

## SILENCING THE ECHO CHAMBER: UNVEILING THE NECESSITY OF SOCIAL MEDIA CENSORSHIP

---

Palak Tyagi\*

### ABSTRACT

*The paper investigates the role that social media censorship can play in creating a welcoming and safe online environment for all users. The paper analyses recent events and real-world examples to illustrate the far-reaching effects of censorship on several facets of communication in the digital realm. The first section of the study explores how censorship can be used to prevent cyberbullying and safeguard users. It exemplifies the efforts of platforms to remove inappropriate material and provide a safe and welcoming virtual space for users. The next section of the study focuses on the role censorship plays in preventing the spread of misinformation and advancing the truth. It delved into how these sites counteract the spread of fake news and other forms of disinformation, ensuring that their users have access to accurate and trustworthy content. Censorship is also examined from the perspective of a safety valve that can shield marginalised people from harm when they go online. It examines efforts to silence bigotry and hate speech so that underrepresented groups can feel welcome and secure. The study also highlights the ways in which censorship encourages diversity and inclusion by advocating for balanced portrayals. It looks at ways to make the internet more diverse and inclusive by eliminating discriminatory material and giving a greater platform to underrepresented groups. The research also touches on how open and equitable censorship procedures should be. The balance between content regulation and freedom of expression is discussed, along with the need for clear guidelines, consistent enforcement, and user feedback methods. The importance of censorship in protecting personal information is also discussed in the study. It explores how platforms implement rules to safeguard user information and sustain public faith in them. Finally, the study looks at how current events, like those in India, have affected online censorship policies. It stresses the significance of appropriate censorship in resolving local issues and maintaining public confidence and security. Read on for more information.*

---

\*FIRST YEAR, SARDAR PATEL SUBHARTI INSTITUTE OF LAW.

## INTRODUCTION

In today's digital era, social media has become a game-changing tool for people of all backgrounds to connect with one another and learn about the world around them. Social media's enormous reach and power have made it an indispensable component of modern society, reshaping public conversation, mobilising groups, and giving a voice to many who were before silence.<sup>1</sup> While social media can have far-reaching benefits, it also comes with a number of risks, such as the easy spread of false or misleading information.<sup>2</sup> Censorship on social media is now the subject of heated debate because of this.

Censorship, in the context of social media, is the monitoring and editing of user-generated content. Content that is judged unsuitable, objectionable, or possibly harmful is subject to standards, policies, and algorithms that monitor, restrict, and remove it.<sup>3</sup> The major goal of social media censorship is to safeguard users from the negative consequences of the unchecked distribution of bad content while still preserving users' rights to free expression.<sup>4</sup>

Even though free speech is essential, it cannot be unrestricted if doing so will endanger others' rights or reputations or threaten national security, public order, or public health. The boundary between free speech and hate speech is often rather thin.<sup>5</sup> If this hate speech is not censored, it will breed violence, prejudice, and the proliferation of dangerous conspiracy theories targeting vulnerable populations, further splintering any nation.

## REDUCING CYBERBULLYING AND KEEPING USERS SAFE

The potential for unrestrained content to do harm is the primary argument in favour of censoring it on social media. Hate speech, cyberbullying<sup>6</sup>, and the dissemination of violent or

---

<sup>1</sup> The NYU Dispatch, 'The Influence of Social Media on Modern Society' (The NYU Dispatch 9 September 2020) <<https://wp.nyu.edu/dispatch/the-influence-of-social-media-on-modern-society/>>.

<sup>2</sup> Priyanka Deka, 'Fake News and Social Media' (Times of India Blog 20 June 2021) <<https://timesofindia.indiatimes.com/readersblog/world-of-words/fake-news-and-social-media-33975/>>.

<sup>3</sup> 'The Impact of Algorithms for Online Content Filtering or Moderation "Upload Filters" Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies PE' (2020) <[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL\\_STU\(2020\)657101\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf)>.

<sup>4</sup> United Nations, 'Moderating Online Content: Fighting Harm or Silencing Dissent?' (OHCHR 23 July 2021) <<https://www.ohchr.org/en/stories/2021/07/moderating-online-content-fighting-harm-or-silencing-dissent>>.

<sup>5</sup> Lee Rainie, Janna Anderson and Jonathan Albright, 'The Future of Free Speech, Trolls, Anonymity and Fake News Online' (Pew Research Center: Internet, Science & Tech 29 March 2017) <<https://www.pewresearch.org/internet/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/>>.

<sup>6</sup> UNICEF, 'Cyberbullying: What Is It and How to Stop It' (www.unicef.org February 2023) <<https://www.unicef.org/end-violence/how-to-stop-cyberbullying>>.

extreme views have all found fertile ground on social media platforms.<sup>7</sup> Users might be subjected to online abuse, psychological anguish, and even physical threats in the absence of adequate filtering measures. In addition, the rapid spread of bad content can have far-reaching societal implications, including the promotion of social unrest, the deepening of polarisation, and the weakening of trust within communities.<sup>8</sup> Social media platforms can safeguard users from harm and promote a more courteous and inclusive online community if they actively monitor and moderate material.

