BEFORE THE
FEDERAL TRADE COMMISSION
WASHINGTON, D.C.

Request for Public Comment Regarding	**)**
Technology Platform Censorship	**)**
FTC-2025-0023	**)**

Comment of the International Justice Clinic of the University of California, Irvine School of Law* and ARTICLE 19

* This document does not reflect the official position of the University of California, Irvine.

I. Introduction: Freedom of Expression and Content Moderation

In February 2025, the Federal Trade Commission (FTC) launched a public inquiry to consider how technology platforms deny or degrade (such as by "demonetizing" and "shadow banning") users' access to services based on the content of the users' speech or their affiliations, including activities that take place outside the platform.[1] The following comments reflect research and conclusions drawn by the International Justice Clinic at UC Irvine School of Law and the global freedom of expression organization, ARTICLE 19.

Generally speaking, it is of course true that the United States has an obligation to resist censorship and to protect the right to freedom of expression for all individuals within its jurisdiction. However, there is little in the call for comments to suggest that the FTC is genuinely committed to advancing human rights or safeguarding online freedom of expression. Instead, its call appears focused on exerting extra-legal pressure to align platforms with administration-preferred viewpoints, rather than promoting content moderation grounded in human rights principles.

Content moderation includes the different sets of policies, measures, and tools that internet companies of all kinds, including social media companies, use to deal with content on their platforms that is either illegal under domestic law or in violation of the companies' own terms of service. The specific policies and enforcement practices can vary from company to company. Generally speaking, content moderation may be influenced by a series of factors, such as a company's business models, pressure from advertisers and regulators to limit particular categories of speech and the goal of facilitating freedom of expression. Although content moderation practices have been criticized for not aligning with global human rights standards concerning freedom of expression, companies often deploy their content moderation practices to protect users and the broader public from harassment, child endangerment, incitement to violence, and privacy interference. The absence of content moderation can also lead to the intimidation and ultimately silencing of certain voices and viewpoints online, in particular those of marginalized communities.

Human rights organizations have consistently called on social media companies to align the ways in which they moderate content – both in terms of their rules as well as their enforcement - with international human rights standards respecting freedom of expression. This requires transparency, consistency, and a focus on user rights to understand and appeal content moderation decisions. It may also require the restriction of certain types of expression, such as hate speech that constitutes incitement to discrimination, hostility, or violence, that may be prohibited under international law.[2] Generally equating content moderation with censorship is an oversimplification and an incorrect way to characterize the

---

[1] FTC, *Request for Public Comment Regarding Technology Platform*, https://www.ftc.gov/system/files/ftc_gov/pdf/P251203CensorshipRFI.pdf.

[2] *See for example*, Article 20 of the ICCPR prohibiting propaganda for war and any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, or Article 3(c) of the Genocide Convention prohibiting "direct and public incitement to commit genocide."

practice. Rights-based content moderation is not inherently censorship – the same way that speech may legitimately be restricted under domestic First Amendment law and international freedom of expression standards - even if it may result in actions against users' posts or accounts.

Extra-legal pressure by the government, such as that seen in the Trump administration's policies, to intimidate social media companies into scaling back their use of content moderation will not result in more freedom of expression on these platforms. Instead, it risks creating an environment prone to politicized moderation, where platforms may feel pressured to suppress viewpoints perceived to oppose those of the government in power. The U.S. government, consistent with international human rights law and domestic law governing free speech, should resist its evident temptation to incentivize viewpoint-based content moderation, an approach that would result in less rights-based moderation and ultimately less, not more, freedom of speech.

II. Background: Sources of Freedom of Expression Rights and its Applications to Content Moderation

Ever since the adoption of Section 230 of the Communications Decency Act of 1996,[3] US law has reflected the understanding that internet platforms of all kinds enjoy rights to develop their brand, host third-party content, and engage in content moderation (or not) in order to establish forums for discussion and innovation, in which the freedom of expression may thrive.[4]

At the dawn of the age of social media, companies were hesitant to utilize content moderation.[5] Social media companies hypothesized that content moderation would hinder the expansion of its user base, hence interfering with their goal of maximizing user-generated content that would attract advertisers. However, the companies quickly realized that without rules, harassment, racism, and incitement of violence was a greater threat to maintaining a large user base and recruiting and maintaining advertisers and advertising revenue, the basis of social media's economic success. In response, social media companies created content moderation rules with the intention to make their platform experience as conducive to as many participants as possible.

