

THE PLATFORMICIDE OF PALESTINE

2021–2025



Content not recommended

This content may go against our [Community Standards](#).



Removed

This content goes against our [Dangerous Organizations and Individuals](#) policies.



Not visible to others

This content is now hidden.



Reach

0

Your content is not distributed.



7amleh - The Arab Centre for the Advancement of Social Media

The Platformicide of Palestine 2021-2025

A data-based analysis of Meta's policy application, moderation enforcement, and communication

June 2026

Principal Investigator: Dr Fabio Cristiano, Utrecht University, f.cristiano@uu.nl

Data collection: Ahmad Qadi and The Palestinian Observatory of Digital Rights Violations (7or)

Data analysis: Yousef Al-Sallal and Youssra Hamdan

Research assistants: Lui Prendergast and Antonio Fabio Romano

Graphic design: Qafzeh - Marketing, Digital, and Production

The editorial content of this publication is licensed under a Creative Commons NonCommercial 4.0 international license

To view a copy of the license, please visit the link below:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Contact us:

Email: info@7amleh.org

Website: www.7amleh.org

Find us on social media: 7amleh       

To cite this report:

Cristiano, F. (2026). *The Platformicide of Palestine 2021–2025: A data-based analysis of Meta's policy application, moderation enforcement, and communication*. 7amleh – The Arab Center for the Advancement of Social Media.

Table of Contents

o. Executive Summary	4
1. Introduction	10
1.1 Research Aim and Background	10
1.2 7or and 7amleh’s Monitoring Activity	11
1.3 Meta’s evolving approach to content moderation	11
1.4 Research design and Methodology	16
1.5 Previous Works and Contributions: Contextualising the Platformicide	18
1.6 Naming the Platformicide	19
1.7 Dataset Description and Coding	21
1.8 Limitations	24
2. Meta’s Policy Application	25
2.1 Dominant policies	25
2.2 Dangerous Organisations and Individuals (DOI)	26
2.2.1 Focus Issue: Is Palestine Newsworthy?	29
2.2.2 Focus Issue: the Term ‘Shaheed’- Are Palestinians Allowed to Mourn?	38
2.3 Hateful Conduct	43
2.3.1 Focus Issue: Zionism: A Problematic Conflation?	44
2.3.2 Focus Issue: Recommended content after October 7	46
3. Meta’s Moderation Enforcement	49
3.1 Dominant Enforcement	49
3.2 Suspension	51
3.2.1 Focus Issue: Suspension of Civil Society, Journalists, and Media Outlets	52
3.2.2 Focus Issue: Suspension of an Israeli activist	57
3.3 Restrictions	59
3.3.1 Focus Issue: Restrictions on Going Live during the Genocide	61
3.4 Content Takedown	64
3.4.1 Focus Issue: The Palestinian flag	65
3.5 Shadowbanning	68
3.5.1 Focus Issue: Shadowbanning activists	69
4. Meta’s Communication	72
4.1. (Not) Responding to 7amleh	72
4.1.1 Focus Issue: ‘Active and Unrestricted’: Targeting Influencers	74
4.1.2 Focus Issue: The US Campaign for Palestinian Rights	78
4.2 Restored with No Explanation	81
4.3 Meta’s Responses After October 7th	83
5. Conclusions	87
5.1 Summary of findings	87
5.2 Policy Recommendations	89
5.3 Future research	91

o. Executive Summary

This report analyses Meta’s moderation of Palestinian users and Palestine-related content on Facebook and Instagram between January 2021 and December 2025. Based on the analysis of 3520 cases submitted by users through 7amleh’s Palestinian Observatory for Digital Rights Violations (7or), the report examines three dimensions of Meta’s platform governance of Palestinian content: policy application, moderation enforcement mechanisms, and communication with users and 7amleh. The dataset is almost evenly split between the two platforms, with 1772 Facebook cases and 1748 Instagram cases. To explore Meta’s moderation approach during the genocide in Gaza, the analysis also captures moderation both before and after 7 October 2023, with 55% of cases occurring before that date and 45% after. Our social media analysis combines quantitative coding with qualitative assessment of screenshots, user reports, and Meta’s responses to appeals.¹ The central and overarching finding is that Meta’s moderation practices systematically silence Palestinian and pro-Palestinian users, producing what the report conceptualises as *the platformicide of Palestine*: the structural and intentional erasure of Palestinian digital presence, visibility, and participation across social media platforms. Meta’s systematic approach to erasing Palestine from its platforms goes beyond traditional content moderation. It also encompasses a wider and exceptional architecture of overbroad policy application, disproportionate enforcement, ad-hoc policy changes, algorithmic suppression, delayed restorations, and poor or absent communication.

Key findings:

1. *Meta’s policy application is ambiguous and lacks transparency.* Of the 3520 cases analysed, only 1144, or 32.5%, contained an identifiable policy that was allegedly violated, as deduced either from Meta explicitly providing this information in their enforcement communication to the user or from our qualitative analysis, in which it could reasonably be identified based on the content provided in the user’s submission. This underscores a clear lack of transparency in Meta’s policy application process.

2. *Meta relies heavily on the Dangerous Organisations and Individuals (DOI) policy when moderating Palestinian content.* Among cases with an identifiable policy, DOI accounts for 654 cases (57.2%), making it by far the most frequently applied policy in the dataset. This indicates a heavy reliance on a policy designed for terrorism, organised violence,

¹ The report’s findings should be read in light of the dataset’s user-reported nature. The cases analysed in this report are not a random sample of all Meta moderation decisions concerning Palestine-related content. They are cases submitted by users who believed that their content, accounts, or visibility had been wrongfully affected, and the analysis is further shaped by the limited information Meta provides about policy application and enforcement decisions. These limitations do not undermine the report’s contribution. Rather, they define its scope: the report documents patterns among reported and escalated cases of moderation affecting Palestinian and Palestine-related content, and shows how these cases reveal recurring problems in Meta’s policy application, enforcement practices, and communication with affected users and 7amleh.

and serious offline harm in the moderation of a broad range of Palestinian content and users, including journalists, media outlets, NGOs, activists, and ordinary users. The application of DOI to such diverse Palestinian content and users suggests that all forms of Palestinian expression are regularly treated through a security framework. Rather than consistently distinguishing between support for DOI-designated organisations and other content, Meta moderates Palestinian content through a security lens that criminalises Palestinian society as a whole.

3. Meta fails to adequately recognise the newsworthiness of Palestinian reporting.

Despite Meta's own stated exceptions for news reporting, the report finds that Palestinian journalists and media outlets are repeatedly penalised for reporting on events involving DOI-designated actors. Of journalists' cases with an identifiable policy, 128 out of 180, or 71.1%, were moderated under DOI. Among media outlets, 102 out of 172 identifiable cases, or 59.3%, were moderated under DOI, with 28 of these cases, or 27.5%, later restored. The qualitative analysis shows that these cases involved neutral news reporting, quotation of public statements, or coverage of unfolding events. Even when content or accounts are eventually restored, the delay reduces the value of time-sensitive reporting, advocacy, and documentation.

4. The term "shaheed" illustrates the linguistic and cultural blind spots of Meta's moderation.

The report identifies "shaheed" as a key example of Meta's failure to account for Arabic linguistic context and Palestinian mourning practices. The term appeared in 17 cases, and among the 14 cases with an identifiable policy, 12, or 85.7%, were moderated under DOI. In many cases, the term was used in mourning, remembrance, or reference to civilians killed by Israeli violence. Meta's treatment of the term as an indicator of dangerous content produced disproportionate enforcement, including content takedowns, suspensions, and restrictions. Following advocacy by 7amleh and others, Meta amended its approach in July 2024, after which cases involving "shaheed" decreased significantly, as shown in our analysis. This confirms both the scale of the earlier problem and the importance of sustained civil society monitoring and advocacy.

5. Meta's treatment of the term "Zionist" raises concerns about the conflation of political critique and legitimate protected identity.

A second major policy finding concerns Hateful Conduct and Meta's post-7 October treatment of the term "Zionist". The report shows that Meta's approach risks conflating criticism of Zionism as a political ideology with attacks on Jews or Israelis as protected groups. The analysis shows that this policy can result in the moderation of posts criticising Zionism, including posts by Jewish organisations or

Israeli users, while failing to address the broader problem of anti-Palestinian incitement and dehumanisation across both platforms.

6. Meta's algorithmic changes intensified the risk of Palestinian content suppression. After 7 October 2023, Meta lowered the threshold for reducing recommendations of potentially violating or borderline content. Given the documented wrongful moderation of Palestinian content, this expanded the space for false positives and algorithmic suppression. The later political-content settings introduced in 2025 further risk reducing the visibility of Palestine-related content, even where it does not violate community standards. The report therefore argues that Palestinian content is suppressed both through explicit moderation and recommendation systems, reduced distribution, and automatic political content filtering.

7. Suspension is the most widely used enforcement mechanism in the dataset. Across the entire dataset, suspensions account for 1187 cases, or 33.7%, followed by restrictions at 29.5% and content takedowns at 25%. Meta, therefore, did not mainly respond to Palestinian content by removing individual posts. It more often uses account-level penalties, especially suspensions and restrictions. While a takedown removes a single piece of content, a suspension can silence an entire account. For journalists, activists, media outlets, and civil society groups, this can interrupt reporting, advocacy, documentation, and communication with audiences. Because suspension removes access to an account and makes it invisible to others, wrongful suspension can have consequences far beyond the specific content that triggered enforcement.

8. Meta wrongfully removes content that displays or references the Palestinian flag. The report identifies the Palestinian flag as a focus issue in content takedown cases, showing that symbolic Palestinian expression can itself become vulnerable to moderation. This indicates that Meta's enforcement also extends to basic symbols of Palestinian identity and collective belonging.

9. Activists, journalists, and media outlets account for a large share of suspensions. The report finds that these groups account for 46.2% of all suspension cases. This shows that core public-interest actors, including those involved in reporting, advocacy, and documentation, are heavily affected by one of Meta's most severe enforcement mechanisms. The suspension of these actors has broader public consequences because it restricts both individual expression and access to information, documentation, and advocacy work.

10. Meta increasingly relies on less visible forms of enforcement, including shadowbanning

and reduced distribution. Our analysis found that cases of shadowbanning increased from 2 cases October 7 (0.1%) to 56 cases October 7 (3.5%). While the hard-to-detect nature of shadowbanning makes it difficult to determine the extent of Meta's increased reliance on this enforcement mechanism, this pattern is nevertheless significant.

11. Restrictions on going live disproportionately affect activists and journalists. Among 235 cases involving restrictions on going live, activists accounted for 67 cases (28.5%), while journalists accounted for 55 (23.4%). The restoration rate for these restrictions was 27.2%, suggesting that many were applied incorrectly or disproportionately. For journalists and activists reporting on the genocide in Gaza and other forms of Israeli violence, live broadcasting is crucial for reporting and mobilising at a time when it is needed the most.

12. Meta regularly fails to respond to 7amleh's appeals. Out of 2828 appeals submitted by 7amleh between 2021 and 2025, Meta provided no response in 1355 cases, or 48.4% of all submissions. This lack of responsiveness undermines Meta's stated commitments to transparency, stakeholder engagement, and accountability. It also leaves 7amleh and affected users without a clear explanation of why content or accounts were moderated, whether the decision was reviewed, or whether any corrective action was taken.

13. When Meta does respond, its explanations are often generic and uninformative. In 664 cases, or 23.5% of appeals, Meta upheld the original moderation decision. However, only 390 of these cases, or 13.9% of total appeals, included a specific policy reference. In the remaining 274 cases, or 9.8%, Meta upheld the decision without identifying the relevant policy. This means that many users were told that their content violated Meta's rules without being informed which rule was violated, which content triggered the decision, or why the enforcement was considered justified.

14. Meta often describes accounts as "active and unrestricted" despite continuing user-reported restrictions. The report identifies 273 of such cases, or 9.7% of appeals, in which Meta stated that accounts were active and unrestricted. The qualitative analysis shows that users often continued to report restrictions despite Meta's claim that the account was functioning normally. This suggests that Meta's communication does not always reflect users' experience of enforcement. It also suggests that some restrictions operate through less visible forms of moderation, including reduced distribution, shadowbanning, or unresolved account-level limitations. Our analysis reveals that this discrepancy is especially significant for influencers and high-reach accounts, where hidden or unresolved restrictions can substantially reduce visibility and engagement.

15. Restorations often occur without explanation. Following 7amleh's escalation, Meta reversed moderation enforcement in 926 cases, representing 26.3% of the dataset. However, in 642 of these restored cases, or 69.5%, Meta provided vague or unclear communication about why the decision was reversed. Only 284 restored cases, or 30.5%, included an explanation. This means that even when Meta restores content or accounts, users are often left without any clear understanding of why the original decision was wrong. Because most restored cases lacked a clear explanation, users and 7amleh could not determine the reasons for the initial moderation error. Restoration without explanation, therefore, limits transparency and prevents meaningful learning and policy change.

16. Meta's communication after 7 October shows a sharp increase in the number of recognised moderation errors. Before 7 October, Meta recognised 14 cases, or 0.8%, as incorrectly removed. After 7 October, this increased to 248 cases, or 23.0%. Incorrectly restricted accounts also increased from 3 cases, or 0.2%, before 7 October to 49 cases, or 4.5%, after 7 October. This indicates that Meta's moderation became less accurate and more error-prone during the genocide, a period in which Palestinian content was most urgent and politically significant.

17. Meta reversed more moderation decisions after 7 October. The report finds that the number of restored cases increased significantly after 7 October. Before 7 October, only 17 cases, or 2.2%, were restored; after 7 October, 330 cases, or 44.2%, were restored. This increase suggests that a substantial number of post-7 October moderation decisions were later found to be incorrect or excessive. It also reinforces the finding that Meta's intensified algorithmic moderation during this period produced a high number of wrongful enforcement decisions.

Taken together, the findings of this report point to a structural accountability problem. Meta's treatment of Palestinian and Palestine-related content relies on broad policy categories, poor automated enforcement, and insufficient communication. These are particularly concerning in the context of genocide, military occupation, and the severe constrictions already placed on Palestinian journalism and reporting. The report, therefore, argues that Meta's remedy cannot be limited to restoring individual posts or accounts after escalation. Meta must address the systems and structural elements that produce the very conditions for the platformicide of Palestine to take place. Based on these findings, the report recommends that Meta conduct heightened human rights due diligence of its content moderation, recommendation, escalation, and appeal systems in relation to Israel and Palestine. This process should be context-sensitive and informed by Palestinian civil

society organisations and affected users. Meta should also reform the application of the DOI policy to ensure that journalism, political analysis, and human rights reporting are not treated as support for designated actors. It should strengthen safeguards for public-interest and newsworthy content, improve Arabic-language and Palestine-specific moderation capacity, provide clear explanations for enforcement and restoration decisions, and make less visible forms of enforcement, including reduced distribution and shadowbanning, subject to notification and appeal. Future work should situate the report's findings within broader accountability frameworks. This includes the UN Guiding Principles on Business and Human Rights, international freedom of expression standards, and conflict-sensitive human rights due diligence. The report's findings also point to the need for further research into how recommendation systems, political-content settings, demotion practices, and automated thresholds shape the visibility of Palestinian content.

1. Introduction

1.1. Research Aim and Background

Extensive previous research by 7amleh² has documented Meta's biased moderation practices concerning Palestinian users and Palestine-related content.³ This data-driven research delves deeper and systematically into these practices, focusing specifically on the period from January 2021 to December 2025. By analysing 3520 cases of Meta's moderation on Facebook and Instagram, it aims to provide a comprehensive understanding of Meta's exceptional and systematic censorship of Palestinian content and users. The analysis is structured around three main areas: 1) policy application, 2) moderation enforcement, and 3) responses and communication. Additionally, this report examines how these patterns and practices evolved in the aftermath of the events of October 7, offering a critical perspective on the shifts in Meta's approach to content moderation of Palestinian content during the genocide⁴ for each of the three researched areas.⁵

More specifically, this report examines the predominant Meta policies governing the enforcement of Palestinian content. It assesses the intensity of their application, including instances of over-enforcement and relevant changes in these policies over time. The analysis further identifies patterns in the types of accounts targeted to uncover broader trends in Meta's enforcement practices. The report also investigates Meta's responses to identify patterns in both restored content and accounts, as well as in cases where moderation decisions were upheld. In doing so, the report scrutinises the clarity and informativeness of Meta's communication, particularly regarding the rationale for removal and the specific policies invoked to justify enforcement decisions. Finally, the report adopts a chronological perspective, highlighting disparities before and after 7 October 2023 across policy application, moderation enforcement, and communication.

