

Silenced online: How social media platforms undermine free expression and human rights

Written by: Laura-Annabell Ambrosch, Shari Lambert, and Zineb Ez Zhak

Revised by Kornelia Gabriela Dąbrowska

Social media platforms are central in molding contemporary discussions, yet their content moderation practices often undermine free expression and human rights. This blog examines how platforms such as Instagram and Facebook disproportionately limit pro-Palestinian content, which breaches Article 19 of the International Covenant on Civil and Political Rights (ICCPR). Factors like automated moderation systems, ambiguous platform regulations, and corporate influences lead to biased censorship that silences marginalized groups. These platforms must implement transparent, rights-focused moderation strategies to promote free expression. It is vital to adopt a balanced approach to ensure that social media remains a forum for various perspectives and activism.

Introduction

Recent disturbances in Amsterdam have sparked public discourse to remove Cestmocro, a popular Dutch Instagram news outlet with over 1.1 million followers, due to allegations of antisemitic content.¹ Legal experts view such a prohibition as a significant infringement on the right

to freedom of expression, while critics argue the platform perpetuates harmful narratives and fails to monitor provocative comments adequately.² This localized debate reflects a broader international issue: the systematic suppression of pro-Palestinian content on social media platforms. Human Rights Watch reports global censorship of Palestine-related content on Facebook and Instagram, disproportionately targeting nonviolent posts highlighting human rights violations and expressing support for Palestinians.³

Between October and November 2023, out of 1,050 cases of censorship reviewed by Human Rights Watch, 1,049 involved the unjust suppression of peaceful pro-Palestinian content, while only one involved pro-Israel content. Such practices, which silence marginalized voices and threaten universal rights to free expression and access to information, are exemplified by the scrutiny faced by Cestmocro, a platform known for its

¹ 'Verbieden van Instagramaccount Cestmocro Is "Enorm Vergaande Maatregel"' (@NOS14 November 2024) <<https://nos.nl/artikel/2544534-verbieden-van-instagramaccount-cestmocro-is-enorm-vergaande-maatregel>> accessed 15 February 2025.

² Phippen Sardjoe, 'Actueel, Snel – En Pro-Palestijns. "Cestmocro: De Banksy van de Media, Niemand Weet Wie Het Is"' (NRC15 November 2024) <<https://www.nrc.nl/nieuws/2024/11/16/actueel-snel-en-pro-palestijns-cestmocro-de-banksy-van-de-media-niemand-weet-wie-het-is-a4873380>> accessed 15 February 2025.

³ Human Rights Watch, 'Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook' [2023] Human Rights Watch <<https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and>>.

coverage of Palestine. This controversy underscores the broader challenge confronting international platforms: moderating content in alignment with human rights.

We argue that social media platforms' content moderation policies and practices, particularly in the context of pro-Palestinian content, are deeply flawed and contribute to the silencing of opposing voices.

Freedom of Expression Under International Law

Article 19 of the International Covenant on Civil and Political Rights (ICCPR) ⁴ guarantees individuals the right to freely express opinions and acquire, seek, and share information, irrespective of national boundaries, extending to online expression.⁵

However, this right is not absolute.⁶ Under Paragraph 3, speech may be restricted only if it is necessary, proportionate, and legally justified to protect the rights of others, public morals, health, order, or national security. These standards are violated through the systematic censorship of non-violent pro-Palestinian content on social media platforms. Social media, as a critical space for advocacy and activism, is undermined when content policies infringe on the right to free expression. Such censorship significantly hinders efforts to raise awareness on Palestinian human

rights issues. This problem is worsened by biased automated moderation systems, which produce discriminatory outcomes, lack transparency, and offer no meaningful recourse.

Platforms Corporate Social Responsibility and Social Media's Impact:

Social media companies have also attempted to establish “platform laws” to govern content moderation. However, these laws are vague and ambiguous, failing to meet the precision required by Article 19(3) of the ICCPR to justify restrictions on expression.⁷ Even when international human rights standards and platform laws align, social media corporations continue unsuccessfully implementing their regulations.