Content moderation practices are constantly evolving. Generally speaking, content moderation includes different sets of measures and tools that social media platforms use to deal with illegal content and enforce their community standards over user-generated content on their service. This generally involves flagging by users, 'trusted flaggers', or 'filters'; removal, labelling, down-ranking, or demonetization of content; or disabling certain features.[6] The process can result in content being removed by a moderator or through automated moderation tools that implement content policies and terms of service. Given the scale of many companies, and opacity of their decision-making, they often make errors even if the general

---

[3] 47 U.S.C.A. § 230.
[4] *See* Jeff Kosseff, *Twenty-Six Words That Created the Internet*, Cornell University Press, 1st edition (April 15, 2019).
[5] David Kaye, *Speech Police: The Global Struggle to Govern the Internet*, Columbia Global Reports (June 3, 2019).
[6] ARTICLE 19, Content moderation handbook [hereinafter *Content moderation handbook*] (Aug. 2023), page 24. https://www.article19.org/wp-content/uploads/2023/08/SM4P-Content-moderation-handbook-9-Aug-final.pdf.

processes of content moderation and ensuring "trust and safety" online are fully legitimate responses to the goal of expanding user-generated content.

A rights based approach to content moderation for global platforms, one in which freedom of expression is guaranteed and limitations are narrow, should be based on Article 19 of the International Covenant on Civil and Political Rights ("ICCPR"), ratified by the United States in 1992, which guarantees the freedom of expression while providing the possibility of only lawful, necessary and proportionate restrictions for legitimate purposes.

Article 19 provides:

1. *Everyone shall have the right to hold opinions without interference.*
2. *Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.*
3. *The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:*
    a. *For respect of the rights or reputations of others;*
    b. *For the protection of national security or of public order (ordre public), or of public health or morals.* [7]

Article 19 protects a robust freedom to seek, receive, and impart information and ideas of all kinds, across borders and through any media of one's choice. However, Article 19(3) clearly establishes that the right to freedom of expression may be limited under narrow, rule-based circumstances. Thus, any limitation on free expression, including those aimed at online platforms, must meet a three-part test focused on legality, necessity and proportionality, and legitimacy. This three part test helps fend off improper limitations, such as politicized restrictions of freedom of expression.

Social media companies present themselves as platforms for freedom of expression. It is clear that in order to protect the right to freedom of expression, social media companies should implement a rights-based approach to content moderation. Some social media and internet companies have at least publicly endorsed their responsibilities, referring to such documents as the United Nations Guiding Principles on Business and Human Rights.[8]

---

[7] International Covenant on Civil and Political Rights (Dec. 16, 1966), 999 U.N.T.S. 171.

[8] The Guiding Principles on Business and Human Rights, which are non-binding guidelines for countries and companies, provide guidance on how to prevent, address, and remedy abuses committed during business operations. Companies have the responsibility to respect the right to freedom of expression by avoiding, causing, or contributing to adverse human rights impacts through company activities. Additionally, companies have the responsibility to seek to prevent or to mitigate adverse human rights impacts that are directly linked to company operations. These principles encourage companies to have policies and procedures in order to identify and remediate any adverse human rights impact of its company's operation or to which they contribute. *See* United Nations, Guiding Principles on Business and Human Rights, HR/PUB/11/04, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

When evaluating whether a particular post or account should be restricted, social media companies should seek to integrate the three-part test of Article 19 into its content moderation practices. Indeed, restrictions, as noted, may be only narrowly permitted, but even this is not categorically different from the law in the United States, where even in the context of public off-line speech the implementation of strict scrutiny enables public authorities to impose narrowly-tailored time, place and manner limitations.[9]