To develop and investigate the above-stated research aims, the report will be structured as follows. In Chapter 2, the report will delve into Meta's policy application of community standards. We will identify the dominant policies governing Palestinian content and track

2. Jalal Abukhater, "Meta's Role in Amplifying Harmful Content During Genocide in Gaza," 7amleh, September

2025, <https://7amleh.org/post/meta-s-role-during-genocide-en>. See also, Meta Let Palestine Speak! www.meta.7amleh.org

3. For brevity, this report refers to 'Palestinian content' as both content posted by Palestinian users and content related to Palestine.

4. The term "genocide" is used in this report as a legal and analytical category grounded in the 1948 Genocide Convention. Its use is informed by the ICJ's January 2024 provisional measures order in South Africa v. Israel, which found the rights of Palestinians in Gaza under the Convention to be plausible; the ICJ Advisory Opinion of July 2024 on the unlawfulness of Israel's continued presence in the Occupied Palestinian Territory; and the UN Special Committee report A/80365/ on Israeli practices affecting Palestinian human rights. See Genocide Convention, Art. II; ICJ, Order of 26 January 2024; ICJ, Advisory Opinion of 19 July 2024; UNGA, A/80365/.

5. The analysis of Meta's approach after October 7th is included into each analytical chapter, thus specifically addressing changes in policy application, moderation enforcement, and communication.

their evolution before and after October 7. In Chapter 3, the report will focus on the specific enforcement mechanisms used to moderate the analysed cases. In Chapter 4, we will discuss Meta's communication with moderated users and with 7amleh. For each analytical chapter, in addition to the quantitative focus that reveals the main patterns, we also examine distinct issues to highlight how Meta's moderation affects the Palestinian digital experience and presence on platforms, ultimately contributing to their erasure. In the final section, the report reflects on Meta's accountability and outlines several policy recommendations to improve the quality of moderation for Palestinian content and beyond.

1.2. 7or and 7amleh's Monitoring Activity

This research has been developed in partnership with the Palestinian Observatory for Digital Rights Violations (7or).⁶ The platform is an online reporting tool developed by 7amleh for collecting submissions from Palestinian users. It enables users to document and submit reports of digital rights violations on social media platforms, including content removals, account restrictions, and other moderation enforcement practices. 7or therefore functions as both a documentation and monitoring mechanism. By aggregating user-submitted cases, it aims to provide insight into trends of platform governance and enforcement practices, especially on Meta's platforms such as Facebook and Instagram. The collected data is used by 7amleh to support research and advocacy efforts focused on digital rights and freedom of expression. By doing so, 7or aims to contribute to broader discussions on transparency and accountability in content moderation by structuring individual user experiences into analyzable data, beyond Palestine as well. This report furthers this aim through a comprehensive analysis of data collected by 7or from 2021 to 2025, identifying the patterns and mechanics of Meta's moderation approach toward Palestinian content and users on Facebook and Instagram.

1.3. Meta's evolving approach to content moderation

Meta's content moderation operates through a layered system of policy rules, automated detection, human review, enforcement actions, and appeals. At the formal level, Meta's Community Standards⁷ define the categories of content that are prohibited, restricted, or allowed under specific exceptions. These include policies on Dangerous Organisations and Individuals (DOI), Hateful Conduct, Violence and Incitement, Bullying and Harassment, Violent and Graphic Content, Human Exploitation, Spam, and related integrity rules. Meta describes enforcement as a combination of automated systems and human review: machine-learning tools identify potentially violating content, route it for review, or apply

6. See 7amleh, "The Palestinian Observatory of Digital Rights Violations (7or)," <https://7or.7amleh.org/>.

7. See Meta, Community Standards, Meta Transparency Center, available at: <https://transparency.meta.com/policies/community-standards/>

enforcement directly when confidence levels are sufficiently high, while review teams assess cases that require contextual or linguistic judgement.⁸ Meta also envisions its platform governance as a multi-stakeholder and participatory endeavour. The Oversight Board and its Trusted Partner Programme are such mechanisms intended to improve accountability, review, and contextual sensitivity in content moderation.⁹ The Oversight Board provides an external appeals and review process for selected enforcement decisions on Meta's platforms, while the Trusted Partner Programme gives selected civil society and human rights organisations a channel to provide local context, flag urgent issues, and inform Meta's policy and enforcement processes.¹⁰

In principle, this moderation architecture is meant to balance safety, expression, and public interest through open communication with involved stakeholders.¹¹ Meta's Community Standards state that the platform allows discussion of public issues and recognises that some otherwise sensitive material may be shared for news reporting or public discussion on sensitive topics. These provisions are particularly important in the context of Palestinian social media presence, where users may need to refer to (mass) violence, armed actors, or graphic harm in order to document events, report abuses, or preserve evidence. Meta also has a specific "newsworthiness allowance", although its own transparency reporting suggests that this allowance is applied relatively rarely.¹² As will be shown in Chapter 2, the most consequential policy for Palestinian content is the Dangerous Organisations and Individuals (DOI) policy. DOI prohibits organisations and individuals that Meta designates as involved in violence or serious offline harm from having a presence on its platforms. It also restricts content that praises, supports, represents, or glorifies these designated entities and individuals.¹³ The problem is that, in the context of ongoing genocide, mass violence, and dispossession under a military occupation such as the one in Palestine, reporting, human rights documentation, mourning, or political commentary may reference designated actors without expressing support for them.

Meta constantly reviews and updates its policies and moderation practices. For the sake of this report, it is important to highlight the major policy changes that occurred after 7

8. Meta, "How Enforcement Technology Works," Transparency Center, 12 November 2024; Meta, "How Review Teams Work," Transparency Center, 12 November 2024.

9. See Meta, "Oversight Board," Meta Transparency Center, <https://transparency.meta.com/oversight/overview/>; Meta, "Bringing local context to our global standards," Meta Transparency Center, <https://transparency.meta.com/policies/improving/bringing-local-context/>.

10. 7amleh is part of Meta's Trusted Partner Programme.

11. A 2023 report published by Internews argues that Meta's Trusted Partner Programme does not function effectively as a safety mechanism: while it operates in practice as an emergency escalation channel, it is neither designed nor resourced as one, with trusted partners often waiting weeks or months for responses, including in cases involving imminent harm. See Internews, Safety at Stake: How to Save Meta's Trusted Partner Program, July 2023, available at: https://internews.org/wp-content/uploads/202307//SafetyAtStake_Internews.pdf.

12. Meta, "How We Enforce Our Policies," Transparency Center; Meta, "Integrity Reports, Second Quarter 2024," Transparency Center, 4 December 2024.

13. Meta, "Dangerous Organisations and Individuals," Community Standards, Transparency Center.

October 2023. On 13 October, Meta announced emergency measures in response to the beginning of the genocide in Gaza and the ensuing public debate on its platforms. The company stated that it had created a special operations centre, expanded Hebrew and Arabic review capacity, removed large volumes of content under the DOI policy, and taken steps to avoid recommending content that may violate or be borderline. Meta also said it was lowering the threshold at which its algorithmic technology would intervene to prevent the recommendation of such content.¹⁴ These ad-hoc emergency measures are significant because lowering automated thresholds may reduce the circulation of genuinely harmful content, but it also increases the risk of false positives and wrongful moderation. This is especially problematic in contexts where Arabic-language or Palestine-related content has already been shown to be vulnerable to over-enforcement. The Oversight Board itself later found that Meta had wrongly removed content about the genocide in Gaza that should have remained available because of its public-interest value. In expedited decisions concerning videos from the conflict, the Board concluded that the content had documentary and awareness-raising significance, even though it was disturbing.¹⁵

Another important policy development concerned the Arabic term “shaheed”. For years, Meta treated the term, often translated as “martyr”, as a strong indicator of praise when used in relation to designated individuals. In March 2024, the Oversight Board found that this approach was overbroad and failed to account for the term’s multiple meanings, including uses connected to mourning, remembrance, and non-violent expression. The Board noted that Meta itself had acknowledged that the policy might be over-enforcing large amounts of speech, particularly among Arabic-speaking users.¹⁶ In July 2024, Meta accepted key Oversight Board recommendations and agreed to end its blanket approach to the term, allowing “shaheed” unless the content otherwise praised violence or violated Meta’s rules.¹⁷ In July 2024, Meta also changed its approach to the term “Zionist” under its Hateful Conduct policy. Meta stated that it would remove some content targeting “Zionists” where the term is used as a proxy for Jews or Israelis and appears alongside dehumanising comparisons, calls for harm, or denial of existence.¹⁸ This specific policy change followed consultations with 145 stakeholders and was intended to address cases where the term “Zionist” functions as antisemitic proxy speech.¹⁹ The change is significant because it formalises a difficult distinction between criticism of Zionism as a political ideology and attacks on Jews or Israelis as protected groups. As our analysis will show, while the policy aims to address

14. Meta, “Meta’s Ongoing Efforts Regarding the Israel-Hamas War,” Meta Newsroom, 13 October 2023.

15. Meta Oversight Board, expedited decisions on Israel-Gaza conflict content, December 2023; The Guardian, “Meta Wrong to Remove Graphic Israel-Gaza Videos, Oversight Board Says,” 19 December 2023.

16. Meta Oversight Board, “Referring to Designated Dangerous Individuals as ‘Shaheed,’” Policy Advisory Opinion, 26 March 2024.

17. Meta Oversight Board, “Meta Accepts Key Oversight Board Recommendations to End Blanket Ban on ‘Shaheed,’” 2 July 2024; Reuters, “Meta to End Ban on the Word ‘Shaheed’ on Oversight Board’s Recommendations,” 2 July 2024.

18. Meta, “Update from the Policy Forum on Our Approach to ‘Zionist’ as a Proxy for Hate Speech,” Transparency Center, 9 July 2024.

19. Reuters, “Meta Expands Hate Speech Policy to Remove More Posts Targeting ‘Zionists,’” 9 July 2024.

antisemitism, it also raises concerns about whether legitimate political critique of Zionism may be mistakenly treated as identity-based hate speech.

The final major policy shift in this period came in January 2025, when Meta announced a broad restructuring of its moderation system under the slogan “More Speech and Fewer Mistakes”. Meta said it would end its third-party fact-checking programme in the United States and replace it with a Community Notes model. It also announced that it would focus enforcement more narrowly on illegal and high-severity violations, reduce some automated demotions, lift restrictions on certain topics described as part of mainstream discourse, and take a more personalised approach to political content.²⁰ Meta also revised Hateful Conduct, including by loosening restrictions on certain forms of insulting or exclusionary speech in political and religious debate.²¹ This 2025 change marked a clear turn in Meta’s moderation philosophy, which more closely aligns with the so-called “free speech” turn adopted by platforms such as X and supported by the Trump administration. On the one hand, Meta presented it as a corrective to over-enforcement and excessive censorship. On the other hand, civil society organisations and civil rights groups warned that the rollback could weaken protections against hate speech, disinformation, and harassment.²² For an overview of the policy changes, see Table 1 below.

20. Meta, “More Speech and Fewer Mistakes,” Meta Newsroom, 7 January 2025.

21. See Kari Paul, “Revisions of ‘Hateful Conduct’: What Users Can Now Say on Meta Platforms,” The Guardian, 8 January 2025.

22. Among others, Amnesty International warned that Meta’s rollback of content moderation safeguards could increase the spread of hate speech, disinformation, and discriminatory content, particularly against already marginalised communities. See Amnesty International, “Meta’s new content policies risk fuelling violence and discrimination,” 12 February 2025, <https://www.amnesty.org/en/latest/news/2025/02/meta-new-policy-changes/>.

Table 1. An overview of relevant Meta policy changes

Name of Update	Date	Policy	Details
Meta's Ongoing Efforts Regarding the Israel-Hamas War	13 October 2023	Dangerous Organization and Individuals	Expansion of review capacity in Hebrew and Arabic; Lower threshold for moderation and increased content takedown of Palestinian content under DOI
PAO on the Treatment of the Word Shaheed	2 July 2024	Dangerous Organization and Individuals	Previously enacted a blanket ban on the term 'Shaheed' as an overbroad indicator of praise when used in relation to a DOI. Following recommendations from the oversight board, Meta acknowledges the overenforcement that may have resulted from this blanket approach and agrees to adopt a more nuanced reading of the term.
Update from the Policy Forum on Our Approach to 'Zionist' as a Proxy for Hate Speech	9 July 2024	Hateful Conduct	Meta will remove content targeting 'Zionists' where the term functions as an antisemitic proxy speech for 'Jews' or 'Israelis', in line with the protected characteristics under hate speech.
More Speech and Fewer Mistakes	7 January 2025	General, Hate Speech, Misinformation	General & broad restructuring of Meta's moderation system, including: removal of 3rd party fact checking in the US, replaced by a community notes model; a more narrow focus of moderation enforcement to 'high-severity' violations; a more personalised approach to political content. Also lifted restrictions on certain topics considered part of mainstream discourse, particularly on exclusionary or insulting speech in certain political and religious debates, under the Hateful Conduct policy.

The implications of these recent policy updates are ambivalent for Palestinian content. Earlier systems appear to have over-enforced Palestinian and Arabic-language speech, especially under DOI.²³ Yet a broader rollback of moderation may not necessarily protect Palestinian users if security-based enforcement remains aggressive while protections against anti-Palestinian hate or incitement are weakened. Across 2021–2025, our study will reveal how Meta’s moderation system appears marked by a persistent tension: the company claims to protect speech and public debate, but its automated systems, emergency thresholds, and policy categories have repeatedly struggled to protect Palestinian expression in a consistent, transparent, and context-sensitive way. When Meta’s recent “more speech” approach is considered alongside the intensified erasure of Palestinian content in recent years, it becomes clear that, now more than ever, Palestine represents an exception within Meta’s broader moderation philosophy. The result is a “more speech” framework that, quite problematically, expands freedoms for some users while leaving Palestinian expression subject to exceptional restriction. In other words, Meta’s recent policy updates create a further imbalance for Palestinians.

1.4. Research design and Methodology

To conduct a thorough analysis of the data collected by 7or for the period 2021-2025, this report employs a methodological approach that integrates quantitative and qualitative social media analysis techniques. This dual-method framework was chosen to provide a balance between systematic data analysis and the nuanced interpretative depth required to address the complexities of content moderation. Given that content moderation heavily depends on the interpretation of textual and visual content, this mixed-methods approach enables the report to engage with the data on multiple levels, ensuring a robust and well-rounded analysis. The quantitative analysis facilitated the structured organisation, categorisation, and statistical evaluation of the dataset, allowing for the identification of patterns, trends, and shifts (over time). Such chronological sensitivity was required to account for Meta’s policy changes over time and the consequent enforcement strategies that derived.

These statistical insights laid the foundation for a more focused qualitative inquiry of individual cases through the analysis of the screenshot content. In their submission to the 7or platform, users can attach a screenshot reporting their specific instance of moderation. In 1809 cases, or 51.4% of the dataset, screenshots were available depicting the enforcement

23. According to an independent review by the Business for Social Responsibility (BSR) commissioned by Meta. See BSR, “Human Rights Due Diligence of Meta’s Impacts in Israel and Palestine in May 2021 | Insights and Recommendations,” September 2022, <https://www.bsr.org/reports/BSR-Meta-Human-Rights-Israel-Palestine-English.pdf>

decision but did not include the specific content in question, while a further 620 cases, or 17.6%, included screenshots showing both the enforcement decision and the content that triggered moderation.²⁴ This allowed our research team to examine both Meta's enforcement actions and the context and communication through which moderation occurred. In turn, the qualitative analysis of the individual cases enabled a deeper exploration of the contextual and semantic layers embedded in the data, offering critical insights into the rationale (or lack thereof) and implications of Meta's practices. This study thus employs a multi-method approach to explore the selected research areas, structured around three key methodological phases: 1) data logging, sampling, and integration, 2) pattern tracing in logged data, and 3) interpretative content analysis.