When platforms fail to adopt a rights-based approach to content moderation, they risk becoming censors rather than facilitators of free expression. They must balance profit-driven motives with the responsibility of upholding human rights, a challenge exemplified by Meta, which owns widely used platforms. Its corporate structure, accountable to shareholders and motivated by profitability, compromises its ability to enforce freedom of expression impartially.⁸

Instead, moderation policies are often guided by market interests

⁴ International Covenant on Civil and Political Rights 1966.

⁵ ‘Human Rights Committee 102nd Session General Comment No. 34’ (2011) <[https://documents.un.org/doc/undoc/gen/g11/453/31/pdf/g1145331.pdf?>](https://documents.un.org/doc/undoc/gen/g11/453/31/pdf/g1145331.pdf?).

⁶ N4.

⁷ Aya Dardari and others, Social Media, Content Moderation and International Human Rights Law: The Example of the Nagorno-Karabakh/Artsakh Conflict (The Promise Institute For Human Rights 2022) <<https://promiseinstitute.law.ucla.edu/wp-content/uploads/2022/05/Social-Media-Content-Moderation-and-Internationals-Human-Rights-Law.pdf>>.

⁸ Andreas Kulick, ‘Meta’s Oversight Board and beyond – Corporations as Interpreters and Adjudicators of International Human Rights Norms’ (ssrn.com 22 September 2022) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4226521>.

rather than a genuine commitment to human rights. The removal or downranking of pro-Palestinian posts during geopolitical tensions highlights how moderation algorithms and governing policies align with powerful political or corporate interests. This approach undermines the equal application of rules designed to safeguard free speech. Despite the creation of the Meta Oversight Board to review controversial moderation decisions, concerns remain. Its reliance on Meta for funding raises doubts about its ability to tackle systemic biases, leaving moderation decisions unclear and unaccountable.⁹

For instance, the justification for removing pro-Palestinian content often cites violations of ‘community guidelines’, yet these justifications are rarely explained, and no remedies are provided.¹⁰ By selectively silencing pro-Palestinian content, Meta violates the principle of non-discrimination as outlined in the ICCPR while demonstrating how financial interests often outweigh the balance between market concerns and freedom of speech.¹¹

This not only reduces trust in these platforms but also creates an unfair digital space where marginalized voices are silenced, and diverse opinions are less

likely to be heard.¹²

Case of Delfi AS v. Estonia

In *Delfi AS v. Estonia*, the European Court of Human Rights (ECHR) established that online intermediaries may be held liable for the content published on their platforms, despite the use of moderators and takedown measures.¹³ The judgment further held that requiring platforms to moderate the content posted on their sites does not constitute a violation of Article 10, the right to freedom of expression. While the judgment aimed to establish liability for “hate speech and direct threats to the physical integrity of individuals”, the precedent may incite social media platforms to over-moderate incendiary content to avoid adverse legal consequences.¹⁴

The pro-Palestine viewpoint is often conflated with antisemitism, an allegation which, when combined with the fear of legal repercussions, can lead to the disproportionate moderation of pro-Palestine content.¹⁵ The ECHR holds that the Internet provides an unprecedented platform for exercising freedom of expression. However, this freedom is put at risk when social media

¹² Barrie Sander, ‘Freedom of Expression in the Age of Online Platforms: Operationalising a Human Rights-Based Approach to Content Moderation’ (2019) 43 SSRN Electronic Journal.

¹³ Case of Delfi AS v Estonia [2015] Coeint (European Court of Human Rights).

¹⁴ Guest Blogger, ‘Delfi as v. Estonia: Grand Chamber Confirms Liability of Online News Portal for Offensive Comments Posted by Its Readers’ (Strasbourg Observers 18 June 2015)

<<https://strasbourgobservers.com/2015/06/18/delfi-as-v-estonia-grand-chamber-confirms-liability-of-online-news-portal-for-offensive-comments-posted-by-its-readers/>>.