Unfortunately, due to their large user bases and widespread influence, social media platforms have become incubators for bigotry, intolerance, and divisive beliefs and practices. When people can hide behind a computer screen and spread their bigotry to a wider audience, they are more likely to engage in hate speech and other forms of discriminatory content. Unchecked hate speech has the potential to deepen existing divisions in society, foment hostility between demographics, and even incite physical violence.<sup>9</sup>

The elimination of hate speech is one method by which social media filtering might limit the spread of bigotry. Any expression of opinion that incites hatred towards a group of people because of their race, religion, ethnicity, gender, or sexual orientation is considered hate speech.<sup>10</sup> Social media platforms can quickly address and prevent the spread of hate speech by introducing stringent policies and algorithms to detect and remove such speech. In addition to shielding those who might be singled out, this also sends a clear message that hostile conduct will not be accepted within their online communities.

Via regional and national laws, various governments have passed legislation that requires social media sites to take action against unlawful, offensive, or dangerous content and speech. Platforms must adhere to the legal requirements of the jurisdictions in which they operate, which can be different from country to country.<sup>11</sup>

---

<sup>7</sup> Zachary Laub, 'Hate Speech on Social Media: Global Comparisons' (Council on Foreign Relations 7 June 2019) <<https://www.cfr.org/background/hate-speech-social-media-global-comparisons>>.

<sup>8</sup> 'Social Media: Prospect and Challenges – Civilsdaily' <<https://www.civildaily.com/story/social-media-prospect-and-challenges/>>.

<sup>9</sup> United Nations, 'Hate Speech and Real Harm' (United Nations) <<https://www.un.org/en/hate-speech/understanding-hate-speech/hate-speech-and-real-harm>>.

<sup>10</sup> Council of Europe, 'Hate Speech and Violence' (European Commission against Racism and Intolerance (ECRI) 2014) <<https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/hate-speech-and-violence>>.

<sup>11</sup> 'Monika Bickert v P, c O N T E N T P O L I c Y F E B R U a R Y 2 0 2 0' <<https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward-Online-Content-Regulation-White-Paper-1.pdf>>.

The development and spread of extremist ideology on social media can be hindered by censorship measures. Due to their convenience and the ease with which propaganda can be disseminated and like-minded people brought together, online platforms have played a significant role in the recruitment and radicalization of individuals into extremist groups<sup>12</sup>. Social media platforms can hinder the recruitment efforts of extremist groups and reduce their influence by aggressively monitoring and deleting extremist content. This has the ability to significantly hinder the spread of harmful concepts and prevent actual harm from occurring.

Instagram, a widely used social media site, has taken measures to prevent cyberbullying and abuse.<sup>13</sup> Several features and policies have been implemented over the past few years to make the site more secure and less inviting to cyberbullying.

Instagram's comment filtering system is a vital defence mechanism. Offensive or abusive comments are identified and removed from the platform using machine learning algorithms. By blocking potentially harmful or harassing comments before they reach users' postings, this proactive censorship method helps mitigate the harmful effects of cyberbullying. Users now have more control over their online interactions with Instagram's new comment filtering and comment disabling features. With this option, people can surround themselves with positivity and safety, reducing their vulnerability to harassment and abuse. Instagram hopes to protect its users from the negative impacts of online harassment by enforcing stronger content control procedures. In addition to making users feel safer, the platform's preventative censorship measures convey a clear signal that cyberbullying will not be tolerated.

Platforms need to find a middle ground between content management and free speech, but filtering measures are essential for stopping online harassment. Censorship practises should be fair, accountable, and in line with the purpose of protecting users from damage, and this can be achieved by the implementation of transparent rules, the provision of chances for appeals, and the incorporation of user feedback. Censorship may play an important role in reducing online harassment, making the internet a safer place for users, and encouraging a more positive online community, as Instagram's comment filtering and anti-bullying measures show.

---

<sup>12</sup> Jens F Binder and Jonathan Kenyon, 'Terrorism and the Internet: How Dangerous Is Online Radicalization?' (2022) 13 *Frontiers in Psychology*.

<sup>13</sup> Instagram, 'Anti-Bullying on Instagram' (about.instagram.com) <<https://about.instagram.com/community/anti-bullying>>.

## STOPPING THE SPREAD OF FALSEHOOD AND PROMOTING ACCURACY

The widespread problem of false information and its proliferation on social media is another important factor to think about<sup>14</sup>. Individuals, communities, and even democratic processes can all suffer when false information spreads freely. Campaigns spreading false information can sway public opinion, change the results of elections, and damage people's trust in official sources of data and knowledge.<sup>15</sup> Social media platforms can prevent the spread of false information and protect their users from being misled by applying censoring mechanisms. The result may be a more educated populace, which is beneficial to the health of our democracy. Fake news and disinformation are more widely disseminated than ever because of the proliferation of social media. False or misleading information can quickly go viral and reach a wide audience thanks to the widespread nature of social media platforms and the ease with which it can be shared without sufficient verification.<sup>16</sup> Since fake news may influence public opinion, weaken faith in reputable sources, and spread misleading narratives, it poses serious threats to individuals, societies, and even democratic processes.