The first phase of this research involved systematic data logging, sampling, and integration, which entailed collecting and consolidating data for each case in the dataset. This phase was critical because Meta provides little information to users and 7amleh about its practices, and our research team needed to infer or triangulate these by revisiting user profiles, particularly regarding policy application and restored content. After acquiring from 7or, our research team logged the data in a centralised database designed to standardise and streamline the integration of cases to identify patterns across them. The integration process also involved identifying thematic overlaps and merging compatible categories while preserving the integrity and context of the original data. The second phase centred on identifying patterns within the integrated dataset. Using a qualitative coding framework, data points were mapped to emerging themes and categories across the three research areas. Pattern tracing also involved iterative cycles of coding and statistical analysis to ensure the patterns were derived inductively from the data. The iterative nature of this methodological process played a key role in enhancing the reliability and depth of the findings. The final phase applied interpretative content analysis to delve deeper into the previously identified patterns. This process involved critically examining data and screenshots from individual cases to uncover underlying meanings, contextual factors, and implications of Meta's practices. This interpretative approach enabled the integration of statistical analysis of Meta's practices and patterns with selected cases that served as examples of broader, more systematic patterns.

Insights from this interpretive endeavour are primarily discussed in the 'focus issue' sections of each analytical chapter and were selected as the most prominent cases to be included in the final writing phase of the report.²⁵ This dynamic approach ensured that the study captured both the systemic dimensions of Meta's practices and the interpretative nuances of how such bias is justified and played out.

24. For further explanation of this distinction, see section 1.7.

25. The final report draft underwent a round of revisions by 7amleh and the research team, as well as language proof and a legal review. Generative AI tools were used in limited ways to support language editing, and restructuring of selected draft passages. The research team and 7amleh retain full responsibility for the analysis, interpretation of data, and verification of evidence.

1.5. Previous Works and Contributions: Contextualising the Platformicide

This research builds upon the extensive analytical groundwork laid by 7amleh in its efforts to critically examine Meta’s approach to moderating Palestinian content. Previous reports by 7amleh have identified and documented Meta’s censorship patterns, shedding light on the pervasive nature of these practices. In the context of the genocide in Gaza, previous work by 7amleh has already addressed Meta’s role in amplifying hate speech against Palestinians.²⁶ Similarly, monetisation schemes on Meta’s platforms have guaranteed economic and material support to Israeli users active in the illegal settlements in the West Bank.²⁷ This report draws on advances in these debates through a systematic analysis. It also builds on the independent human rights due diligence exercise conducted by BSR on Meta’s impacts in Israel and Palestine following the May 2021 escalation. Commissioned by Meta, the BSR assessment examined whether Meta’s policies and processes were applied without bias and offered recommendations to align Meta’s operations with its human rights commitments and responsibilities under the UNGPs. Its findings are significant because they show that concerns regarding Arabic-language over-enforcement, inadequate contextual moderation, and unequal protection from harmful content were not new after 7 October 2023. Our report extends these earlier assessments by showing how these risks persisted and intensified across a longer period, from 2021 to 2025, through patterns of policy application, moderation enforcement, and communication.

In December 2023, Human Rights Watch (HRW) published a report that corroborates many of the patterns identified in 7amleh’s previous analyses of Meta’s content moderation practices, and in this report. The HRW report highlights how Meta’s approach systematically targets and erases Palestinian voices and narratives, shedding light on the structural and intentional nature of these practices. Our study expands this research through a distinctly Palestinian-centred approach that draws on analysis of a large dataset. It also prioritises the lived experiences and specific contexts of Palestinian users subjected to Meta’s interventions.²⁸ Another distinctive contribution of this report lies in its dedicated exploration of the communication between Meta, users, and 7amleh - two dimensions that have not been previously examined in such depth. For the first time, this research delves into the interactions between the platform and 7amleh, offering a unique perspective on the mechanisms, processes, and dynamics of Meta’s engagement with a monitoring partner. By doing so, the report not only expands on the patterns of Meta’s content moderation

26. Abukhater, J. (2025, September). Meta’s role in amplifying harmful content during genocide in Gaza. 7amleh – The Arab Center for the Advancement of Social Media.

27. Qadi, A. (2026, April). “Monetizing Occupation: Meta’s Financial Enablement of Settlement Activity and Violent Rhetoric Against Palestinians,” 7amleh – The Arab Center for the Advancement of Social Media.

28. On this issue, please see Erased and suppressed: Palestinian testimonies of Meta’s censorship: Testimonies from Palestinian influencers, journalists and media outlets.

but also provides a critical understanding of the organisational and procedural context in which these patterns unfold. This focus marks a significant step forward in scrutinising the structural dimensions of Meta's exceptionalism on Palestinian content.

1.6 Naming the Platformicide

Meta's exceptionalism in relation to the moderation of Palestinian content is reflected in the title of this report. The concept which best captures the systematic silencing and erasure of Palestine from social media platforms: platformicide. The concept draws inspiration from and parallels the idea of spaciocide. The notion of spacio-cide was introduced by Palestinian sociologist Sari Hanafi to describe a distinct modality of Israeli colonial violence in Palestine.²⁹ For Hanafi, the Israeli colonial project is "spacio-cidal" because it targets land and the conditions of inhabitable space in order to make the "voluntary" transfer of Palestinians appear inevitable. In this formulation, violence is directed not only at people as such, but at the spatial foundations of Palestinian social, political, and everyday life: the land, infrastructures, homes, routes, and geographies through which collective existence is made possible. It helps describe the Palestinian condition before and beyond the current genocide in Gaza. The need for such concepts is reflected in the growing use of related terms to describe the different dimensions of destruction experienced by Palestinians. Scholars, lawyers, journalists, and activists have referred to educide, scholasticide, domicide, ecocide, and medicide, among others, while some have also proposed treating the Nakba as a legal concept. These concepts extend the somewhat limited definition of genocide but they indicate that the destruction of Palestinian life operates across multiple domains: land, housing, education, health, ecology, memory, and the infrastructures of social reproduction. It is within this broader conceptual field that platformicide should be understood.

Just as spaciocide dismantles the physical and symbolic infrastructures of a people, Meta's exceptional approach to Palestinian content represents more than an issue of biased content moderation. Similar to the idea of spaciocide, Meta's platformicide of Palestine signifies the systematic and further erasure and suppression of digital Palestinian spaces where already marginalised voices and narratives seek visibility and recognition. As this analysis reveals, this extends beyond Meta's acts of censorship and 'broken promises' to encompass the broader, structural exclusion of the entire Palestinian community from meaningful participation, agency, and presence online.

It is thus important to situate this study within the broader historical context in which

29. Sari Hanafi, "Spacio-cide and Bio-politics: Israeli Colonial Project, from 1948 to the Wall," in Michael Sorkin, ed., *Against the Wall: Israel's Barrier to Peace* (New York: The New Press, 2005). The concept was later developed in Sari Hanafi, "Explaining Spacio-cide in the Palestinian Territory: Colonization, Separation and State of Exception," *Current Sociology* 61, no. 2 (2013): 190–205.

Palestinian digital rights have long been denied. Palestinian digital life has developed under conditions of territorial fragmentation and infrastructural dependency, where internet access and digital infrastructure are shaped by Israeli control over territory, borders, spectrum, mobility, and service provision.³⁰ In this sense, Palestinian digital rights cannot be understood only as a question of platform governance and content circulation/moderation. They are also embedded in a longer colonial history of infrastructural domination, where Palestinians have been denied full control over the material and regulatory conditions of their digital existence. This historical condition has become even more urgent in the context of the genocide in Gaza and the intensified assault on Palestinian journalism and media infrastructure. Palestinian journalists have carried the primary burden of documenting the genocide because Israel has continued to deny most international media outlets independent access to Gaza.³¹ At the same time, Palestinian journalists and media workers have been killed at an unprecedented rate. The Committee to Protect Journalists describes the genocide in Gaza as the deadliest conflict for journalists since it began collecting data in 1992, and its current casualty tracker records 263 Palestinian journalists and media workers killed in Gaza by Israel.³²

In this context, the visibility and presence of Palestinian stories, documentation, and voices online become even more politically and ethically significant. When journalists are killed, communications infrastructure is damaged, and foreign reporters are excluded, social media platforms become one of the few remaining spaces through which Palestinian testimony can circulate globally. The wrongful moderation of Palestinian content by Meta, therefore, does not occur in a neutral or ordinary information environment. It adds a further layer to an already existing system of silencing and erasure. The findings of this report show that Meta's moderation practices have systematically undermined Palestinian presence on social media platforms. Taken together, these findings suggest that Meta's moderation of Palestinian content functions as the digital counterpart to a wider politics of erasure. Its systematic removal and suppression of news, reporting, mourning, political expression, and documentation of Palestine; its use of enforcement tools to silence or restrict Palestinian content regardless of its message; and its failure to communicate clearly with affected users all contribute to the narrowing of Palestinian digital space. In this sense, Meta's moderation does not merely reflect existing inequalities. It reproduces and intensifies them, operating as a digital extension of the broader spaciocidal logics that structure Israeli policies towards Palestine.

30. Cristiano, F. (2019) "Internet Access as Human Right: A Dystopian Critique from the Occupied Palestinian Territory," in *Human Rights as Battlefields: Changing Practices and Contestations* (Palgrave Macmillan, 2019), 249–268.

31. Associated Press, "Media Organizations Call on Israel to Allow Foreign Reporters Independent Access to Gaza," 30 April 2026.

32. Committee to Protect Journalists, "Israel-Gaza War"; Committee to Protect Journalists, "Journalist Casualties in the Israel-Gaza War."

1.7. Dataset Description and Coding

The primary data corpus for this research consists of the data collected by 7amleh via the 7or platform. The reports submitted by users to 7amleh include detailed information about the user profiles, flagged content, and the corresponding intervention by Meta. Additionally, each data point captures details about the interactions between Meta and 7amleh, as indicated by 7amleh in the dataset. 7amleh provided our research team access to the 7or database, which includes cases spanning from 1st January 2021 to 18th December 2025. The dataset comprises 3,520 data strings, each referred to in this report as a case and temporally assigned a unique hashtag number (#). Of these, 1772 cases (50.3%) originated from Facebook users, while 1748 cases (49.7%) were from Instagram (Figure 1).

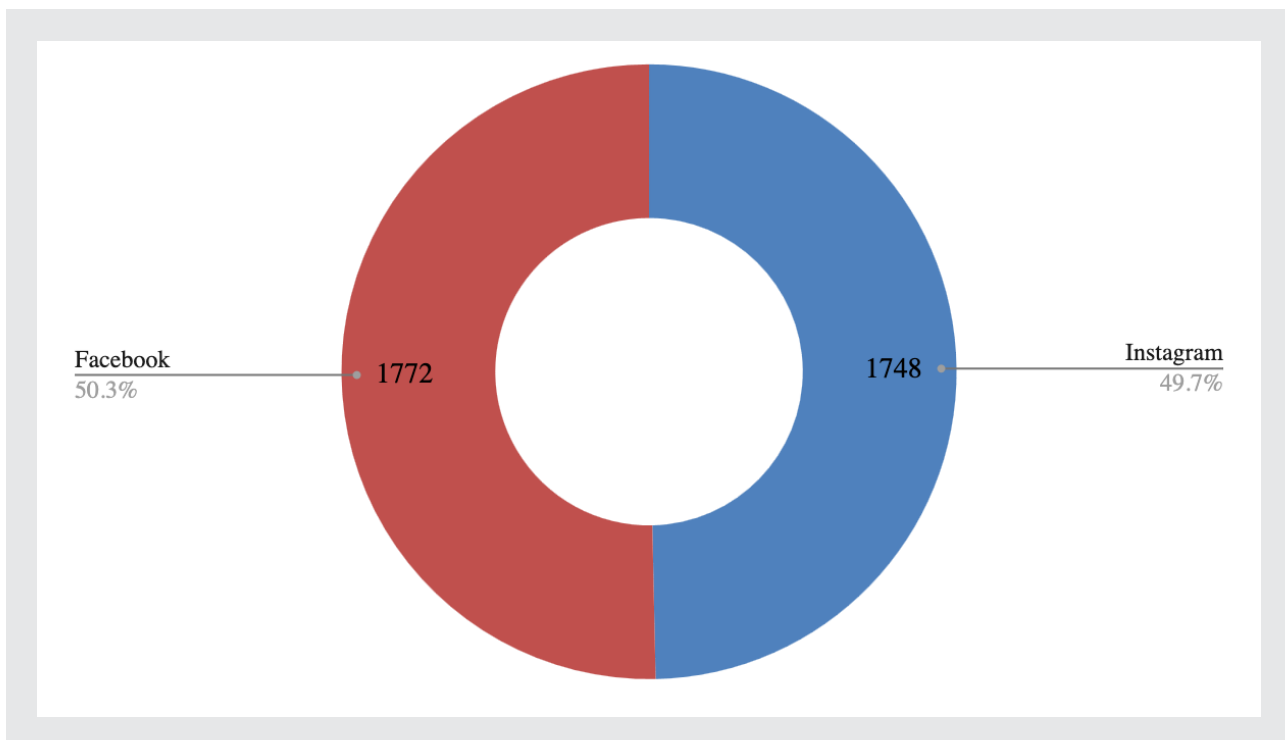


Figure 1. Distribution of Cases per Platform.

Additionally, the dataset was temporally coded according to whether cases occurred before or after 7 October 2023. Of the total cases, 1,938 (55.1%) occurred before 7 October, while 1,582 (44.9%) occurred after 7 October, across both platforms (see Figure 2).

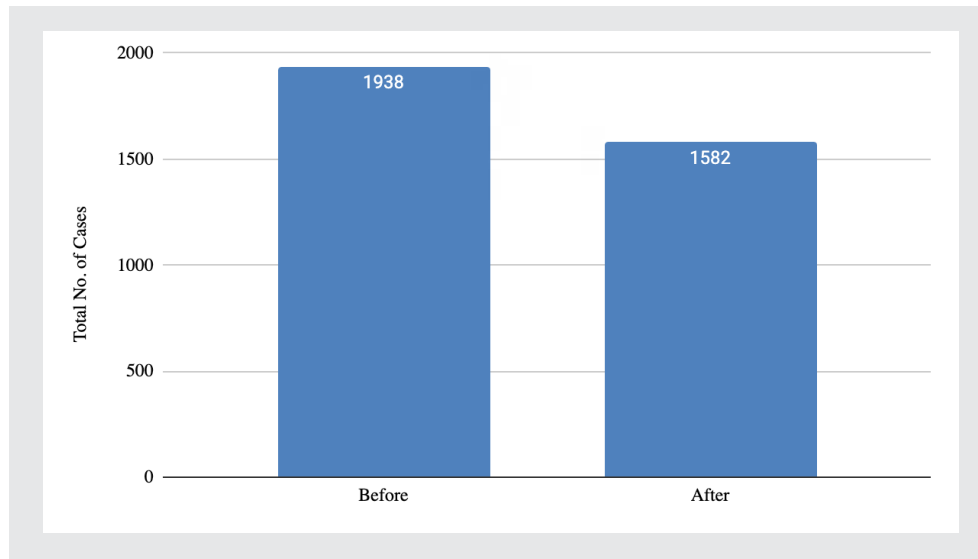


Figure 2. Distribution of Total Cases Before and After 7th October 2023.

For the qualitative analysis in this research, we examined whether screenshots accompanied each case, as these provided critical evidence of interactions between Meta and users. These were, however, not present in all cases. For data management and analytical purposes, the cases have been categorised into three groups based on the availability and nature of the screenshots, see Figure 3 below. The first group, labelled 'available', included 1809 cases (51.4%) with screenshots that provided evidence of the type of restriction or intervention imposed by Meta. However, they did not disclose the specific content the user posted. This group provided a general understanding of Meta's actions but lacked contextual depth on the fairness of Meta's intervention. The second group, 'available with content', comprised 620 cases (17.6%) in which screenshots included both the restriction details and the user's posted content. These cases were particularly valuable, as they enabled a deeper investigation into the specific nature and context of the posts, offering richer insights into how and why certain content or users were flagged or restricted by Meta. Finally, the third group, 'not available', consisted of 1091 cases (31.0%) without screenshots³³, making it impossible to directly assess the content or context of the user's post. However, within this group, 464 cases included documented interactions between 7amleh and Meta. These records offered alternative means of analysis, providing insight into Meta's interventions even without visual evidence. This three-part categorisation was integral to structuring the qualitative analysis and understanding the limitations and opportunities presented by the dataset.