¹⁵ Center for Security, Race and Rights, ‘Presumptively Antisemitic: Islamophobic Tropes in the Palestine–Israel Discourse’ (Rutgers University Law School 2023)

<<https://csrr.rutgers.edu/wp-content/uploads/2023/11/csrr-presumptively-antisemitic-report.pdf>>.

⁹ Securing Ongoing Funding for the Oversight Board | Oversight Board’ (www.oversightboard.com 22 July 2022) <<https://www.oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board/>>.

¹⁰ N3.

¹¹ Oliva Thiago Dias, ‘Content Moderation

Technologies: Applying Human Rights Standards to Protect Freedom of Expression’ (2020) Volume 20 Human Rights Law Review 607.

platforms prioritize their legal interests over protecting their users' right to freedom of expression.

Conclusion

Social Media platforms moderate content at their discretion, following rules known only to them. Platforms like Cestmocro have become indispensable tools for communication and sharing news. It is high time for more

transparent oversight as to how these pages and their content are moderated. International content moderation disputes must be resolved, and social media platforms must adhere to international human rights standards, particularly concerning the freedom of expression. Only then can we ensure a fair and inclusive digital space.

Bibliography

Blogger G, 'Delfi as v. Estonia: Grand Chamber Confirms Liability of Online News Portal for Offensive Comments Posted by Its Readers' (*Strasbourg Observers* 18 June 2015) <<https://strasbourgobservers.com/2015/06/18/delfi-as-v-estonia-grand-chamber-confirms-liability-of-online-news-portal-for-offensive-comments-posted-by-its-readers/>>

Center for Security, Race and Rights, 'Presumptively Antisemitic: Islamophobic Tropes in the Palestine-Israel Discourse' (Rutgers University Law School 2023) <<https://csrr.rutgers.edu/wp-content/uploads/2023/11/csrr-presumptively-antisemitic-report.pdf>>

Dardari A and others, *Social Media, Content Moderation and International Human Rights Law: The Example of the Nagorno-Karabakh/Artsakh Conflict* (The Promise Institute For Human Rights 2022) <<https://promiseinstitute.law.ucla.edu/wp-content/uploads/2022/05/Social-Media-Content-Moderation-and-Internationals-Human-Rights-Law.pdf>>

'Human Rights Committee 102nd Session General Comment No. 34' (2011) <<https://documents.un.org/doc/undoc/gen/g11/453/31/pdf/g1145331.pdf>>

Human Rights Watch, 'Meta's Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook' [2023] Human Rights Watch <<https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and>>

Kulick A, 'Meta's Oversight Board and beyond – Corporations as Interpreters and Adjudicators of International Human Rights Norms' (*ssrn.com* 22 September 2022) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4226521>

Pippen Sardjoe, 'Actueel, Snel – En Pro-Palestijns. "Cestmocro: De Banksy van de Media, Niemand

Weet Wie Het Is ”(NRC15 November 2024) <<https://www.nrc.nl/nieuws/2024/11/16/actueel-snel-en-pro-palestijns-cestmocro-de-banksy-van-de-media-niemand-weet-wie-het-is-a4873380>> accessed 15 February 2025

Sander B, ‘Freedom of Expression in the Age of Online Platforms: Operationalising a Human Rights-Based Approach to Content Moderation ’(2019) 43 SSRN Electronic Journal

‘Securing Ongoing Funding for the Oversight Board | Oversight Board ’(www.oversightboard.com22 July 2022) <<https://www.oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board/>>

Thiago Dias O, ‘Content Moderation Technologies: Applying Human Rights Standards to Protect Freedom of Expression ’(2020) Volume 20 Human Rights Law Review 607

‘Verbieden van Instagramaccount Cestmocro Is “Enorm Vergaande Maatregel ”’(@NOS14 November 2024) <<https://nos.nl/artikel/2544534-verbieden-van-instagramaccount-cestmocro-is-enorm-vergaande-maatregel>> accessed 15 February 2025

Case of Delfi AS v Estonia [2015] Coeint (European Court of Human Rights)
International Covenant on Civil and Political Rights 1966