While social media companies are in principle free to restrict content on the basis of freedom of contract, they should still respect human rights, including the rights to freedom of expression, privacy, and due process in line with the UN Guiding Principles.[10] In order to meet the legality principle, social media companies should thoroughly explain their rules and explain what factors are used to assess what types of content will be restricted.[11] In order to ensure that users understand a social media company should disclose data and examples that provide insight into the factors they assess in determining a violation, its severity, and the action taken in response.[12] In order to satisfy the necessity and proportionality principle, a social media company should take the least restrictive action necessary.[13] As described in General Comment 34, "restrictive… must be appropriate to achieve their protective function… they must be proportionate to the interest to be protected."[14] To meet the legitimacy principle, social media companies should only restrict content to protect the rights or reputations of others and for the protection of national security, of public order (*ordre public*), or of public health or morals. Essentially, when social media companies identify "a legitimate ground for restriction of freedom of expression, it must demonstrate in specific and individualized fashion the precise nature of the threat… in particular by establishing a direct and immediate connection between the expression and the threat."[15]

Furthermore, as a safeguard to abusive content moderation and to protect freedom of expression, there should be transparency at all stages of a social media company's operation, from its rule-making process to the analysis as to why a certain post was taken down. Transparency demands that social media companies provide notice to individuals whose content has been moderated as well as an appeal process. Finally, companies should provide effective remedy for individual's whose content has been wrongfully restricted, such as reinstatement of their profiles or posts (Article 2(3)).

A rights-based approach to moderation would help regulate content that would otherwise undermine the rights of others to participate in public-facing platforms. Furthermore, rather than social media companies deferring to governments on issues of expression rights, with some implementing rules that might at times compromise free speech, aligning the terms of service with a rights-based approach will give the platforms a basis to resist government-driven demands for unlawful censorship.

---

[9] *See generally* Victoria L. Kilion, *Freedom of Speech: An Overview*, Library of Congress (Sept. 13, 2024).
[10] Content moderation handbook, at 27.
[11] David Kaye, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/38/35 (Apr. 6, 2018), https://docs.un.org/en/A/HRC/38/35.
[12] *Id*.
[13] *Id*.
[14] See Human Rights Committee, General Comment No. 34, Article 19: Freedoms of Opinion and Expression, UN Doc. CCPR/C/GC/34 (Sept. 12, 2011), https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.
[15] *Id*.

In order to fulfill its positive obligation to protect the right to freedom of expression, the United States has a specific role: promote a rights-based approach to content moderation.

III. Social media companies need to adopt a rights-based approach

The following two examples demonstrate why it is critical that content moderation is done properly and in accordance with corporate responsibility to promote and protect fundamental rights.

The first concerns Facebook's role in Myanmar. In 2017, the Myanmar military forced hundreds of thousands of Rohingya Muslims to leave Myanmar. The UN High Commissioner for Human Rights called it a "textbook example of ethnic cleansing."[16] To help facilitate this ethnic cleansing, the Myanmar military and militant anti-Muslim Burmese groups utilized Facebook to reinforce discrimination and violence against the Rohingya community. Anti-Rohingya groups flooded Facebook with false information that the Rohingya were invaders. In 2018, the UN Human Rights Council stated in a report, "Facebook has been a useful instrument for those seeking to spread hate, in a context where, for most users, Facebook is the internet."[17] In November 2018, Meta contracted with Business for Social Responsibility (BSR) to provide an assessment on the role of Meta's services in Myanmar. BSR conducted documentation review, direct consultation with around 60 potentially affected rights holders and stakeholders during two visits to Myanmar by BSR staff, and interviews with relevant Facebook employees.[18] The assessment concluded that the prevalence of hate speech, disinformation, and bad actors on Facebook had a negative impact on freedom of expression, assembly, and association for Myanmar's most vulnerable users and that Facebook had become a useful platform for those seeking to incite violence and cause offline harm.[19]

The case of Myanmar is unfortunately not unique – social media companies have been criticized in the contexts of other conflicts as well for insufficient moderation of inciting content or 'disinformation' that exacerbated conflict dynamics and offline violence.[20] Efforts to address such impacts should not be seen as censorship but as parts of a policy to address online and offline harms to others' rights, including to others' freedom of expression.