33. Depicting interaction between the User and Meta.

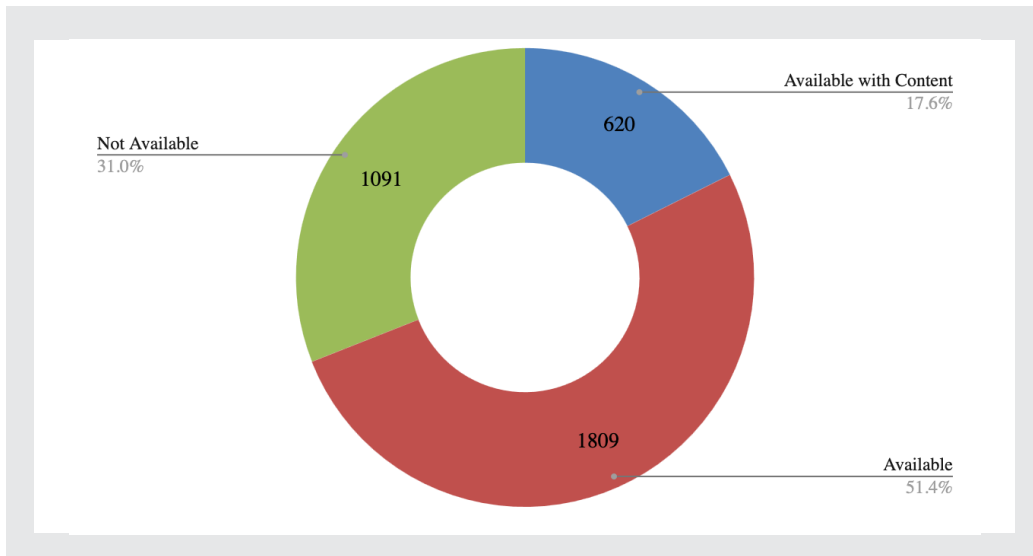


Figure 3. Distribution of Screenshot Availability Categories.

In terms of the classification of types of users as self-reported in their submissions to 7or, not including those self-labels as Others³⁴, we see a large proportion represented by activists (714, 20.3%), followed by Journalists (475 cases, 13.5%) and Media Outlets (465 cases, 13.2%). Importantly, these figures refer to the 2632 cases in which the user provided a classification in their submission (See Figure 4 for more details).

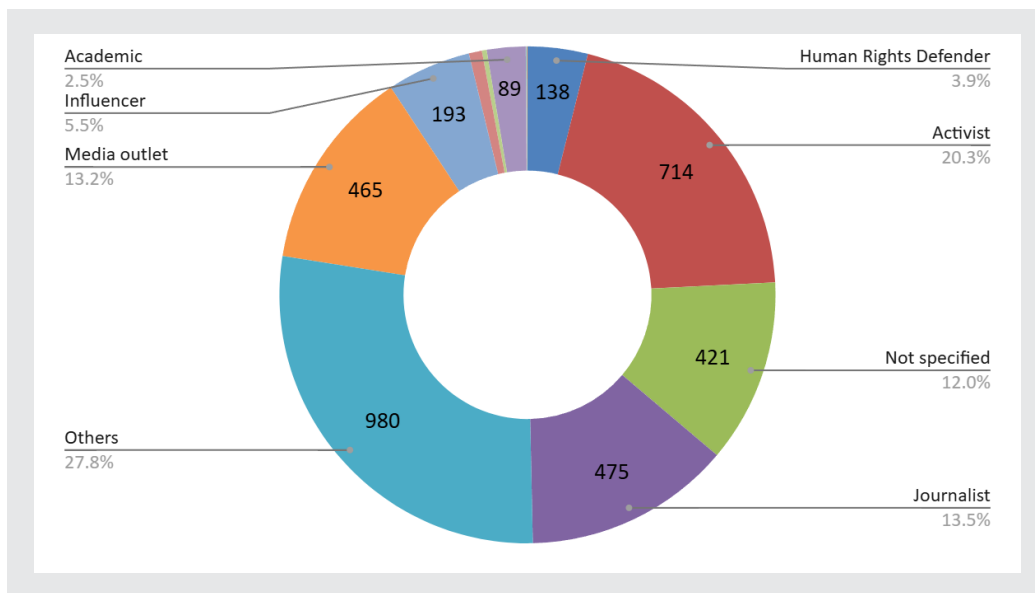


Figure 4. Distribution of Classifications of Users Across all Cases with a Distinguished Classification.

34. The category "Others" encompasses users who, when submitting their case to 7or, did not identify under one of the specific classifications provided.

1.8. Limitations

Stemming from the user-report-based nature in which the 7or's database was constructed, and as a direct consequence of the lack of transparency in moderation and policy application on Meta's end, two key limitations to this research arise. Firstly, the dataset is not based on a neutral or randomised sample representative of users across Meta platforms. Palestinian or Pro-Palestinian voices self-reported cases of censorship or wrongful moderation through the 7or online platform, based on their subjective judgment. Therefore, the database is constructed from cases in which the user believes they have been wrongfully moderated by Meta. Additionally, this submission usually occurs at the point of moderation, when content has already been removed or the account has been suspended. This limits the contextual elements of the previous violations that preceded many of the cases. This sampling method does not, however, invalidate the report's findings that Meta's bias in moderation concerns Palestine-related content. Rather, this report speaks to the frequency of wrongful moderation of Palestine-related content as reported by users, not to the likelihood of wrongful moderation of Palestinian content relative to other content. In practical terms, the latter is impossible to assess given the closed-door approach to Meta's policy and moderation.

Likewise, this lack of transparency in Meta's policy application, moderation enforcement, and communication hinders the assessment of the fairness of moderation - a second key limitation. As Meta's moderation team's policy applications and communications are often insufficient or entirely lacking, our research team had to infer the reason for moderation from the user's context and Meta's official community standards at the time. We acknowledge that fairness is rarely an objective variable and that moderation errors are permissible and to be expected. However, the report reveals Meta's approach as one of systematic censorship of Palestinian content and users. As it will be shown in the analytical chapters that follow, such systematic censorship is clearly revealed in both quantitative and qualitative terms.

2. Meta's Policy Application

This chapter examines Meta's predominant policies governing the enforcement of Palestinian content and assesses their intensity of application, highlighting instances of over-enforcement. We begin with a quantitative analysis to identify Meta's main policies governing content moderation for our dataset. Building on these findings, our analysis zooms in on Meta's two most used policies for Palestinian content moderation: Dangerous Organisations and Individuals (DOI) and Hateful Conduct. Before analysing Meta's policy application present in the dataset, it is important to emphasise again that of the 3520 analysed cases, only 1144, or 32.5%, contained an identifiable policy or policies that were allegedly violated. Our research team deduced this information either from cases in which the moderation or subsequent communication from Meta explicitly stated the policy that was violated, or, in many other cases, through our qualitative analysis of the individual cases, which could reasonably infer which policy Meta was invoking based on the content provided in the user's submission to 7or. Hence, the majority of the cases in the dataset, 2376 or 67.5%, do not have an identifiable policy to be analysed. This underscores a clear lack of transparency throughout Meta's overarching policy application process.

2.1. Dominant policies

When analysing the policies applied in Meta's moderation of Palestinian content, the DOI policy plays a very significant role. According to Meta's Community Standards, the DOI policy generally screens content that supports or promotes violence, terrorism, or dangerous entities, preventing "organisations or individuals that proclaim a violent mission or are engaged in violence" from having a presence on their platform.³⁵ Meta claims to be particularly stringent on DOI filtering, as the presence of designated groups on the platform could lead to instances of offline harm.³⁶ Above all, this also implies that Meta defines who is considered to be a dangerous individual or organisation across its platforms.³⁷ As shown in Table 2, 654 cases (57.2% of the cases with an identifiable policy) were moderated under the DOI policy. When considering the sheer diversity of users who have submitted their reports through 7or – journalists, media outlets, NGOs, human rights activists, etc. – Meta's over-reliance on the DOI policy for moderating Palestinian content might seem to be indicative of a criminalisation of Palestinian society as a whole.

35. Meta. "Dangerous Organisations and Individuals." Community Standards. Effective March 19, 2026.

<https://transparency.meta.com/policies/community-standards/dangerous-individuals-organizations/>

36. Meta. "Dangerous Organisations and Individuals." Effective August 26, 2021

37. In crafting the DOI policy, Meta originally adopted the US government lists for Narcotic Organisations, Foreign Terrorist Organisations, and Specially Designated Global Terrorists.

Table 2. Distribution of policies¹ applied in Meta's moderation 2021 - 2025

Type of Policy Applied	Number of Entries	Proportion of Total Entries(%)
Dangerous Organisations and Individuals	654	57.2
Use of Multiple Policies ²	115	10.1
Hateful Conduct	78	6.8
Violence and Incitement	68	5.9
Violent and Graphic Content	61	5.3
Intellectual Property	22	1.9
Fraud, Scams, and Deceptive Practices	22	1.9
Others	124	10.9
Total	1144	100

¹ Only included those with an identifiable policy based on our coding metrics.

² Of the cases where multiple policies were identified, 45(39.1%) of them included DOI.

2.2. Dangerous Organisations and Individuals (DOI)

Before focusing on Meta's application of DOI and the related targeted categories, it is important to clarify some of the evolving policy changes that have underscored the platform's moderation of DOI cases. According to Meta's Community Standards, content related to DOI is subject to stringent scrutiny, particularly for entities involved in severe offline harm, such as terrorist organisations and hate groups. The policy also requires the removal of content that glorifies, supports, or represents these entities and individuals, including vague references if the user's intent is unclear. However, as mentioned briefly in the introduction, Meta does recognise the importance of allowing discussions about these entities within civic and political debates in its DOI policies:

“We recognise that users may share content that includes references to designated dangerous organisations and individuals in the context of social and political discourse. This includes content reporting on, neutrally discussing or condemning dangerous organisations and individuals or their activities.”³⁸

38. Meta. “Dangerous Organisations and Individuals.” Effective March 19, 2026.

As quoted above, Meta allows content that reports on, neutrally discusses, or condemns dangerous organisations and individuals, acknowledging the need for such discourse in the public domain. Meta also purports the changes regularly implemented to this policy as serving to improve the policy's language and bolster this news-based exception through clearer differentiation.³⁹ Before the beginning of the genocide, Meta had been implementing such changes to allow for human rights discussions surrounding the people and organisations labelled as DOI and introduced policy changes to allow for raising awareness and for news reporting.⁴⁰ In the DOI policy effective from August 2023, Meta included the following paragraph:

“News reporting includes information that is shared to raise awareness about local and global events in which designated dangerous organisations and individuals are involved. E.g. “Breaking news: Al-Shabab claimed responsibility for the attack in Somalia” E.g. “Timeline and expert analysis: How the shooting at the Buffalo Supermarket unfolded and what the perpetrator said in court.”⁴¹”

In addition, in the same policy update, Meta clarified:

“We are updating the Community Standards to reflect the first phase of implementation of the praise policy revamp, which entails allowing social and political discourse at scale.”⁴²”

This was done with a clear intention of facilitating discussions of human rights and/or reporting on DOIs, further complementing their purported specification and refinement of this policy. In their first update to the DOI policy following October 7, effective from December 2023, Meta reverted its approach and started granting increased moderation for less severe risks of offline harm, thus enlarging the scope of the DOI enforcement. By widening the range of actions sanctionable with moderation under DOI, Meta effectively rigidified their moderation policy language: in practice, their DOI policy updates during the genocide mean that more content will be more heavily moderated, which therefore increases scrutiny on accounts discussing conflict-related issues. It is interesting to account for these policy changes in our analysis of the DOI policy application to the cases we analysed. Our analysis reveals that DOI consistently remained the predominant policy for moderating Palestinian content both before and after October 7. Of all cases with an identifiable policy applied before October 7, DOI accounted for 439 cases, or 62.7%; after October 7, DOI accounted for 215 cases, or 48.4% of the total (see Table 3 and Figure 5 below).

39. Ibid.

40. See: Meta. “Dangerous Organisations and Individuals.” Effective December 23, 2021.

41. Meta. “Dangerous Organisations and Individuals.” Effective August 29, 2023

42. Ibid.

Table 3. Changes in distribution of policy application, Before & After Oct 7th of the cases with an identifiable policy

Type of Policy Applied	Before Oct. 7th		After Oct. 7th	
	Number of Entries	Proportion of Total Entries (%)	Number of Entries	Proportion of Total Entries(%)
Dangerous Organisations and Individuals	439	62.7	215	48.4
Hateful Conduct	66	9.4	12	2.7
Multiple Policies Applied	59	8.4	56	12.6
Violence and Incitement	47	6.7	21	4.7
Violent and Graphic Content	24	3.4	37	8.3
Other policies	65	9.3	103	23.2
Total	700	100	444	100

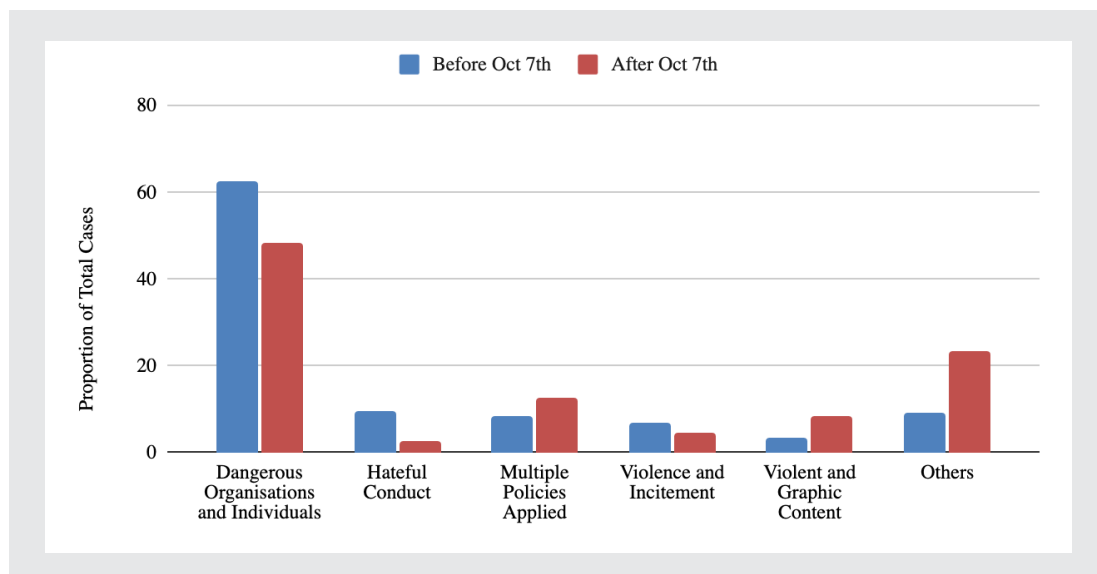


Figure 5. Distribution of Policies Applied Before and After October 7th.

In addition, the overall inaccuracy of DOI applications is straightforward throughout the period 2021-2025. When appealed, the platform restored 102 cases (15.6%) of content and/or accounts flagged under DOI. Moreover, moderation under DOI disproportionately affected certain user classifications (Figure 6): with 58% of the total including activists, journalists, and media outlets. As these actors are central to the functioning of a healthy and responsive public debate on sensitive issues, the staggeringly high proportion targeted under DOI signals worrying tendencies in Meta's moderation of Palestinian content.

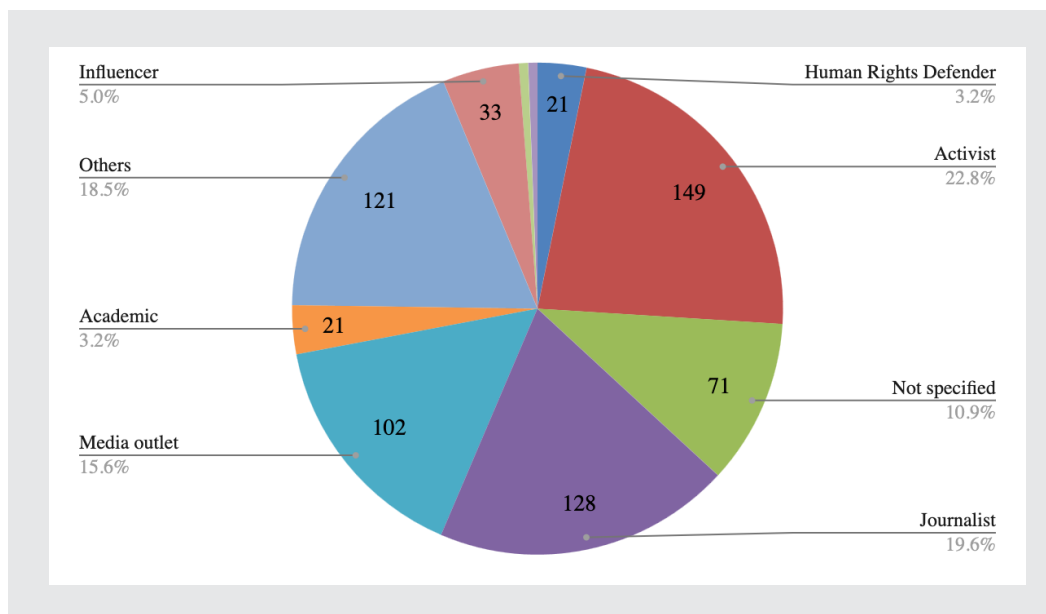


Figure 6. Distribution of Classifications Wrongfully Moderated under the DOI policy.