In order to prevent the proliferation of false information, censorship is an essential tool for managing online communities. The spread of disinformation can be slowed if social media sites provide ways to detect, flag, and remove incorrect or misleading posts. Patterns, keywords, and sources commonly used to disseminate false information can be identified with the help of algorithms and artificial intelligence techniques.<sup>17</sup> When such material is uncovered, it may be checked for accuracy using sophisticated methods, and if it is determined to be erroneous, it can be swiftly removed or labelled as unreliable. This method not only reduces the impact of fake news, but also gives people access to more reliable sources of data.

---

<sup>14</sup> Filippo Menczer and Thomas Hills, 'Information Overload Helps Fake News Spread, and Social Media Knows It' (2020) 323 *Scientific American* <<https://www.scientificamerican.com/article/information-overload-helps-fake-news-spread-and-social-media-knows-it/>>.

<sup>15</sup> Sadiq Muhammed T and Saji K Mathew, 'The Disaster of Misinformation: A Review of Research in Social Media' (2022) 13 *International Journal of Data Science and Analytics* <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8853081/>>.

<sup>16</sup> Przemysław Majerczak and Artur Strzelecki, 'Trust, Media Credibility, Social Ties, and the Intention to Share towards Information Verification in an Age of Fake News' (2022) 12 *Behavioral Sciences* 51 <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8869166/>>.

<sup>17</sup> 'Artificial Intelligence (AI) Algorithms: A Complete Overview' (Tableau) <<https://www.tableau.com/data-insights/ai/algorithms>>.

Another useful tactic in the fight against fake news is collaboration between social media companies and independent fact-checking bodies.<sup>18</sup> Social media sites can improve their accuracy ratings and verification processes by collaborating with established fact-checking organisations. By working together, censorship efforts can be better coordinated, and content can be evaluated more thoroughly and confidently. By prominently displaying verified fact-checking labels next to potentially misleading content, we can provide people with quick context and help them become more critical consumers of the information they find online.

The spread of false information on social media can be stifled through a combination of reactive and preventative actions. This includes encouraging viewers to learn how to critically consume media. Social media platforms may help people become more critical information consumers by funding educational projects and providing tools to help users distinguish between reliable and false information<sup>19</sup>. This kind of work can make the Internet a safer and more secure place by making people less likely to believe false information.

However, a middle ground must be found between overly restrictive censorship and inadequate safeguards for free expression. To prevent the stifling of free speech or the formation of echo chambers, censorship measures should be handled carefully and openly. Censorship practices can be more in line with democratic norms and tolerance for varied opinions with the use of regular assessments, periodic audits, and stakeholder dialogues.<sup>20</sup> To address concerns and ensure that censorship activities are performed in a responsible and accountable manner, open conversation and feedback channels are essential.

False information, conspiracy theories, and misleading claims about the COVID-19 virus, its origins, treatments, and vaccines have proliferated throughout the course of the epidemic. The task of combating this pervasive misinformation and halting its destructive effects has fallen to social media companies.<sup>21</sup> Social media sites like Facebook, Twitter, and YouTube have created systems to detect and eliminate harmful or misleading content. In order to verify

---

<sup>18</sup> 'Indian Government Gives Itself the Power to "Fact-Check" and Delete Social Media Posts' (Rest of World 12 April 2023) <<https://restofworld.org/2023/indian-government-fact-check-delete-social-media/>> accessed 26 June 2023.

<sup>19</sup> 'Social Media Must-Haves: Proactive and Reactive Social Strategies' (Zen Media 31 May 2017) <<https://zenmedia.com/blog/social-media-must-haves-proactive-reactive-social-strategies/>>.

<sup>20</sup> Carne Colomina and others, 'The Impact of Disinformation on Democratic Processes and Human Rights in the World @Adobe Stock EN STUDY' (2021) <[https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO\\_STU\(2021\)653635\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf)>.

<sup>21</sup> Ingjerd Skafle and others, 'Misinformation about Covid-19 Vaccines on Social Media: Rapid Review' (2022) 24 Journal of Medical Internet Research.



the information and counteract the spread of misconceptions about COVID-19, they have partnered with fact-checking organisations, health authorities, and specialists.<sup>22</sup>

For instance, websites have begun removing content and profiles that promote ineffective or harmful treatments, such as herbal remedies or pharmaceuticals that have not undergone rigorous testing. They've also taken action against conspiracies that attribute the virus's genesis to human intent or other unfounded claims. Social media sites are doing their best to protect truth and accuracy by actively banning and eliminating such misinformation, giving their users access to only the most credible and up-to-date information regarding the epidemic that is backed up by science. By limiting access to misleading content, we can lessen the likelihood that people will make poor health choices based on erroneous information or become confused in the first place.<sup>23</sup>