The second example pertains to Syria and the role of YouTube. In 2011, videos of the conflict in Syria were uploaded to YouTube. The videos often contained images of violent death and scenes of war. During this time, YouTube's policy prohibited the sharing of videos or images of "dead bodies or similar things intended to shock or disgust."[21] YouTube took down these videos swiftly on the basis that they

---

[16] United Nations, *UN human rights chief points to 'textbook example of ethnic cleansing' in Myanmar* (Sept. 11, 2017), https://news.un.org/en/story/2017/09/564622-un-human-rights-chief-points-textbook-example-ethnic-cleansing-myanmar.
[17] United Nations, *Report of the independent international fact-finding mission on Myanmar*, (Sept. 12, 2018), https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf.
[18] BSR, *Human Rights Impact Assessment: Facebook in Myanmar* (Oct. 2018), https://about.fb.com/wp-content/uploads/2018/11/bsr-facebook-myanmar-hria_final.pdf.
[19] *Id.*
[20] See ARTICLE 19, *Clearing the Fog of War* [hereinafter *Clearing the Fog of War*] (2024), at 36, https://www.article19.org/wp-content/uploads/2021/07/Clearing-the-Fog-of-War-3-December-2024.pdf.
[21] *Supra* note 5 at 22-23.

may be viewed as "celebratory of the violence itself."[22] However, this led to destruction of evidence that was working to bring attention to the human rights abuses occurring in Syria or that could have one day been used to indict war criminals behind such vicious attacks. In response to the takedowns, volunteers created the Syrian Archive. However, YouTube's algorithmic automation to take down videos made it difficult for the Archive to capture these videos.[23] Documentation of crimes and human rights violations that focus on conflicts are vulnerable to the same over-removal of content.

Engaging in rights-based content moderation is particularly challenging during armed conflicts, where the visibility of extremely violent content can be deeply distressing to some users. At the same time, particularly in contexts marked by severe repression and censorship, social media can become a primary channel for reporting events on the ground and documenting potential violations of international humanitarian law and international human rights law. There is a compelling public interest in understanding how hostilities are conducted, identifying potential human rights violations, and preserving evidence of possible atrocity crimes. This makes content moderation a challenging task for social media companies, which requires a process of heightened human rights due diligence, an in-depth understanding of the conflict dynamics and actors, sufficient allocation of resources, including human moderators, limited reliance on automated tools and ensuring regular verification of their accuracy and impartiality.[24]

These two examples illustrate the critical importance of an intentional approach to engaging in content moderation—and of doing so in a manner that respects human rights of users and the broader public. Content moderation practices, especially in times of armed conflict or in the context of political unrest and political repression, can have far-reaching implications on human rights. A rights-respecting approach requires that moderation practices be transparent, accountable, and grounded in international human rights standards. This is what the US government should promote.

IV. Evidence does not indicate that content moderation operates to limit specifically conservative viewpoints.

Shortcomings in companies' content moderation practices will no doubt have affected users with conservative viewpoints - as it has users with liberal viewpoints. Yet there is no scientific evidence finding that content moderation decisions are motivated by animus toward conservative viewpoints and ideologies.[25] Even where speech reflecting political viewpoints is restricted, it does not automatically mean that this was because of the speaker's position on the political spectrum. Two examples, often cited as evidence of social media companies' engaging in viewpoint based restrictions of speech, in fact show the opposite.

---

[22] *Id.* at 27.
[23] *Id.* at 25.
[24] Clearing the Fog of War, at 36.
[25] *See, e.g.,* Barrett, Paul M. and Sims, J. Grant, *False Accusations: The Unfounded Claim that Social Media Companies Censor Conservatives* [hereinafter *Unfounded Claim*], (Feb. 2021), NYU/STERN: Center for Business and Human Rights, at 6.