2.2.1. Focus Issue: *Is Palestine Newsworthy?*

In relation to news-reporting contexts, Meta introduced the so-called ‘newsworthiness allowance’ in October 2016 to enable open discussion of newsworthy topics, claiming to recognise the importance of allowing such discussions within civic and political debates.⁴³ Meta’s newsworthiness determination focuses on whether ‘content surfaces an imminent threat to public health or safety, or gives voice to perspectives currently being debated as part of a political process’. Despite these multi-layered provisions to protect journalistic work, our analysis finds that Meta’s DOI application on Palestinian content does not adequately differentiate between journalistic reporting from other user-generated content. This lack of distinction can lead to excessive enforcement of journalistic content, thus stifling freedom of expression and the public’s right to information.

Of all the journalists reporting their content moderation cases to 7or, these have overwhelmingly been moderated under DOI, with 128 (71.1%) of the 180 journalists’ cases with an identifiable policy falling under this category (see Figure 7). Moreover, Meta’s specific reasoning for the moderation of these cases is unclear, including which clauses were invoked, due to Meta’s lack of communication during enforcement. However, the proportion of journalists whose moderated content was later restored (17.3%) suggests the same pattern of wrongful moderation and excessive reliance on DOI discussed earlier in this section. Our qualitative analysis indicates that most of the content of these cases concerned the journalist posting non-opinionated reports of live news coverage, that is, instances in which the newsworthiness allowance should have been straightforwardly applied.

43. Meta. “Dangerous Organisations and Individuals.” Effective March 19, 2026.

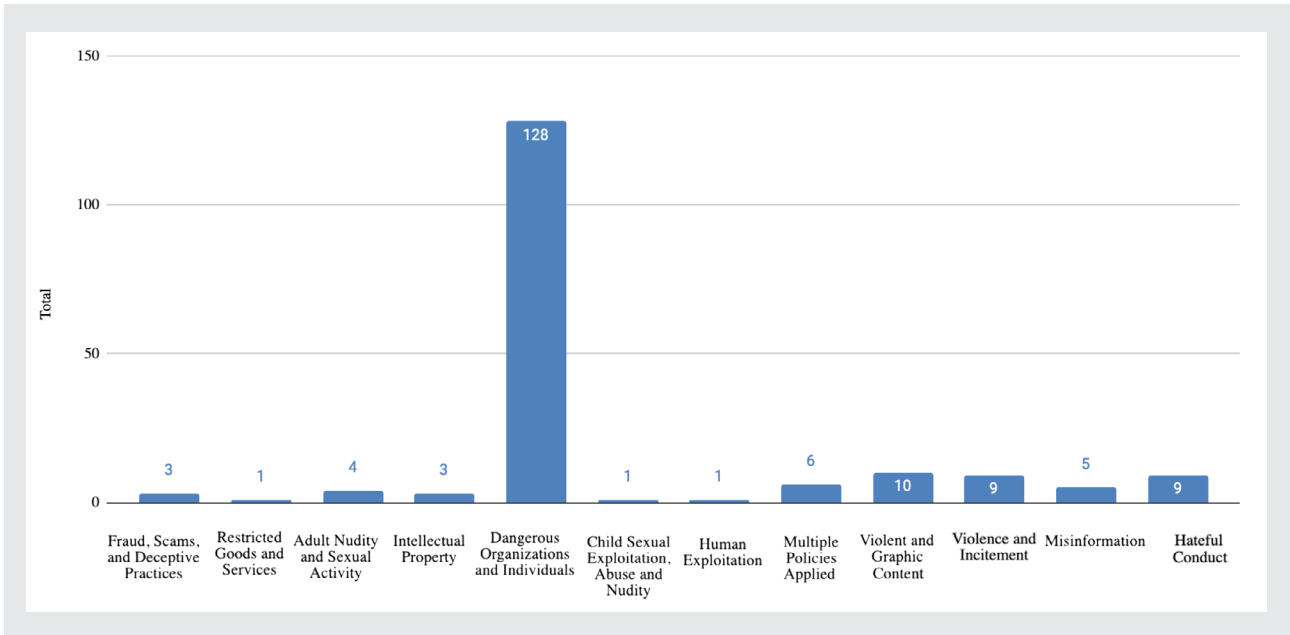


Figure 7. Distribution of Policies Applied to Journalists.

The cases in Figure 8a and Figure 8b are a clear example of how journalists are regularly moderated under DOI when posting about topics related to Palestine. In these specific instances, the journalist was moderated under DOI for reporting direct quotes by designated entities, which clearly included quotation marks and attribution.



Translation:

The party's new secretary-general in Lebanon, Naim Qassem:

If the Israelis decide to end the war, we will accept that, but on terms we deem appropriate; we will not beg for a ceasefire

Naim Qassem: We are not asked why we support Gaza; rather, others are asked why they did not support Gaza



Figure 8a. Post 1 by Mohamadghnama dated Oct 30th 2024.



Translation:

“Hamas Movement:

The movement’s leadership is studying, with a high sense of national responsibility, the proposal it received from the mediators and will provide its response as soon as possible, once the necessary consultations regarding it have been completed.

The movement reaffirms its unwavering position that any future agreement must achieve a permanent ceasefire, the complete withdrawal of occupation forces from the Gaza Strip, the conclusion of a genuine prisoner exchange deal, the launch of a serious process to rebuild what the occupation has destroyed, and the lifting of the unjust siege on our people in the Gaza Strip.”



Figure 8b. Post 2 by Mohamadghnama dated April 14th 2025.

Both posts clearly do not praise or support the individual's quotes, and thus do not constitute inappropriate content under Meta's DOI policy. Nevertheless, these posts received a warning and were subsequently taken down. All posts were later restored, further confirming that Meta's original moderation was indeed wrongful. Our analysis indicates that the same trend extends to different types of journalistic content: from relaying official statements on ceasefires to neutral reporting of events as they unfold. The case in Figure 9 is another exemplary and very common instance of journalistic content moderated under DOI. This post provided a breaking news update on the events of October 7 as they unfolded. Although Hamas, a designated organisation, is mentioned in the post, there are no indications of support or praise whatsoever.



“

Translation:

“#LatestNews

At least 22 Israelis killed and more than 500 were injured in a Hamas attack on a settlement...

You shared this content on someone's profile”

”

Figure 9. A journalist's post reporting on October 7 was flagged for DOI.

Our analysis also reveals that Meta fails to apply its newsworthiness allowance to both independent journalists and established Palestinian media outlets. The latter should be more easily recognised and protected as journalistic work. Like for the journalist classification, media outlets have been disproportionately moderated under DOI, accounting for 102 cases (59.3%) of the 172 cases with an identifiable policy in the dataset (Figure 10). Likewise, of these 102 cases, 28 cases (27.5%) were restored, further indicating Meta's own acknowledgement of wrongful moderation.

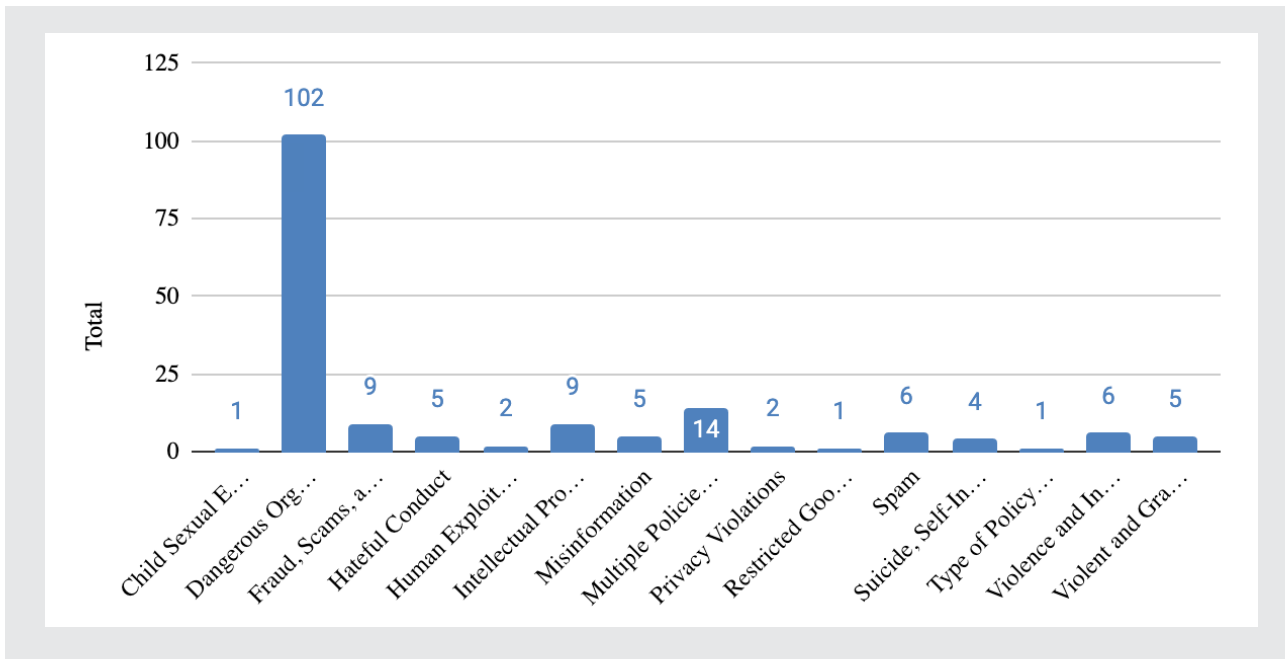


Figure 10. Distribution of Policies Applied to Media Outlets.

While the classification of individual users as journalists can be complex for moderation and algorithms to detect, the newsworthiness of content produced by established Palestinian media pages should be more straightforward and thus easier for Meta to detect. Our analysis indicates that Palestinian media outlets such as Ramallah News, Eye On Palestine, Palestine Info Center, and Ask Jerusalem are systematically censored under the DOI. Despite reporting news on current events in Palestine through solid journalistic standards, these accounts are regularly subjected to moderation under DOI due to the terms they use, such as references to Hamas or the Jenin brigade (the military wing of the Islamic Jihad), without any assessment by Meta of the newsworthiness of the post. Figure 11a and Figure 11b show two of the cases submitted by Ramallah News in December 2023. They are indicative of how Meta systematically silences - in these cases, through account restrictions - Palestinian media outlets.

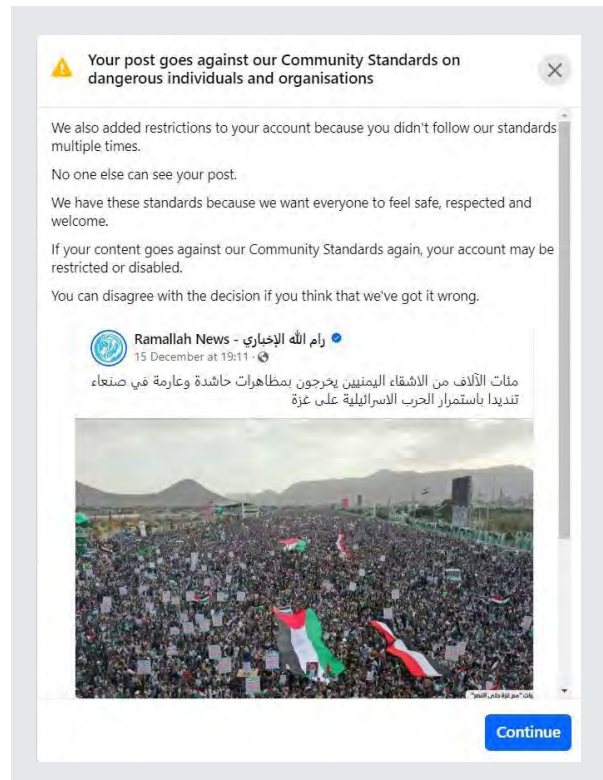


Figure 11a. First post by Ramallah News Figure 11b. Second post by Ramallah News

The first post (Figure 11a) reports the words of the former President of Iran, Ebrahim Raisi: *“Palestine was and still is the first issue for Islamic countries and for all free people in the world”*. It was moderated under DOI and taken down. Similarly, taken down under DOI, the second post (Figure 11b) reports on a demonstration in Sanaa (Yemen) against the genocide in Gaza. Translation: *“Hundreds of thousands of Yemeni brothers take out massive demonstrations in Sanaa denouncing the continuation of the Israeli war on Gaza”*. Following appeals by 7amleh, Meta restored four of the nine posts (44.4%) that had been taken down from the Ramallah News page, including the two mentioned above.

Regarding Meta’s moderation of media outlets, our qualitative assessment indicates that of a total of 465 cases of media outlets moderated under DOI, only 20 cases (4.3%) included sufficient grounds for DOI application (see Figure 12).

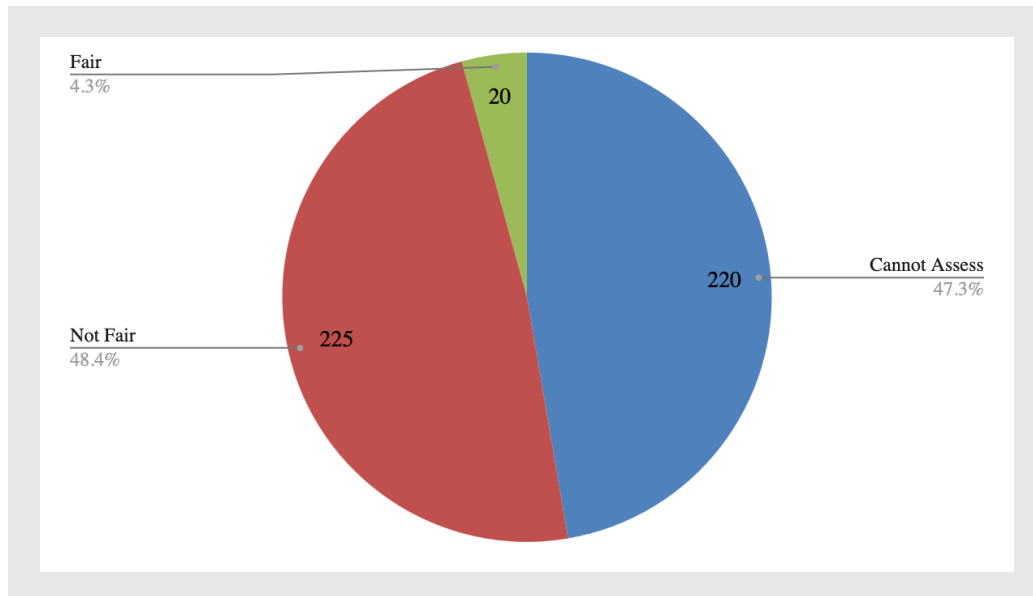


Figure 12. Qualitative Assessment of Media Outlet Cases Moderated under DOI.

Moreover, our analysis also reveals that such disproportionate DOI application to media outlets remained unchanged over time, despite Meta's various amendments to the policy. Figure 13a shows another post by Ramallah News reporting about the funeral of a Palestinian man killed during an Israeli military operation in the Old City of Nablus on 2 April 2025. The content was taken down and later restored after an appeal.



Translation:

“Journalist Amiid Douikaat writes: The funeral procession of martyr Hamza Khmash from Nablus!! We have reached a really hard stage, not only is the West Bank out of touch with Gaza, but also Ramallah is out of touch with Jenin...”



Figure 13a. Third Post by Ramallah News

Similarly, Figure 13b shows a post by Palestinian radio Raya FM, which relayed a global news piece about the Qassam Brigades claiming responsibility for rocket attacks. Meta took down this content and later restored it. While the post mentions a DOI-designated entity indeed, it does not contain any praise or glorification.