It is critical, however, that platforms keep their censorship practises open and accountable to the public. A middle ground must be found between limiting the spread of false information and protecting people's right to say what they want without interference. When it comes to policing and removing potentially damaging or offensive posts, social media companies frequently rely on content moderation.<sup>24</sup> Certain legal frameworks may provide platforms with protection from liability, such as United States Code Section 230<sup>25</sup>, which prevents platforms from being held liable for user-generated content. Censorship can play an important role in preventing the spread of incorrect information, safeguarding public health, and encouraging accurate knowledge, as evidenced by the actions of social media platforms in the case of COVID-19. The positive effects of disinformation on individuals and society can be reduced if platforms adhere to standards of truth and accuracy.

---

<sup>22</sup> 'Twitter Launches Labels, Warnings on Misleading COVID-19 Information' Reuters (11 May 2020) <<https://www.reuters.com/article/uk-health-coronavirus-twitter-idUKKBN22N2EG>> accessed 4 April 2023.

<sup>23</sup> OECD, 'Transparency, Communication and Trust: The Role of Public Communication in Responding to the Wave of Disinformation about the New Coronavirus' (OECD3 July 2020) <<https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/>>.

<sup>24</sup> Prasad Banerjee, 'Inside the Secretive World of India's Social Media Content Moderators' (https://www.livemint.com18 March 2020) <<https://www.livemint.com/news/india/inside-the-world-of-india-s-content-mods/amp-11584543074609.html>> accessed 26 June 2023.

<sup>25</sup> Legal Information Institute, '47 U.S. Code § 230 - Protection for Private Blocking and Screening of Offensive Material' (LII / Legal Information Institute2018) <<https://www.law.cornell.edu/uscode/text/47/230>>.

## CENSORSHIP AS A SAFETY VALVE: PROTECTING SOCIALLY EXCLUDED GROUPS ONLINE

In addition, filtering on social media is essential for safeguarding vulnerable individuals and underprivileged groups. Harassment<sup>26</sup>, cyberbullying, and prejudice can be perpetuated and exacerbated through the use of online platforms. To protect vulnerable populations from content that discriminates on the basis of race, gender, religion, or other protected characteristics, censorship can be a useful tool. Censorship of social media sites reduces harm to vulnerable people and advances equality and social justice by creating a more courteous and accepting online community. Some examples of marginalised groups that are disproportionately affected by online harassment, abuse, and violence are members of underrepresented racial and ethnic groups, the LGBTQ+ community, women, and persons with disabilities.<sup>27</sup> By rapidly addressing and eliminating content that promotes hate speech or attacks individuals based on identification, censorship protects marginalised communities. Safer online environments can be established on social media platforms by using algorithms and reporting procedures to identify and remove hate speech. Social media platforms may convey a strong message that hates speech and harassment will not be tolerated by enforcing clear norms and regulations that prohibit such speech and conduct.

Censorship is also essential in preventing the spread of biased and damaging stereotypes<sup>28</sup>. The public's thoughts and attitudes can be influenced by what they see on social media. Harmful content and misinformation, if unchecked, can further marginalise already vulnerable communities by reinforcing stereotypes and biases. By enforcing strict rules for content removal, platforms may prevent the propagation of toxic narratives and instead encourage positive portrayals of underrepresented groups.

Also, members of minority groups may feel more secure and empowered when censorship is in place. Social media companies demonstrate they care about their users' well-being and security by removing content quickly when it is shown to be discriminatory or harassing on the basis of users' identities. As a result, more people may feel safe enough to contribute to online forums, tell their stories, and express their opinions without worrying about being

---

<sup>26</sup> Emily A Vogels, 'The State of Online Harassment' Pew Research Center (13 January 2021) <<https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/>>.

<sup>27</sup> Ilan H Meyer, 'Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence' (2003) 129 Psychological Bulletin 674.

<sup>28</sup> Country Navigator and Ognjen Varadanin, 'Stereotypes, Bias, and Culture' (Country Navigator 13 February 2023) <<https://www.countrynavigator.com/blog/stereotypes-bias-and-culture/>>.



subjected to harassment or discrimination. Censorship, in this sense, contributes to making the Internet a place where everyone feels safe and valued, no matter their identity or where they come from.

The hazards and difficulties of censorship should not be underestimated, and issues of free speech and abuse of power<sup>29</sup> should be among them. Finding a middle ground between shielding vulnerable populations and allowing for free expression calls for thoughtful deliberation and a dedication to openness and responsibility. These issues can be addressed and censorship practises kept in line with democratic norms and societal diversity through routine audits, community involvement, and stakeholder participation in decision-making.

The prominent social media site Twitter has been making strides to prevent hate speech and defend vulnerable groups from cyberbullying.<sup>30</sup> Content that advocates discrimination or incites violence based on race, ethnicity, religion, gender, or sexual orientation has been subject to tougher regulations and removal from the platform in recent years.