Consider first the restriction of Donald Trump's account in 2021. On January 6, 2021, a mob forcibly entered the Capitol Building, causing five deaths and numerous injuries.[26] During these events, President Trump, who had a strong presence across different social media platforms, published a post on Facebook, a post on Instagram[27], and several tweets on Twitter in support of the mob.[28]

Meta removed the post on Facebook and Instagram within hours for violation of Meta's "Community Standards on Dangerous Individuals and Organizations" and later that day blocked Trump's account for 24-hours. The next day on January 7, 2021, Meta extended the block it placed on Trump's accounts "indefinitely and for at least the next two weeks until the peaceful transition of power is complete." On January 8, 2021, Twitter issued a "permanent suspension" of Trump's account, justifying its suspension due to two tweets Trump made on January 8, 2021 that violated Twitter's "Glorification of Violence Policy". On January 12, 2021 Youtube suspended Trump's account indefinitely due to concerns for "violence".[29]

Meta's "Dangerous Individuals and Organizations Policy" prohibits "content that praises, supports, or represents events that Meta designates as terrorist attacks, hate events, mass murders or attempted mass murders, serial murders, hate crimes and violating events" (emphasis added).[30] Here, Meta deemed the storming of the Capitol as a "violating event" due to the violence that ensued. Trump's words of "we love you. You're very special", "great patriots", and "remember this day forever", targeted at the rioters, qualified as praise or support of individuals involved in a violating event. Moreover, when the case was referred to Meta's Oversight Board, the Board confirmed that, in context of the continuing riots where people died, lawmakers were put at serious risk of harm, and a key democratic process was disrupted, these supportive posts severely violated its ban on supporting violating events.[31]

Twitter similarly justified its ban by stating that its "Glorification of Violence policy" prohibits the glorification of violence that could inspire others to replicate violent acts. Here Twitter assessed that the posts made on January 8, 2021 that those who voted on him "will have a GIANT VOICE long into the future" and that they "would not be disrespected or treated unfairly in any way, shape, or form!!!" and that he would "not be going to the Inauguration on January 20th", in context of the violence on January 6,

---

[26] *See generally* U.S. Department of Justice, *Final Report of the Special Counsel under 28 C.F.R. § 600.8* (Jan. 7, 2025), https://www.justice.gov/storage/Report-of-Special-Counsel-Smith-Volume-1-January-2025.pdf.
[27] Meta, *Oversight Board Upholds Former President Trump's Suspension, Finds Facebook Failed to Impose Proper Penalty*, Meta Oversight Board (May 5, 2021), https://www.oversightboard.com/news/226612455899839-oversight-board-upholds-former-president-trump-s-suspension-finds-facebook-failed-to-impose-proper-penalty/.
[28] The American Presidency Project, *Tweets of January 6, 2021,* (Jan. 06, 2021), https://www.presidency.ucsb.edu/documents/tweets-january-6-2021.
[29] Kari Paul, *YouTube extends ban on Trump amid concerns about further violence*, The Guardian (Jan. 26, 2021), https://www.theguardian.com/us-news/2021/jan/26/youtube-trump-ban-suspension.
[30] Instagram's Dangerous Individuals and Organizations Policy mirrors that of Facebook's Dangerous Individuals and Organization Policy.
[31] However, the Oversight Board noted that the "indefinite" restriction violates the principles of freedom of expression and it failed to transparently use the "newsworthiness allowance."

2021, were "highly likely to encourage and inspire people to replicate the criminal acts that took place at the U.S. Capitol on January 6, 2021."[32]

YouTube stated the reason for the ban was "in light of concerns about the ongoing potential for violence" and that YouTube had taken similar actions previously with other posts "involving safety concerns."[33] In a subsequent interview with TechPolicy.Press, then-CEO Susan Wojciciki noted that Trump's account strike was routine and that a possible reinstatement would be considered if the "risk of violence has decreased."[34]

ARTICLE 19 argued in its submission to the Meta Oversight Board that the indefinite suspension of any account raises serious freedom of expression concerns – and that it is even more problematic where political speech is concerned as under international freedom of expression standards, political speech is awarded a higher protection and strong reasons must be given in order to restrict it. However, ARTICLE 19 also argued that in "Trump-like cases, the suspension is likely to be proportionate as long as there is risk of imminent violence/imminent lawless action. ARTICLE 19 concludes that while suspensions will be justified whilst a significant risk of violence/imminent lawless action persists, users should have the possibility to have their account reinstated when those conditions are no longer met".

This is ultimately what happened - the decisions were reversed in the aftermath, well before the reelection of President Trump. Most importantly, the pattern of restriction highlights that the restrictions were not based on Trump's political views but rather on the application of general policies to his content.