“

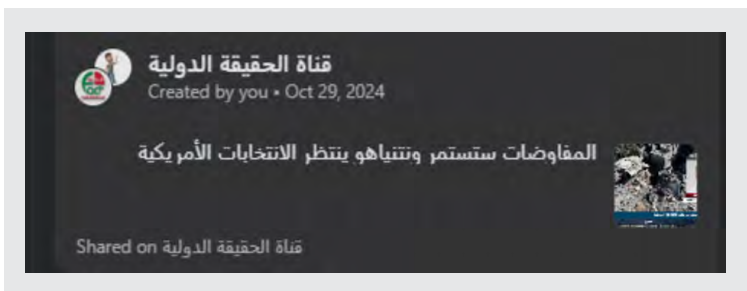
Translation:

“The Qassam Brigades claim responsibility for shelling Netivot, Okafim, Ashdod and Ashkelon with dozens of rockets in response to Israeli aggression”

”

Figure 13b. Raya FM’s post relaying the global news segment

Furthermore, Meta’s extensive use of DOIs extends to international media outlets. Figure 14 shows a post by a Jordanian-based newspaper reporting on the stalemate of ceasefire negotiations in October 2024. In this case, Meta issued a warning and enforced a limited distribution before eventually restoring the content after appeal.



“

Translation:

“Negotiations will continue, and Netanyahu is waiting for the US elections.”

”

Figure 14. A post by a Jordanian Outlet discussing negotiations.

Another interesting case is included in Figure 15 below. This case pertains to the Palestinian Sada News Agency, which was moderated under the DOI for sharing an Israeli news update about Hamas, which included a link to the original Israeli source. As with the previous cases, the content was taken down and restored after an appeal.

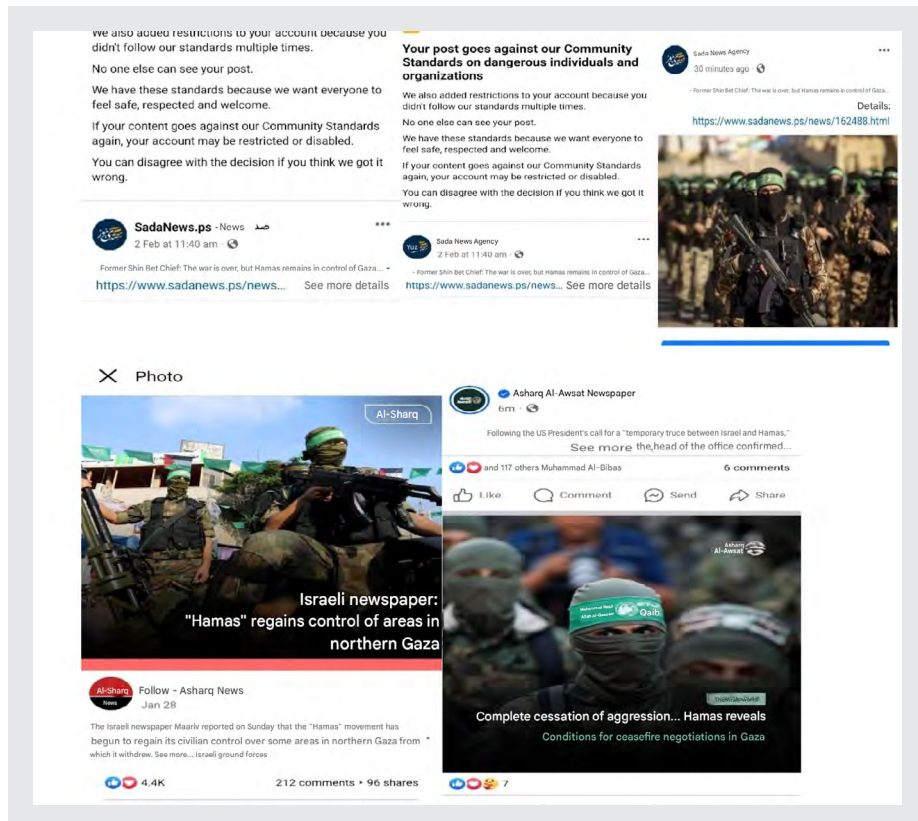


Figure 15. A Palestinian Media Outlet relaying Israeli News, Jan 2022.

All the cases analysed in this section indicate a clear pattern. None of these cases presents any ambiguity that might reasonably be interpreted as support or praise for DOI entities, yet they have been moderated by Meta under DOI, with only a small percentage being restored after appeal. With regard to restorations, it is important to note that, especially for media outlets, a later restoration nevertheless creates a temporal silencing that might render the news item obsolete or irrelevant by the time the restoration occurs.

2.2.2. Focus Issue: the Term ‘Shaheed’- Are Palestinians Allowed to Mourn?

Another important example of Meta’s disproportionate use of the DOI policy involves the term “shaheed”. Deriving from the Arabic verbal root shahada, meaning to ‘see’ or ‘witness’, Shahid or Shaheed is a person who sees and witnesses, mirroring the original meaning of the English translation “martyr”. However, where the usage of martyr has evolved to specifically describe those who suffered persecution and death for their religious or political beliefs, the Arabic “shaheed” more closely reflects the original meaning, and is therefore used much more commonly to describe anyone who has died for a good cause, in war, or whose death was not in vain. The moderation of the term “shaheed” has been a known issue for Meta. In March 2023, the Oversight Board accepted a request to review Meta’s moderation approach to the term.⁴⁴ The Board highlighted concerns about Meta’s blanket removal policy, which stifled freedom of expression and disproportionately affected Arabic-speaking users. Following the Board’s advice, Meta refined its approach to the moderation of the term and now defines it as “an honorific term with various meanings, including referring to those who die honourably or unexpectedly”.⁴⁵

In our analysis, 17 cases pertained to the moderation of the term “shaheed”. Of these, 14 have an identifiable policy. As shown in Table 4 and Figure 16, DOI is the primary policy applied in these cases (85.7%).

Table 4. Distribution of policy application of cases containing the use of shaheed

Type of Policy	Number of Entries	Proportion of Total Entries(%)
Dangerous Organisations and Individuals	12	85.7%
Hateful Conduct	2	14.3%
Total	14	100%

44. Oversight Board “Oversight Board Announces a Review of Meta’s Approach to the Term ‘Shaheed,’” News. 9 March, 2023.

<https://www.oversightboard.com/news/1299903163922108-oversight-board-announces-a-review-of-meta-s-approach-to-the-term-shaheed/>

45. Meta. “Expanding Free Expression.” Tracking Impact. Updated August 28, 2025. <https://transparency.meta.com/governance/tracking-impact/expanding-free-expression>; See also Meta. “PAO on the treatment of the term ‘shaheed.” Oversight Board Cases. Updated July 02, 2024.

<https://transparency.meta.com/oversight/oversight-board-cases/shaheed-pao>

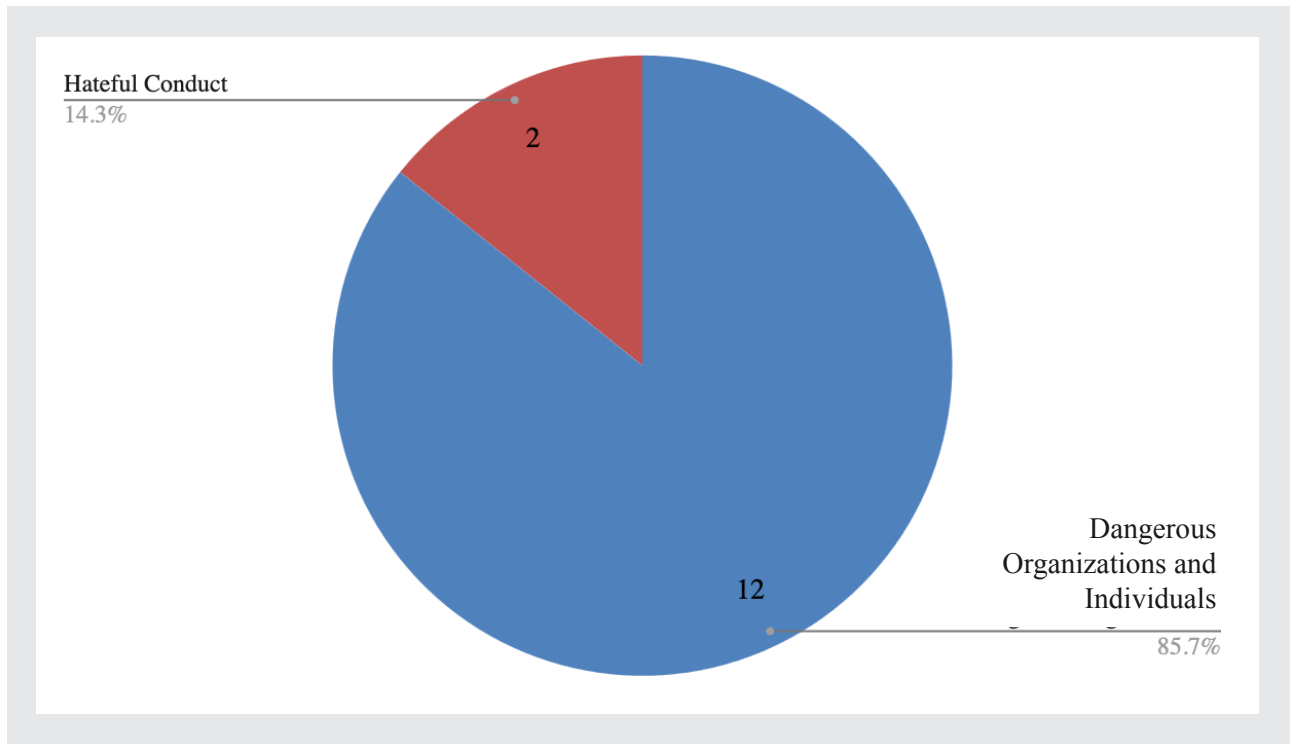


Figure 16. Distribution of Policies Applied by Meta to the term Shaheed.

The qualitative assessment reveals that 70.6% (12 cases) were wrongfully moderated because they did not include any element of praise or glorification of DOI-designated entities. Notably, the over-enforcement of these cases occurred primarily on Facebook, with fewer cases appearing on Instagram.

Table 5. Assessment of the fairness of policy application to the shaheed cases

Assessment policy	Number of Entries	Proportion of Total Entries(%)
Fair	5	29.4%
Not Fair	12	70.6%
Total	17	100%

Turning to the types of enforcement applied in cases where the term “shaheed” was mentioned, content takedowns emerge as the most common sanction imposed by Meta (Table 6 and Figure 17). However, harsher enforcement measures, including restrictions and suspensions, were also applied. This is significant because such measures extend the sanction beyond the term’s specific use in a single post, targeting the account more broadly.

Table 6. Types of enforcement applied to shaheed cases

Type of Enforcement	Number of Entries	Proportion of Total Entries (%)
Content Takedown	12	70.6
Suspension	2	11.8
Restriction	2	11.8
Reduced Distribution	1	5.9
Total	17	100

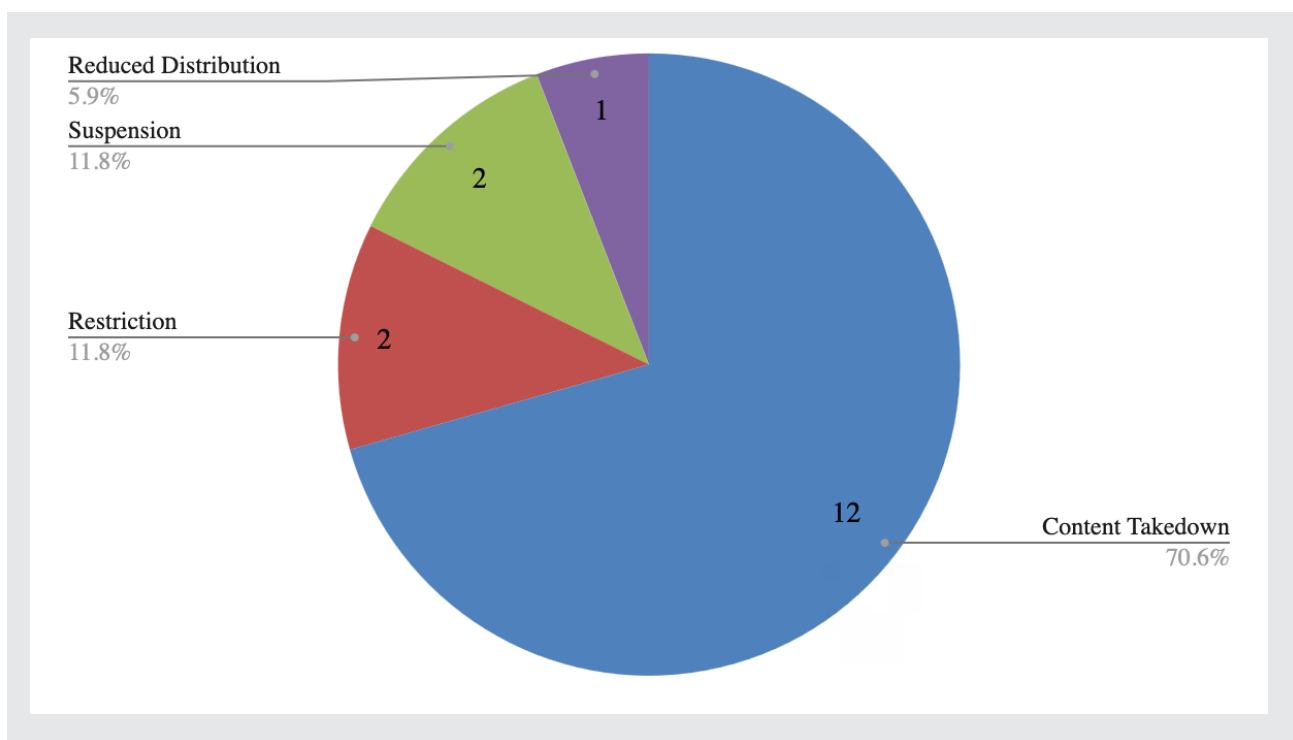


Figure 17. Distribution of Types of Enforcement on Shaheed Cases.

Looking at the actual content of these posts, the two cases below exemplify how Meta approached their moderation. The case in Figure 18 contains a post in which the user pays posthumous homage to his deceased mother. Meta took this post down under DOI.



Figure 18. A User Paying Homage to Their Deceased Mother.



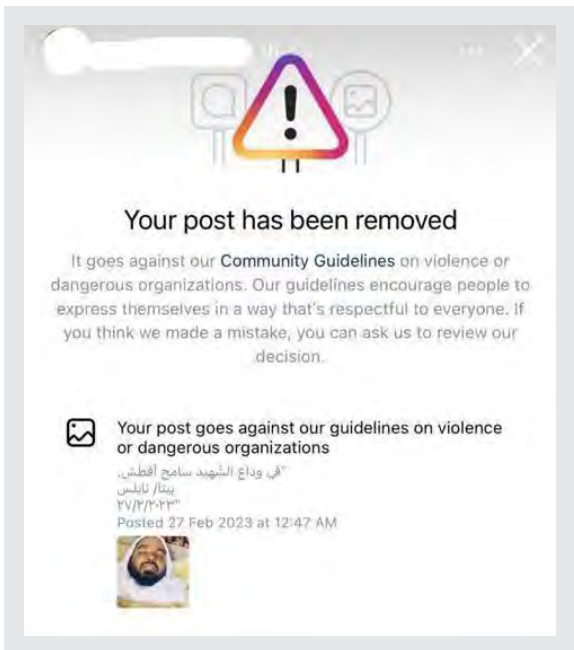
Translation:

“May God bless me with the blessing of my mother and father’s prayers at all stages, and I always saw myself as having abundant luck, ease and success, until Israel targeted the homes of my relatives in the city of Deir al-Balah, to which we were displaced after the evacuation of the Karama neighbourhood in the northern Gaza Strip.

I asked about them while we were looking for the rest. My aunt lied and said that she had called my mother and checked on her. I believed her, and I went to check on my father. I saw him on the emergency bed surrounded by five doctors and more. He was calm, stable, and silent. I entrusted him to God and to the prayers of my friends. He entered intensive care and has remained there until now. Now, I went back to make sure my mother was safe, and I heard a voice saying, “A woman wearing a double-strap ring. Do you know her?” She said, “My aunt is your mother.” He said, “Go to the martyrs’ tent and say goodbye to her.” I did not go. I saw this picture of her in front of me, and I wanted it to be the last. I took it on the Friday morning before the war. She was in her last days, like a bird that had given up her spirit for a while, but she did not tell us about it. Waiting Abdullah Aziz, son of Halima, from the occupied town of Simsim, Gaza District, was born in Jabalia, lived 56 years, was a good, simple, ascetic, strong and faithful woman. She gave birth to and raised 7 daughters and 3 sons. She hoped for peace and a good ending, and that she would die at the age of 60, not be humiliated, nor weakened, or degraded by old age or need, and she achieved what she wished for. May God have mercy on my mother.. to our hearts and to your pure soul”



Similarly, the case in Figure 19 further reveals a lack of nuance in Meta’s policy application regarding the term “shaheed.” This case involved a post in which a user bid farewell to Sameh Aqtash, on 27 February 2023, who was referred to as a martyr.