In one famous instance, Twitter took a stand against content that promoted white supremacy.<sup>31</sup> White supremacist accounts and tweets were deleted as the platform began strictly enforcing its regulations. This move was made to shield people from abuse, dehumanisation, and the spread of bigotry. Twitter's goal in eliminating white supremacist content was to make the platform more welcoming to marginalised groups that are frequently the targets of hate speech and other forms of discrimination. By taking such preventative measures to silence hate speech, we can lessen the emotional and psychological damage caused by online harassment and create a more welcoming and safe online environment for all.

The suppression of white nationalist information on Twitter is an example of how censorship can protect vulnerable internet groups. Platforms can reduce the spread of dangerous ideologies, shield users from online abuse, and make marginalised communities feel more welcome and secure if they adopt and strictly enforce standards against hate speech and discriminatory content. To make sure that censorship policies are effective in addressing the

---

<sup>29</sup> Ilan H Meyer, 'Prejudice, Social Stress, and Mental Health in Lesbian, Gay, and Bisexual Populations: Conceptual Issues and Research Evidence' (2003) 129 *Psychological Bulletin* 674.

<sup>30</sup> Sophia Bernazzani, 'How Twitter Is Fighting Harassment & Cyberbullying' (Hubspot.com2017) <<https://blog.hubspot.com/marketing/twitter-harassment-cyberbullying>>.

<sup>31</sup> 'Twitter Won't Treat White Supremacy like ISIS Because It'd Have to Ban Some GOP Politicians Too' (www.vice.com) <<https://www.vice.com/en/article/a3xgg5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too>>.

unique obstacles encountered by marginalised communities, social media platforms should continue to refine their censorship practices, engage with affected communities, and solicit advice from advocacy groups. By doing so, platforms may help foster an online community where people from all walks of life are treated fairly and their opinions are valued.

## **PROMOTING DIVERSITY AND INCLUSION BY FIGHTING FOR FAIR REPRESENTATION**

Online social networking When used to aggressively challenge discriminatory content and negative stereotypes, censorship can help promote inclusiveness and diversity.<sup>32</sup> Content that specifically targets users because of their race, ethnicity, gender, sexual orientation, religion, or other protected characteristics can be censored by platforms. By actively combating discriminatory and hateful content, platforms foster an atmosphere that values the worth and rights of all users. This encourages more diverse voices to be heard, gives voice to underrepresented groups, and makes the internet a more welcoming place for everybody.

YouTube's Efforts to Counter Harmful Speech and Content Regarding the LGBTQ+, YouTube has taken action to prevent the spread of hate speech and advance diversity in its user base. The platform has rules in place to delete anything that promotes bigotry or targets vulnerable communities, like the LGBTQ+ population.<sup>33</sup> In one high-profile instance, YouTube was criticised for allowing content that was both disrespectful and discriminatory towards the LGBTQ+ community. The platform updated its community guidelines to outlaw hate speech and harassment based on sexual orientation or gender identity after realising the harm such content can cause. YouTube's goal is to create a more welcoming community for its users, and it does this by aggressively banning and eliminating videos that promote hate speech. This method is useful in preventing online discrimination, harassment, and abuse against marginalised groups. It also sends a strong signal that the platform values diversity of opinion and is actively working to eliminate bigotry.<sup>34</sup>

---

<sup>32</sup> 'Freedom of Expression Philosophy - Diversity & Inclusion - Gettysburg.edu' (www.gettysburg.edu) <<https://www.gettysburg.edu/offices/diversity-inclusion/freedom-of-expression-policy>> accessed 26 June 2023.

<sup>33</sup> 'Social Media Platforms Fail to Stop LGBTQ Hate Speech, according to GLAAD Report' (PBS NewsHour 13 July 2022) <<https://www.pbs.org/newshour/amp/nation/social-media-platforms-fail-to-stop-lgbtq-hate-speech-according-to-glaad-report>>.

<sup>34</sup> 'Updates on Our Efforts to Make YouTube a More Inclusive Platform' (blog.youtube) <<https://blog.youtube/news-and-events/make-youtube-more-inclusive-platform/>>.

In addition, YouTube has added functions like restricted mode<sup>35</sup>, which lets users restrict access to content they deem objectionable. By limiting access to potentially hazardous or offensive content, this feature allows users to craft a viewing environment that is both welcoming and safe for all viewers. Actions like YouTube's crackdown on hate speech and discriminatory videos show how censorship may help promote diversity and inclusion. By censoring messages that promote bigotry, platforms can become safer and more inclusive places for marginalised groups to express themselves. This allows people to feel safe participating in online communities and forums without worrying about being singled out or harassed for their opinions.

To prevent the unjustified silencing of legitimate content and dissenting ideas, it is essential that platforms find a middle ground between censorship and freedom of expression. To guarantee that censoring practices are equitable, responsible, and in line with the purpose of encouraging inclusivity and diversity, there must be clear regulations, transparency in enforcement, and regular contact with affected populations.