A second example indicates that the approach of content moderation may fall across a political spectrum. Louis Farrakhan, the leader of the Nation of Islam (NOI), who has cultivated a large number of followers both in-person and on social media, faced deplatforming across several platforms due to findings of violation of the platform policies. In 2019, Facebook implemented a ban of six individuals they deemed "dangerous individuals", one of which included Louis Farrakhan.[35]

Facebook's decision was based on its "Policy Against Dangerous Individuals and Organizations."[36] Facebook's 2018 policy on "Dangerous Organizations and Individuals" stated that those who "proclaim a violent mission" are not allowed a presence on Facebook including acts such as "organized hate" and

---

[32] Robert Farley, *Trump's Dubious Claim About 'Hidden' Tweets Exonerating Him for Jan. 6 Capitol Attack* FactCheck.org (Feb. 17, 2023), https://www.factcheck.org/2023/02/trumps-dubious-claim-about-hidden-tweets-exonerating-him-for-jan-6-capitol-attack/.

[33] Jaclyn Diaz, *YouTube Joins Twitter, Facebook In Taking Down Trump's Account After Capitol Siege*, NPR (Jan. 13, 2021), https://www.npr.org/sections/insurrection-at-the-capitol/2021/01/13/956317191/youtube-joins-twitter-facebook-in-taking-down-trumps-account-after-capitol-siege.

[34] Justin Hendrix, *Youtube CEO says Donald Trump suspension isn't permanent because of "grace period"*, Tech Policy Press (March 04, 2021), https://www.techpolicy.press/youtube-ceo-says-donald-trump-suspension-isnt-permanent-because-of-grace-period/.

[35] Matthew S. Schwartz, *Facebook Bans Alex Jones, Louis Farrakhan And Other 'Dangerous' Individuals*, NPR, (May 3, 2019), https://www.npr.org/2019/05/03/719897599/facebook-bans-alex-jones-louis-farrakhan-and-other-dangerous-individuals; Casey Newton, *Facebook bans Alex Jones and Laura Loomer for violating its policies* [hereinafter *Facebook bans Alex Jones and Laura Loomer Article*], The Verge, (May 2, 2019), https://www.theverge.com/2019/5/2/18526964/facebook-ban-alex-jones-laura-loomer-milo-louis-farrakhan.

[36] This was the same policy later used in Meta's ban of Trump regarding the January 6, 2021 riots.

"organized violence".[37] Although it did not disclose all of the instances that led to an accounts removal, Facebook stated that it was a mix between the individual's behavior both on and offline. General factors included engaging in acts of hate or violence; calling for or carrying out acts of violence rooted in racial or ethnic prejudice; describing themselves as followers of a hateful ideology; or using hate speech or slurs in their profiles.[38] Farrakhan had previously been found to have engaged in hate speech and unfounded conspiracies towards a certain racial group and the LGBTQ+ community along with incitements of violence offline.[39]

On October 2, 2020 the NOI channel was taken down due to violating YouTube's hate speech policy, specifically for "advancing a claim that members of a group are part of an evil conspiracy theory."[40] Farrakhan and NOI's removal was based on YouTube "strict policies prohibiting hate speech on YouTube" and its practice of removing "channel[s] that repeatedly or egregiously violates those policies."

As with the previous examples, the content moderation policies and practices that led to the restriction of Farrakhan and the Nation of Islam may not have been beyond reproach. Hate speech policies are often formulated too broadly and inconsistently applied. ARTICLE 19 has also repeatedly criticized Meta's DIO policy as it is not available to the public and often leads to over removal of protected speech.[41] Yet, it is also clear that international human rights law allows – and at times demands – restrictions of speech, in particular when it comes to inciting hate speech.

V. Content moderation needs more consistent human rights protection, not government incentives to take a political stance

Despite the prior advancements of content moderation and content policies that have trended towards greater adherence to human rights, recent shifts amongst social media companies have signaled a change for the worse, driven by political demands of the Trump administration.