“

Translation:

“Last moments of goodbye for martyr Sameh Aqdash Bayta/Nablus 27 Feb 2023”

”

Figure 19. A User Bidding Farewell to Sameh Aqdash.

Aqtash was shot by Israeli illegal settlers outside his home near Hawara in February 2023, shortly after returning from volunteering in earthquake relief efforts in Turkey. The shooting occurred near the perimeter fence of the Palestinian village of Zaatara, where a crowd of Israeli settlers and soldiers had gathered.⁴⁶ Sameh Aqtash was an ordinary Palestinian man with no ties to organisations that Meta had labelled as dangerous or violent. As mentioned, thanks to 7amleh and other NGO’s advocacy work, Meta acknowledged and amended its problematic moderation approach to the term “shaheed” in July 2024.⁴⁷ Our analysis reveals that no cases in the dataset were moderated for containing the word ‘shaheed’ from July 2024 and throughout 2025.

46. Bateman T., “Hawara West Bank: ‘What Happened Was Horrific and Barbaric,’” BBC News, 27 February 2023, <https://www.bbc.com/news/world-middle-east-64757990>; Obel A. and Fabian E., “Relatives of Palestinian Killed during Settler Riot Say He Was Shot by ‘IDF Troops,’” The Times of Israel, 27 February 2023, updated 28 February 2023, <https://www.timesofisrael.com/relatives-of-palestinian-killed-during-settler-riot-say-he-was-shot-by-idf-troops/>.

47. Meta. “Expanding Free Expression.” Tracking Impact. Updated August 28, 2025. <https://transparency.meta.com/governance/tracking-impact/expanding-free-expression>; See also Meta. “PAO on the treatment of the term ‘shaheed.’” Oversight Board Cases. Updated July 02, 2024. <https://transparency.meta.com/oversight/oversight-board-cases/shaheed-pao>

2.3. Hateful Conduct

While the application of Hateful Conduct in Meta's moderation is not as prolific as DOI, this policy ranks as the second most commonly employed policy⁴⁸ with 78 cases representing 6.8% of all cases with an identifiable policy. As mentioned in the introduction, this policy is designed to protect individuals from hate speech, including harmful stereotypes, dehumanising speech, and other attacks based on their protected characteristics. Initially referred to as 'Hate Speech', the policy underwent several updates through the years and, as mentioned in Section 1.3 of the introduction, it received a substantial relaxation since January 2025 in the name of Meta's ambition to "expand free expression" and "get back to our roots".⁴⁹ Alongside a rollback of the 3rd-party moderation program in the US, Meta implemented a major rewording of the hate speech policy, including changing the title to 'Hateful Conduct' and relaxing some of its embedded protections.⁵⁰ Most notably, despite Meta centring the protected characteristics in the policy's original construction, the steady relaxations are particularly pronounced for certain groups and topics, which "allow room" for "call[s] for exclusion or use [of] insulting language in the context of discussing political or religious topics, such as when discussing transgender rights, immigration or homosexuality".⁵¹ These changes, in general, reflect a broader trend towards aligning Meta's policies with more conservative American political discourse and 'free speech'.

Going back to our analysis, we identified that among the cases moderated under the Hateful Conduct policy, activists are the largest represented classification (Figure 20).⁵² This finding is significant because it suggests that enforcement under the Hateful Conduct policy falls heavily on activists' accounts. Given that activists use social media for advocacy, their overrepresentation under Hateful Conduct raises concerns that Meta's moderation may restrict the broader visibility of Palestinian rights-based advocacy and engagement.

48. Excluding cases where multiple policies were applied.

49. Meta. "More Speech, Fewer Mistakes" Newsroom. January 07, 2025. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>

50. Meta. "Hateful Conduct" Effective January 7, 2025.

51. Meta. "Hateful Conduct." Effective January 7, 2025.

52. Other encompasses users who, when submitting their claim to Meta, did not identify under one of the specific classifications provided.

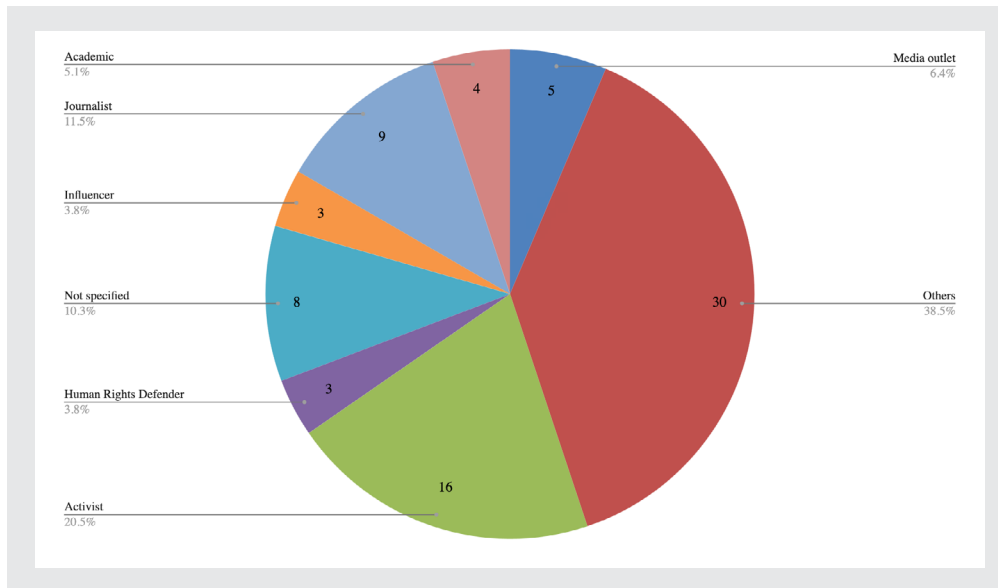


Figure 20. Distribution of Classifications Moderated under the Hateful Conduct Policy.

2.3.1. Focus Issue: Zionism: A Problematic Conflation?

During the genocide, in July 2024, Meta announced and implemented a new policy regarding the use of the term 'Zionism' on its platforms⁵³. According to this announcement, and following a review from 'different perspectives' in March 2024, the Hateful Conduct policy will be applied to "remove speech targeting 'Zionists' in several areas where our process showed that the speech tends to be used to refer to Jews and Israelis with dehumanising comparisons, calls for harm, or denials of existence".⁵⁴ In this announcement, Meta does acknowledge the multiple layers behind the term, stating that "*the word "Zionist" has layers of meaning based on its origins and usage today, and may also be highly dependent on context*".⁵⁵ Here, Meta recognizes that it "*often refers to supporters of a political movement, which is not itself a protected characteristic under our policy*". However, the newly-introduced policy also clarifies that "*in some cases may be used as a proxy to refer to Jewish or Israeli people, which are protected characteristics under our Hate Speech policy*".

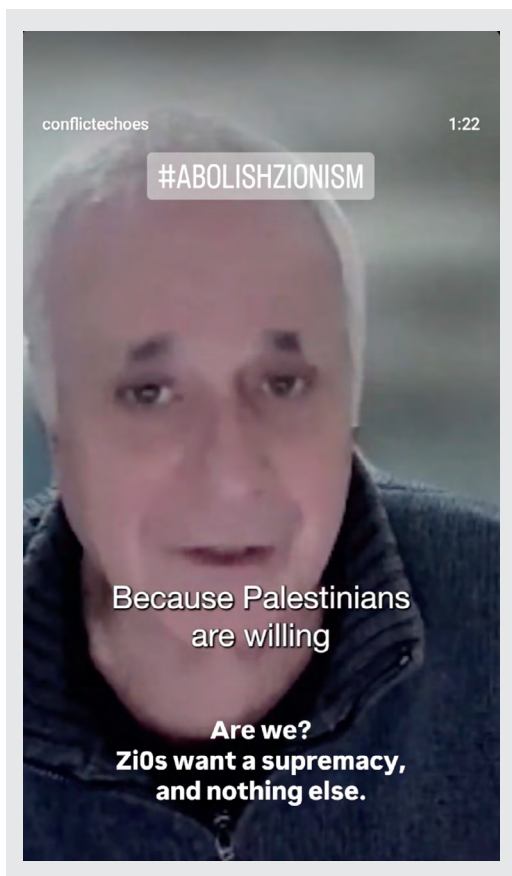
This is reinforced by Meta's own stated rationale for its Hateful Conduct policy, which is intended to prevent attacks against people on the basis of protected characteristics. It is important to notice that, while Meta has relaxed protections for some vulnerable groups in January 2025, including immigrants and women, it has simultaneously expanded protection around the term "Zionist" by treating certain uses of the word as potential proxies for protected characteristics.

53. See: Meta. "Update from the Policy Forum on our approach to 'Zionist' as a proxy for hate speech" Policies. Updated July 9, 2024,

54. Ibid

55. Ibid

A clear case (#348) showing the ramifications of this over-protection of the word Zionism can be seen by the moderation under Hateful Conduct of two separate Instagram stories from November 2025. Leading to the suspension of the account, the stories in question were reposting material that discussed the very issue of conflating Zionism and Judaism. Posted by a pro-Palestinian Israeli activist, the first post (Figure 21a) is a snippet from a video of Israeli historian Ilan Pappè discussing Zionism, arguing that Palestinians only require equal citizenship while Israel demands domination. In the video, the Historian says, *“The Palestinians are asking for equal citizenship, it is up to the Jews to do the same.”* The story is captioned with “Are we? Zio’s want supremacy and nothing else”. In this case, the term Zionism is not being used as a dehumanising or exclusionary reference to Jews or Israelis, but as the same criticism of Zionism as a political project.



“

Transcript:

“Are Jews willing to live as equal citizens? Because Palestinians are willing to live as equal citizens. All their life they are fighting to be equal. That’s what they are asking. They are asking for natural, normal life. They are not, they don’t ask....”

”

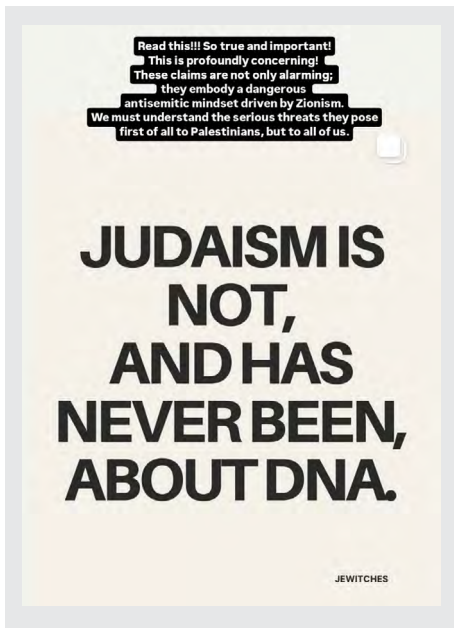
“

Caption:

*“#ABOLISHZIONISM
Are we? Zio's want a supremacy, and nothing else”*

”

Figure 21a. A Repost of a Video Snippet Of Israeli Historian Ilan Pappè Criticising Zionism. The same user reposted a separate post (Figure 21b), written and posted by a Jewish organisation account, specifically criticising dangerous rhetoric applying race theory ideologies to Judaism, where one’s identity as Jewish is related to their DNA. The repost encourages users to read the post, highlighting the alarming antisemitic mindset embodied in the claims and their implications. Again, there is no hateful discourse targeting a nationality or religion in this post.



“

Caption:

“Read this!!! So true and important! This is profoundly concerning! These claims are not only alarming; they embody a dangerous antisemitic mindset driven by Zionism. We must understand the serious threats they pose first of all to Palestinians, but to all of us.”

”

Figure 21b. A Repost Highlighting a Dangerous Claim and Mindset Propagated by Zionism.

2.3.2. Focus Issue: Recommended content after October 7

Our analysis of Meta’s policy application indicates that algorithms operate on two different levels. First, algorithms are used to moderate content and enforce community standards. Second, they also operate by recommending content to users. In this section, we focus on Meta’s policy changes on recommended content after October 7. With regards to Facebook, Meta defines recommended content as follows:

“

“Facebook’s goal is to make sure that you see posts from the people, interests and ideas that you find valuable, whether that content comes from people you’re already connected to or from those you may not yet know. When you open Facebook and see Feed in your Home tab, you experience a mix of “connected content” (e.g. content from the people you’re friends with or are following, groups you’ve joined and Pages you’ve liked), as well as “recommended content” (e.g. content we think you’ll be interested in from those you may want to know). We also show you ads that are tailored to you.”⁵⁶

”

Through the logic of recommended content, Meta’s algorithms regulate the distribution of specific content and influence what users see in their feeds. Following the beginning of the genocide in October 2023, Meta took an unprecedented effort to moderate content on its platforms through algorithmic scanning.⁵⁷

56. Meta. “Our approach to Facebook Feed ranking.” Features. Updated June 11, 2025. <https://transparency.meta.com/features/ranking-and-content/>

57. Meta. “Meta’s Ongoing Efforts Regarding the Hamas-Israeli War.” Newsroom. October 13, 2023. <https://about.fb.com/news/2023/10/metass-efforts-regarding-hamas-israel-hamas-war/>

“We already use technology to avoid recommending potentially violating and borderline content across Facebook, Instagram and Threads. We’re working to further reduce the possibility of this happening by lowering the threshold at which our technology will take action to avoid recommending this type of content.”

Lowering the algorithm’s sensitivity creates space for more false positives to be moderated. This is particularly relevant given the vague language used in this policy change, defined as “potentially violating and borderline content.” From this poorly specified clause, we can infer that Meta’s moderation does not necessarily tackle only content that is considered to be in violation of community standards.

Meta’s algorithmic moderation has repeatedly been shown to disproportionately affect Palestinian content. This needs to be understood alongside Meta’s documented under-moderation of Israeli hate speech. As 7amleh’s previous research found, the company “failed to allocate sufficient resources to address Hebrew-language incitement.”⁵⁸ Taken together, the under-moderation of Hebrew-language incitement and the wrongful moderation of Palestinian content suggest an uneven enforcement environment. In this context, lowering Meta’s detection threshold for potentially harmful content, combined with the vagueness of the relevant policy language, most likely increased the risk that Palestinian content and users would be wrongfully moderated, including in cases where the content did not violate Meta’s Community Standards.

To assess whether this threshold change was associated with increased wrongful moderation of Palestinian content, we conducted a quantitative analysis of Meta’s restoration rate before and after this policy change on the 13th October 2023. As shown in Figure 22, the restoration rate following appeal increased from 24% to 30%. Moreover, Table 11 (section 4.2) shows a marked increase in Meta’s own acknowledgement of wrongful moderation. The share of cases in which Meta recognised that content had been wrongly removed increased from 0.8% to 23%, while the share of cases involving incorrectly restricted accounts increased from 0.2% to 4.5%. Taken together, these findings indicate that the lowered sensitivity threshold of Meta’s algorithms contributed to higher levels of wrongful moderation affecting Palestinian content and users.

58. Abukhater, J. (2025, September). Meta’s role in amplifying harmful content during genocide in Gaza. 7amleh – The Arab Center for the Advancement of Social Media.

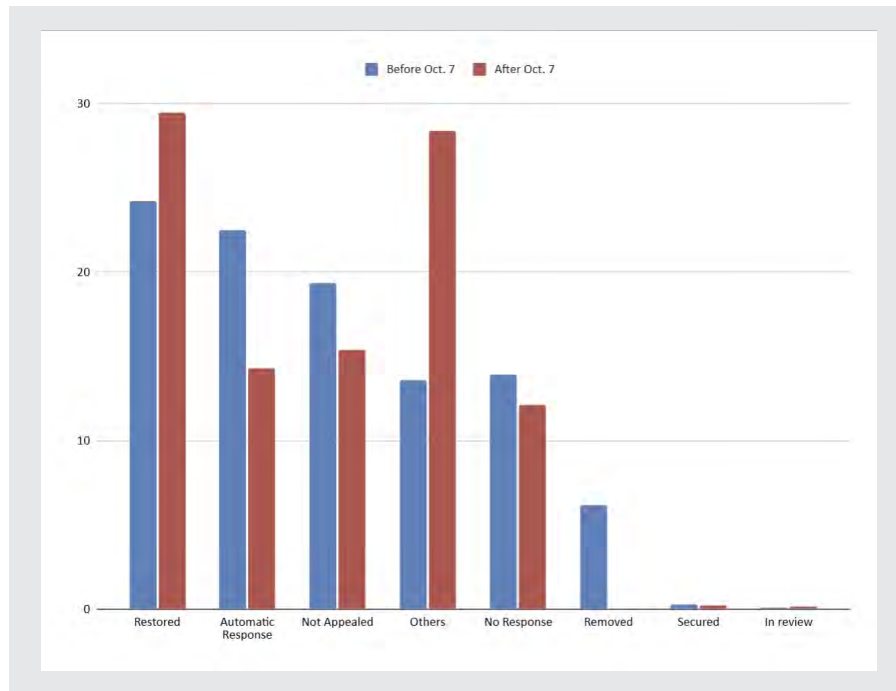


Figure 22. Distribution of Meta’s Responses Pre and Post October 7.

