freedom of expression. This shows Meta's evolving content moderation policies.

These developments underscore the dynamic and often contentious relationship between political figures, social media platforms and speakers, in the modern pluralist model. The initial bans on Nigerian President and Trump illustrated the capacity of platforms to enforce standards against powerful individuals. In contrast, the subsequent policy shifts by Twitter and Meta highlight the growing emphasis on freedom of expression, even as platforms face scrutiny for enabling the spread of harmful content. Further, the Governments power to suppress its citizens voice as seen in the first instance by arresting its citizens as they express their opinions on digital platforms, followed by the Government suspending a social media platform when its own freedom of speech is violated; thereby further violating the freedom of expression of its millions of citizens on that

<sup>10</sup> Ebenezer Obadare, *Twitter Ban Shows Limits of State Power in Nigeria*, Council on Foreign Relations dated 18<sup>th</sup> Jan, 2022, last visited on 21<sup>st</sup> Jan, 2025 - <https://www.cfr.org/blog/twitter-ban-shows-limits-state-power-nigeria>

<sup>11</sup> Bobby Allyn, Tamara Keith, *Twitter Permanently Suspends Trump, Citing 'Risk of Further Incitement of Violence'*, NPR dated 8<sup>th</sup> January 2021, last visited on 21<sup>st</sup> January 2025 - <https://www.npr.org/2021/01/08/954760928/twitter-bans-president-trump-citing-risk-of-further-incitement-of-violence>

<sup>12</sup> *Facebook's Oversight Board upholds ban on Trump. At least for now*, The Washington Post, dated 5<sup>th</sup> May 2021, last visited on 21<sup>st</sup> January 2025 - <https://www.washingtonpost.com/technology/2021/05/05/facebook-trump-decision/>

<sup>13</sup> Elon Musk, *Reinstate former President Trump on Twitter*, Twitter poll, dated 19<sup>th</sup> November 2025, last visited on 21<sup>st</sup> January 2025 - <https://x.com/elonmusk/status/1593767953706921985?mx=2>

<sup>14</sup> *Meta lifts restrictions on Trumps Facebook, Instagram accounts*, Al Jazeera, dated 13<sup>th</sup> July 2024, last visited on 21<sup>st</sup> January 2025 - <https://www.aljazeera.com/news/2024/7/13/meta-lifts-restrictions-on-trumps-facebook-instagram>

respective platform, probably is a perfect fit of the Golden Rule – “*Do unto others as you would have them do unto you.*”

## **Free Speech in 2025**

In the digital era, social media platforms serve as pivotal arenas for the exercise of the right to freedom of speech. The recent initiative by social media platforms to promote freedom of speech in alignment with the UDHR. Companies like Meta and X have implemented various policies and tools aimed at balancing the promotion of free expression with the need to mitigate harmful content.

Meta announced a significant shift in its content moderation strategy by discontinuing its third-party fact-checking program in the United States in January 2025 as mentioned earlier, and introduced a Community Notes system, similar to X.<sup>15</sup>

Elon Musk has been a pioneer in community-driven content moderation through X. Formerly known as Birdwatch, the platforms – community notes allow users to identify and annotate potentially misleading information. These contributors must meet specific eligibility requirements and only notes rated as helpful by a diverse range of users are displayed. Also, X does not directly write, rate or moderate these notes, ensuring that the content reflects community input rather than corporate viewpoints.

This welcome shift to community-driven moderation offers several advantages<sup>16</sup> – it reduces centralised control over content, thereby fostering a more open environment where diverse viewpoints can be shared without undue interference; it empowers users to take an active role in content moderation, promoting a sense of responsibility and engagement within the platforms ecosystem and promotes transparency as decisions are made collectively rather than behind closed doors.

However, there are certain challenges which may arise such as spread of misinformation, manipulation risks and lack of consensus.

## **Conclusion**

This pluralist model, where the governments seem to control freedom of speech on social media platforms varies globally. The recent actions in the United States highlight this dynamic – as briefly discussed President Donald Trump has now signed an executive order aimed at prohibiting federal authorities from violating American citizens free speech and mandated an investigation into potential infringements during the previous administration.<sup>17</sup> This move underscored the complex interplay between governmental authority and corporate policies in regulating online speech.

<sup>15</sup> Ash Johnson, *Meta Community Notes and Content Moderation in a Free Market*, Information Technology and Innovation Foundation dated 16<sup>th</sup> January 2025, last visited on 21<sup>st</sup> January 2025 - <https://itif.org/publications/2025/01/16/meta-community-notes-and-content-moderation-in-a-free-market/>

<sup>16</sup> Joel Kaplan, *More Speech and Fewer Mistakes*, Meta dated 7<sup>th</sup> January 2025, last visited on 21<sup>st</sup> January 2025 - <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>

<sup>17</sup> Ali Swenson, *Trump orders government not to infringe on Americans speech, calls for censorship investigation*, AP news, dated 20<sup>th</sup> January, 2025, last visited on 21<sup>st</sup> January, 2025 - <https://apnews.com/article/trump-executive-order-censorship-misinformation-social-media-b344d8c75463990ff6c782bc836bb337>

While on the other hand Meta and X have committed to enhance efforts to combat illegal hate speech under a revised European Commission code integrated with the Digital Services Act in the European Union. This code of conduct on countering illegal hate speech online plus requires increased transparency in hate speech detection and a commitment to review a significant portion of hate speech notices within 24 hours. Such regulatory frameworks illustrate how governments can influence platform policies to align with broader social values.<sup>18</sup>

These approaches though offer promising avenues for enhancing free expression and user engagement, they also present challenges that require careful navigation. Perhaps, the balance between fostering open dialogue and preventing the spread of harmful content remains delicate. Further, the role of the governments in regulating online speech continues to be a contentious issue, necessitating ongoing dialogue among stakeholders to ensure that the right to freedom of expression is upheld in the digital age.



<sup>18</sup> Jess Weatherbed, *X, Facebook, Instagram and YouTube sign EU pledge to tackle hate speech*, The Verge, dated 20<sup>th</sup> January 2025, last visited on 21<sup>st</sup> January 2025 - [https://www.theverge.com/2025/1/20/24347876/x-facebook-instagram-youtube-eu-hate-speech-dsa-code?utm\\_source=chatgpt.com](https://www.theverge.com/2025/1/20/24347876/x-facebook-instagram-youtube-eu-hate-speech-dsa-code?utm_source=chatgpt.com)