### **MAINTAINING FAIRNESS AND TRANSPARENCY BY ENSURING ACCOUNTABLE SOCIAL MEDIA CENSORSHIP**

While it's clear that social media filtering is necessary, there are valid worries about its potential for misuse and its effect on free expression. It's undeniably difficult to find a happy medium between censorship and free speech. Therefore, it is essential to set forth clear standards, include relevant parties in the decision-making process, and guarantee accountability in the rollout of any censorship policies. Social media platforms can reduce the likelihood of censorship being used to silence dissent or stifle lawful expression if they adhere to the values of transparency, fairness, and proportionality.<sup>36</sup>

It is crucial when imposing restrictions on social media platforms to strike a balance between censorship and the protection of free expression. Censorship is intended to prevent the spread of dangerous information, but it must be carefully monitored to prevent it from being used to silence valid criticism or speech. Openness is a cornerstone of ethical censorship policies and procedures. It is imperative that social networking sites lay out their content moderation

---

<sup>35</sup> Ben Stockton, 'What Is YouTube Restricted Mode and How to Enable or Disable It' (18 May 2021) <<https://www.online-tech-tips.com/computer-tips/what-is-youtube-restricted-mode-and-how-to-enable-or-disable-it/amp/>> accessed 26 June 2023.

<sup>36</sup> United Nations, 'Moderating Online Content: Fighting Harm or Silencing Dissent?' (OHCHR23 July 2021) <<https://www.ohchr.org/en/stories/2021/07/moderating-online-content-fighting-harm-or-silencing-dissent>>.

policies and guidelines in detail so that users can fully grasp what is and isn't allowed on the site. Sharing these guidelines publicly helps platforms gain users' trust and empowers them to make well-informed decisions about their interactions with others online.

Furthermore, censorship initiatives should put fairness first. Without regard to a user's race, religion, gender, sexual orientation, nationality, or socioeconomic background, platforms should apply the same set of rules and regulations to everyone. Discrediting censorship measures with favouritism or manipulation allegations due to biased or selective enforcement. Censorship policies can be implemented more fairly if content moderators are subjected to regular internal audits, external assessments, and diversity training. The principle of proportionality<sup>37</sup> is also important in ethical censorship. When taking action, platforms should consider the gravity of the content or behaviour at hand. A nuanced strategy is required to strike a balance between user safety and the right to free expression. When deciding whether or not to censor information, platforms should take into account elements such as context, intent, and possible impact. A blanket policy<sup>38</sup> could end up stifling free speech; therefore, it's important to look at each situation individually.

Censorship practices can only be trusted if those responsible are held accountable. Users who are negatively impacted by censorship choices should be able to provide comments and file appeals with social media networks. To ensure that mistakes are corrected, false positives are addressed, and platforms learn and improve their moderation procedures, it is important to have explicit avenues for users to voice issues or seek redress. In March 2019, a gunman opened fire on two mosques in Christchurch, New Zealand, resulting in the sad loss of 51 lives and injuries to many more<sup>39</sup>. This instance demonstrates the true relevance of censorship. The entire attack was broadcast live on Facebook, terrifying and grossing out viewers.<sup>40</sup>

This incident highlighted the critical necessity for online communities to tackle extremist information and curb the spread of violent incitement. As a result, many sites came under fire

---

<sup>37</sup> 'The Principle of Proportionality' (Diakonia International Humanitarian Law Centre) <<https://www.diakonia.se/ihl/resources/international-humanitarian-law/ihl-principle-proportionality/>>.

<sup>38</sup> 'The Bullet: The Convenience of Blanket Policies — How Thailand's One-Size-Fits-All Approach Falls Short' (KrASIA1 February 2023) <<https://kr-asia.com/the-bullet-the-convenience-of-blanket-policies-how-thailands-one-size-fits-all-approach-falls-short>> accessed 26 June 2023.

<sup>39</sup> 'Christchurch Gunman "Planned Third Mosque Attack"' (BBC News23 August 2020) <<https://www.bbc.com/news/world-asia-53861456.amp>> accessed 26 June 2023.

<sup>40</sup> Graham Macklin, 'The Christchurch Attacks: Livestream Terror in the Viral Video Age' (Combating Terrorism Center at West Point18 July 2019) <<https://ctc.westpoint.edu/christchurch-attacks-livestream-terror-viral-video-age/>>.

for their part in spreading the live stream and its subsequent material. To discover and remove violent or extremist content more quickly, social media sites like Facebook, YouTube, and Twitter have revised their content moderation standards. This incident emphasised the need for balanced censorship that safeguards the public from harmful content while upholding the values of free speech. Since then, platforms have spent money on AI algorithms, human moderation teams, and collaborations with outside organisations in order to detect and delete extremist and violent content. There is now more openness in reporting and enforcement, as well as stronger community guidelines and broader definitions of what constitutes prohibited content.