On January 7, 2024, Facebook CEO Mark Zuckerberg gave a public statement affirming Meta's new stance regarding issues such as fact checking, algorithms, and freedom of expression.[42] As ARTICLE 19 immediately called out "[w]hile claiming to protect freedom of expression, the shift, and its timing, appear politically motivated, reflecting an attempt to cater to specific political interests rather than

---

[37] Transparency Center, Meta, *Dangerous Organizations and Individuals*, at "Dec 29, 2018", https://transparency.meta.com/policies/community-standards/dangerous-individuals-organizations/.
[38] Alex Jones and Laura Loomer Article.
[39] For example, CNN, *Nation of Islam leader Farrakhan delivers anti-Semitic speech* (Feb. 28, 2018), https://www.cnn.com/2018/02/28/politics/louis-farrakhan-speech; https://j0nathan-g.medium.com/louis-farrakhan-again-61e629706f4c.
[40] The Times of Israel, *YouTube removes Louis Farrakhan's Nation of Islam channel* (Oct. 7, 2020), https://www.timesofisrael.com/youtube-removes-louis-farrakhans-nation-of-islam-channel/.
[41] ARTICLE 19, *Iran: Meta must overhaul Persian-language content moderation on Instagram* (Jun. 9, 2022), https://www.article19.org/resources/iran-meta-persian-language-content-moderation-instagram/; *see also* Clearing the Fog of War, footnote 260.
[42] Meta, *More Speech and Fewer Mistakes* (Jan. 7, 2025), https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/.

genuinely enhancing this fundamental right…. It suggests an effort to appease a political faction that has in the past accused Meta of suppressing conservative viewpoints and calls into question the integrity of Meta's commitment to freedom of expression."[43] ARTICLE 19 also called out the specific targeting of the efforts by the European Union to regulate Big Tech, which gave rise to concerns that this represented a politically expedient way to undermine any attempts at accountability through tech regulation. EFF, one of the predominant free speech organizations in the United States, further criticized Meta's new hateful conduct policy – which was also changed shortly after Mark Zuckerberg's announcement – finding that it would allow dehumanizing statements to be made about certain vulnerable groups, with these changes revealing "that Meta seems less interested in freedom of expression as a principle and more focused on appeasing the incoming U.S. administration".[44]

For X (formerly Twitter), the changes are not as recent. Since it has been taken over by Elon Musk, despite the claims of combatting bots and decreasing hate speech, the result since the takeover has been a 50% increase in hate speech with no relevant decrease in the number of inauthentic accounts and activities.[45] Immediately after his takeover, Musk also fired Twitter's human rights team.[46] If the current trends are to continue – and are further encouraged and incentivized by the U.S. government - all they will achieve is to harm the human rights and freedom of expression of users of the platforms.

As human rights defenders, the authors of this comment have repeatedly criticized content moderation of social media for falling short of their human rights responsibilities. Since the current shift in policies are towards less content moderation and decreased transparency for content moderation choices, freedom of expression will be hampered as social media companies will not be evaluated against a consistent international standard. Worse, the lack of proper content moderation has, and will continue to, lead to and bolster real-world acts of violence.

Extra-legal pressure by state actors or the government to further specific political positions does not only undermine the many policies meant to consistently moderate content by human right standards, but it distracts from substantive discussions about improving content moderation processes when it falls short of meeting its human rights obligations.

---

[43] ARTICLE 19, Meta: Prioritise human rights, not politics (Jan. 7, 2025), https://www.article19.org/resources/meta-prioritise-human-rights-not-politics/.

[44] Electronic Frontier Foundation, Meta's New Content Policy Will Harm Vulnerable Users. If It Really Valued Free Speech, It Would Make These Changes (Jan. 9, 2025), at https://www.eff.org/deeplinks/2025/01/metas-new-content-policy-will-harm-vulnerable-users-if-it-really-valued-free.

[45] *See* Daniel Hickey, Daniel M.T. Fessler, Kristina Lerman, Keith Burghardt, *X under Musk's leadership: Substantial hate and no reduction in inauthentic activity* (February 12, 2025), at 15-16, https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0313293&type=printable.

[46] Independent, Elon Musk fires Twitter's human rights team as part of sweeping layoffs at platform (Nov. 4, 2022), https://www.independent.co.uk/tech/elon-musk-twitter-employees-layoffs-b2218097.html.