Responsible censorship is essential in preventing the spread of violent and extremist information, as seen by the recent massacres at mosques in New Zealand. By eliminating such content proactively, social media may help stop the normalisation and instigation of violence, shield the vulnerable from damage, and create a more secure virtual space. However, platforms also have a responsibility to be open and honest about their policing procedures. Censorship measures must be enforced uniformly and without bias; thus, it's important to conduct regular audits, invite external evaluations, and provide user feedback methods. Platforms can better moderate their services, respond to user complaints and correct errors by maintaining open lines of communication with users and other stakeholders.

### **PRIVACY AND DATA SECURITY: KEEPING USER DATA SAFE**

Censorship on social networking sites is also very important for keeping people's private information private. Without adequate censoring mechanisms, huge amounts of user data collected by online platforms might be accessed, misused, or exploited by malicious actors. Platforms can aid in preventing the disclosure of personally identifiable information (such as addresses, bank details, or intimate images) by enforcing stringent content moderation policies. This ensures the safety of the online community by preventing incidents of identity theft, harassment, and other forms of digital abuse.

In this regard, in the 2018 Cambridge Analytica Scandal<sup>41</sup>, it was revealed that the political consulting firm Cambridge Analytica was involved in a massive data privacy scandal. In this case, millions of Facebook users' private information was improperly obtained and used for

---

<sup>41</sup> Nicholas Confessore, 'Cambridge Analytica and Facebook: The Scandal and the Fallout so Far' The New York Times (4 April 2018) <<https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>>.

electoral purposes. Through a third-party app called "This Is Your Digital Life"<sup>42</sup>, Cambridge Analytica gathered private information on Facebook users and their friends without their knowledge or approval. Not only did the software steal information from its users, but it also gained unauthorised access to the profiles of their friends.

Cambridge Analytica is accused of using the collected data to create psychological profiles of users and then sending them targeted political ads during the 2016 U.S. presidential election campaign and the 2016 Brexit vote.<sup>43</sup> The issue prompted serious questions about data security, privacy, and the possibility of political manipulation through data exploitation. Calls for tougher data privacy legislation arose in response, and agencies like the US Federal Trade Commission (FTC)<sup>44</sup> launched investigations. Facebook's data protection practises and the way it handled user information came under severe scrutiny and criticism after the Cambridge Analytica crisis. Many consumers, policymakers, and tech businesses were alerted to the need to protect users' personal information after the tragedy. This real-world scenario shows how serious the consequences may be when private information is compromised. Strong privacy legislation, openness about data practices, and individual agreement for the storage and use of personal information by businesses and other organisations are all stressed.

### **RECENT INCIDENTS IN INDIA HIGHLIGHT THE NEED FOR SOCIAL MEDIA FILTERING TO PRESERVE TRUST AND SAFETY.**

Seventy percent of people who took the 2020 survey early stated they were using social media more often. Twenty-five percent of respondents to the survey thought life was basically the same as it had been before the epidemic. The vast majority of the 2021 survey's second-wave respondents (89%) said they were using social media more frequently or at the same level as before. Social networking has become a major time sink for many young people.<sup>45</sup> In 2018, the Pew Research Centre polled more than 750 American teenagers aged

---

<sup>42</sup> 'Facebook App "This Is Your Digital Life" Collected Users' Direct Messages: Report - National | Globalnews.ca' (globalnews.ca) <<https://globalnews.ca/news/4143810/aleksandr-kogan-this-is-your-digital-life-messages/amp/>> accessed 26 June 2023.

<sup>43</sup> 'Cambridge Analytica Helped "Cheat" Brexit Vote and US Election, Claims Whistleblower' Politico Europe (8 September 2022) <<https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/amp/>>.

<sup>44</sup> Federal Trade Commission, 'About the FTC' (Federal Trade Commission 7 February 2019) <<https://www.ftc.gov/about-ftc>>.

<sup>45</sup> Anshu Siripurapu and William Merrow, 'Social Media and Online Speech: How Should Countries Regulate Tech Giants?' (Council on Foreign Relations 9 February 2021) <<https://www.cfr.org/in-brief/social-media-and-online-speech-how-should-countries-regulate-tech-giants>>.



13 to 17, finding that over half of them go online virtually nonstop and that nearly all of them use at least one kind of social media.<sup>46</sup> Is this the best time to introduce children to explicit sexual content and violent media? Are we to encourage such naive individuals to participate in riots, wars, and atrocities like torture, murder, and rape? Because they are humans, all children are vulnerable to negative influences. Minority youngsters, whose delicate minds are still growing, benefit greatly from censorship.

As an example, let's look at China. As we have seen, censorship is the path to peace and prosperity. China is home to 56 distinct ethnic groups<sup>47</sup>, each with its own culture, traditions, and beliefs. Even with these differences, however, religious and ethnic tensions in the country do not explode into violence to the same degree that they do in India<sup>48</sup>. Can they not have a say in the matter? No, they just have reasonable censorship standards in place to maintain social harmony. More recent examples include demonstrations at Jamia Millia and Aligarh Muslim University in opposition to India's citizenship amendment statute. Nobody knows why political gangs burned down everything from public buses to police stations when they made it an issue. Overnight, social media administrators failed to foresee that Muslims may interpret the act and law negatively if they were exposed to the way they were being portrayed online.<sup>49</sup>

Another example of the need for censorship is the recent demonstration by Khalistani, which took place even in Canada, Australia, and the United States.<sup>50</sup> The Farm law of 2021<sup>51</sup> is at the heart of the movement to separate Punjab from India. Everyone in six circles around the world was furious with India, but nobody knew why. After arresting over a hundred people,

---

<sup>46</sup> Monica Anderson and Jingjing Jiang, 'Teens, Social Media & Technology 2018' (Pew Research Center 31 May 2018) <<https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/>>.

<sup>47</sup> China Discovery, 'Chinese Ethnic Groups, Ethnic Groups in China' (Chinadiscovery.com 2019) <<https://www.chinadiscovery.com/ethnic-minority-culture-tour/ethnic-minorities-in-china.html>>.

<sup>48</sup> 'Indian Minister Says Protests Stop Country Becoming like China' (South China Morning Post 22 January 2019) <<https://amp.scmp.com/news/china/society/article/2183210/too-many-protests-prevent-india-becoming-china-or-singapore-says>> accessed 26 June 2023.

<sup>49</sup> 'Jamia Violence, Amended Citizenship Act and Speculation over Nationwide NRC' (India Today) <<https://www.indiatoday.in/amp/news-analysis/story/jamia-violence-amended-citizenship-act-and-speculation-over-nationwide-nrc-1628621-2019-12-16>> accessed 26 June 2023.

<sup>50</sup> 'Protests over Action against Amritpal Singh Erupt in 4 Countries, India Objects' (Hindustan Times 21 March 2023) <<https://www.hindustantimes.com/india-news/protests-over-action-against-amritpal-singh-erupt-in-four-countries-india-objects-101679339467237-amp.html>> accessed 26 June 2023.

<sup>51</sup> 'The Farm Laws Repeal Bill, 2021' (PRS Legislative Research) <<https://prsindia.org/billtrack/the-farm-laws-repeal-bill-2021>>.

authorities were forced to disable mobile internet. People like Amritpal Singh<sup>52</sup> use social media to their advantage. Inciting sectarian strife among Indians and other minorities through the use of free speech by criminals and terrorists forces us to face the fact that we are not a cohesive nation. Chandigarh's video leak case was particularly heinous because it involved the unauthorised dissemination of obscene footage showing roughly 60 female students bathing in their college dorm bathroom. Several of the students who witnessed the footage later committed suicide as a result of the incident.<sup>53</sup> Censorship on social media platforms would have prevented that.

## CONCLUSION

Hatred bleeds out when this machine that we built gets too powerful and global. By censoring, we will limit the flow of trolling, bullying, and, most importantly, violence. Social media has the potential to spark a world war, and if the government does not take care, doomsday will come any time soon. Censorship allows us to put an end to this hatred before it begins. This limiting attitude creates more opportunities for equality. The United Nations chief has acknowledged that a sizable fraction of humanity possesses the capability of causing mass destruction<sup>54</sup>. Now what about the laws that require us to protect our planet? Let's start by censoring one of them.

To the extent that it is necessary to maintain public safety, safeguard vulnerable populations, or prevent widespread disturbance of accepted social norms, the law permits censorship of this kind. Censorship on social media is a powerful tool for tackling issues on a local level, such as cyberbullying, misinformation, exclusion, accountability, privacy, and security. Finding a happy medium between unrestricted free speech and reasonable content control is essential if we want the internet to be a place where everyone can feel secure, respected, and included. To keep up with the changing demands and expectations of users all across the world, social media platforms must constantly improve their filtering procedures. Since people from all over the globe use social media platforms, it is imperative that those who run

---

<sup>52</sup> 'Amritpal Singh: Who Is He and Why Was He Arrested?' (www.aljazeera.com) <<https://www.aljazeera.com/amp/news/2023/4/23/amritpal-singh-who-is-he-and-why-was-he-arrested>> accessed 26 June 2023.

<sup>53</sup> '12 Clips Recovered, 1 More Accused Identified in Chandigarh University Video Leak Case' (India Today) <<https://www.indiatoday.in/amp/india/story/chandigarh-university-video-leak-case-12-clips-recovered-1-more-accused-identified-2002347-2022-09-20>> accessed 26 June 2023.

<sup>54</sup> 'As Critical Nature Talks Open, UN Chief Says 'Humanity Has Become a Weapon of Mass Destruction' - World News' (www.wionews.com) <<https://www.wionews.com/world/as-critical-nature-talks-open-un-chief-says-humanity-has-become-a-weapon-of-mass-destruction-540741/amp>> accessed 26 June 2023.

them are familiar with and adhere to the legal and cultural norms of each country in which they conduct business. Platforms should consider regional variances in laws and regulations when applying censorship rules to ensure consistency and fairness.