A second relevant policy change to algorithmic moderation occurred in January 2025, as part of the “More Speech, Fewer Mistakes” update. Through this new setting of its algorithms, Meta automatically reduced the recommendation of political content unless users manually changed it in their personal settings. In other words, political content started receiving reduced visibility even when it did not violate Meta’s Community Standards. By introducing a default reduction in the recommendation of political content, Meta creates an additional barrier for Palestinian and solidarity voices. This is especially significant because the filtering function is enabled by default rather than activated by users themselves, thus creating effects similar to shadowbanning for accounts discussing political issues. Case #659 offers a useful example of how Meta applies reduced distribution to content concerning politically and socially debated topics. As shown in Figure 23, the user received significant restrictions after posting about issues classified as socially debated.

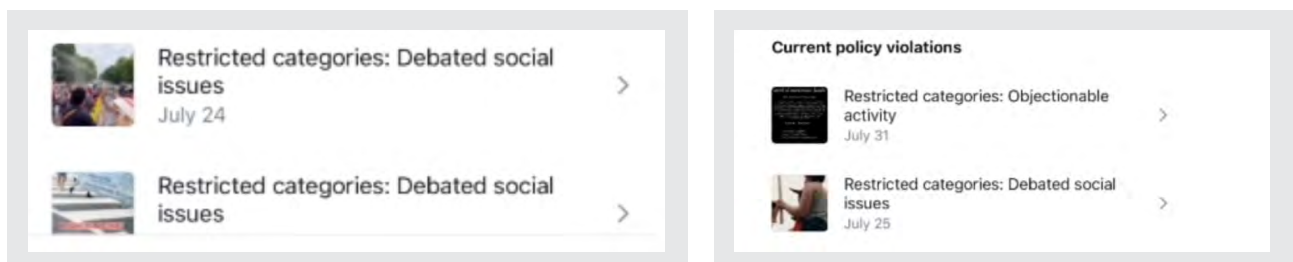


Figure 23. User Being Restricted Due to Posting About Socially Debated Issues.

3. Meta's Moderation Enforcement

The third chapter of this report provides an in-depth examination of Meta's moderation enforcement practices concerning Palestinian content and users. The primary objective is to identify patterns in the enforcement measures applied, the types of accounts and content most frequently targeted, and the broader implications of these practices for Palestinian presence and visibility across platforms. Special attention is given to how Meta's enforcement practices have evolved over time, with a chronological examination that sheds light on significant shifts in moderation approach. A key focus of this analysis is to highlight the disparities in Meta's handling of Palestinian content before and after October 7, examining how events during this period may have influenced moderation enforcement and decision-making.

3.1. Dominant Enforcement

For the entire dataset (including both Facebook and Instagram), Meta's moderation enforcement types appear to be distributed fairly evenly across its three primary categories: suspensions (33.7%), restrictions (29.5%), and content takedowns (25%), as shown in Table 7 and Figure 24. According to Meta, suspensions refer to the temporary or permanent disabling of user accounts due to violations of policies.⁵⁹ This enforcement action prevents users from accessing their accounts. Restrictions involve limiting certain functionalities of an account, such as the ability to post, comment, or interact with others, for a specified period. This measure is typically applied as a less severe penalty than suspension. Content takedowns entail the removal of specific posts, images, videos, or other content deemed to violate Meta's community standards, such as those related to hate speech, misinformation, or incitement to violence (including DOI).

59. Meta. "Disabling Accounts." Enforcement. Updated July 17, 2025.
<https://transparency.meta.com/enforcement/taking-action/disabling-accounts/>

Table 7. Distribution of enforcement mechanisms in Meta's moderation 2021 - 2025

Type of Moderation Enforcement	Number of Cases	Proportion of Total Cases(%)
Suspension	1187	33.7
Restriction	1039	29.5
Content Takedown	879	25.0
Warning ⁶⁰	163	4.6
Reduced Distribution	95	2.7
Shadowbanning	58	1.6
Page Unpublished	43	1.2
Other	56	1.6
Total	3520	100

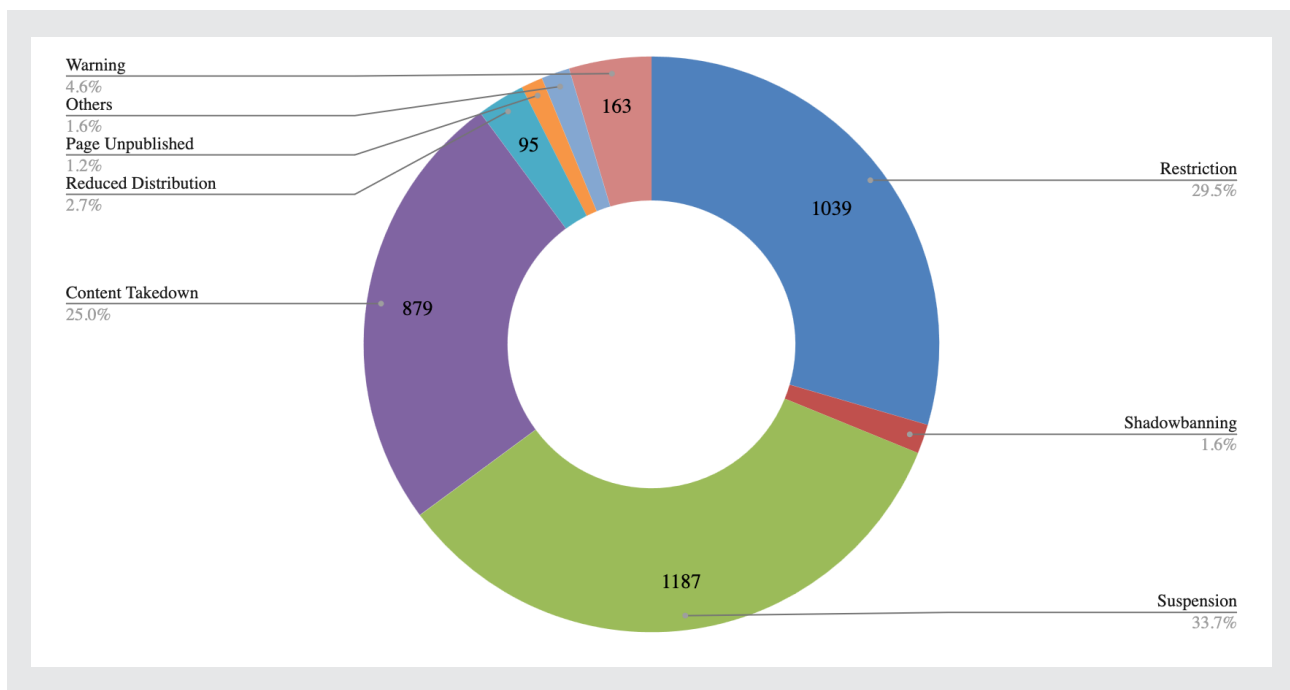


Figure 24. Distribution of Enforcement in Meta's Moderation 2021 - 2025

Considering Meta's various policy changes over time, our analysis reveals notable differences in the enforcement tools used before and after October 7, as detailed in Table 8. One particularly significant change pertains to the increase in shadowbanning, which rose from 0.1% to 3.5%. Shadowbanning remains a relatively underexplored and less visible form of moderation; for this reason, a dedicated section later in this chapter examines its effects in greater detail.

60. Refers to an action similar to content removal but which applies specifically to organisations' pages on Facebook.

Table 8. Distribution of enforcement in Meta's moderation before and after Oct 7th 2023

Type of Enforcement	Before Oct 7th		After Oct 7th	
	Number of Cases	Proportion of Total Cases(%)	Number of Cases	Proportion of Total Cases(%)
Suspension	689	35.6	498	31.5
Restriction	582	30.0	457	28.9
Content takedown	447	23.1	432	27.3
Warning	106	5.5	57	3.6
Reduced distribution	57	2.9	38	2.4
Others	30	1.5	26	1.6
Page Unpublished	25	1.3	18	1.1
Shadowbanning	2	0.1	56	3.5
Total	1938	100	1582	100

3.2. Suspension

As shown in Table 8, suspension constitutes the most widely used enforcement tool for moderating Palestinian content. This enforcement measure comes in two forms: user and device suspension. Both can be applied from a minimum of twenty-four hours up to a permanent suspension. Meta recognises that suspension is particularly powerful; thus, it is recommended to moderators to use it carefully and only after multiple warnings and other, less severe, tools.⁶¹ A suspension means the user is unable to log in to their profile, and other users cannot see it. In recent years, many users have complained about the surge in suspensions in Meta's moderation, with many finding their profiles disabled overnight and unable to retrieve information about their suspension and moderation.⁶² The surge in suspension waves is arguably tied to a general surge in the use of algorithmic models for content moderation. As some experts point out, the scale of algorithmic moderation application for suspension is particularly worrying, as they can eliminate whole digital ecosystems without sufficient scrutiny.⁶³ This happens because, in practice, suspensions

61. For example, as a recommendation given to VR developers with apps on the platform, see: Meta Horizon. "Suspension Service." Resources. Updated February 20, 2026. <https://developers.meta.com/horizon/resources/suspension-service/>.

62. Harry Richt, "Meta Account Suspensions: Understanding the 2025 AI Moderation Crisis | Richt Law Firm," Richt Law Firm, November 15, 2025, <https://richtfirm.com/meta-account-suspensions-understanding-the-2025-ai-moderation-crisis/>.

63. Ibid.

are often enacted at scale, targeting both the moderated user in violation of community standards and their related and connected profiles.

3.2.1. Focus Issue: Suspension of Civil Society, Journalists, and Media Outlets

As shown in the graph in Figure 25, activists, journalists, and media outlets account for 46,2% of the total cases suspended from Facebook or Instagram.

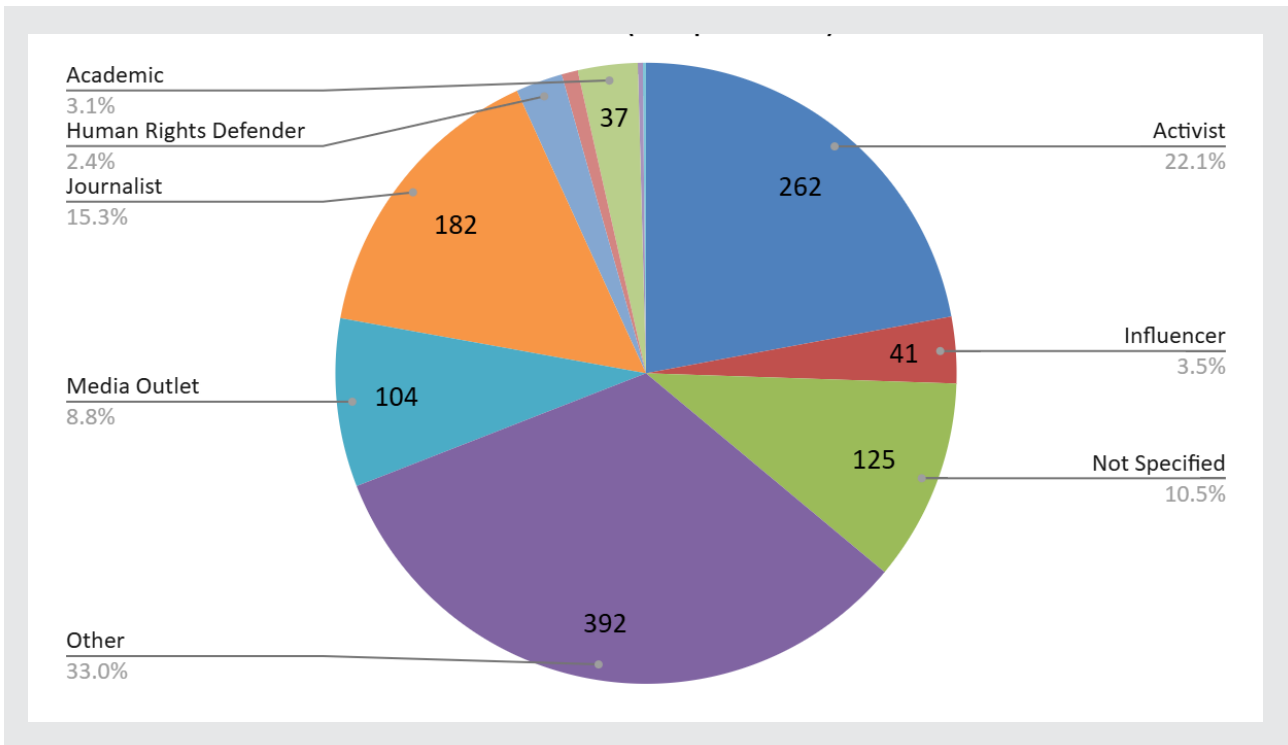


Figure 25. Distribution of Affected Classifications in Cases of Suspension.

Case #235 (Figure 26a) is one of the most exemplary data entries we have to showcase Meta’s wrongful enforcement through suspension. The user in question uploaded a post showing the death of newborns caused by Israeli forces during the genocide. While an argument could be made that the post’s fair moderation falls outside Meta’s policy on graphic content, the user here was moderated for child nudity infringement.

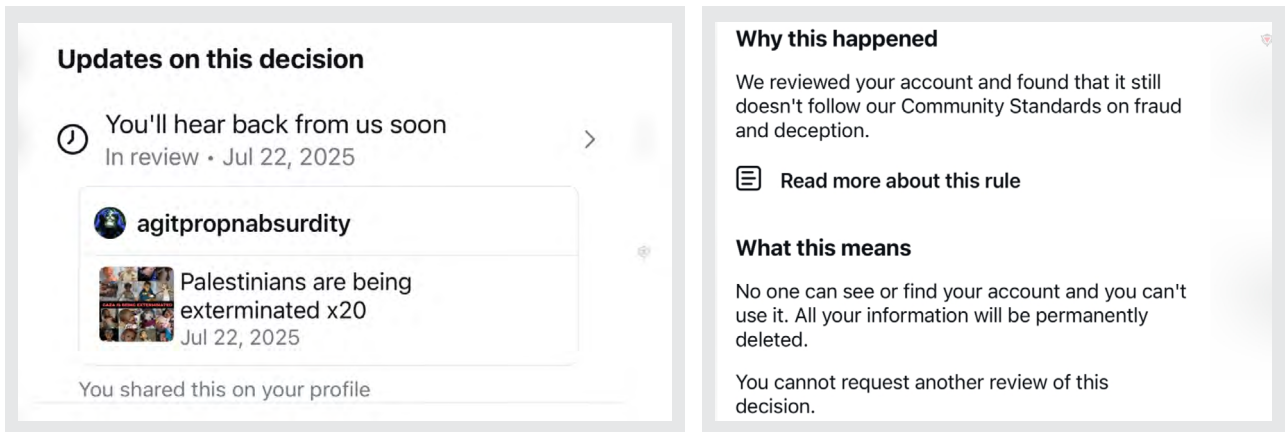


Figure 26a. A Post Raising Awareness of the Murder of Palestinian Newborns Under Israel's Genocide Being Suspended for Child Nudity Infringement.

The user, an activist, received a permanent suspension, without possibility of appeal and lost the account. The use of a permanent suspension underscores the magnitude of this enforcement tool and how misapplying it can significantly hinder users' efforts to raise awareness on Meta's platforms. In this case, the use of powerful moderation tools for violations that would usually receive less severe restrictions is evident. The impossibility of appealing the decision also shows the unilateral nature of Meta's moderation and a lack of communication. Without the ability to communicate with the platform and ensure clear, transparent application of its policies, Meta's overreliance on suspensions normalises overenforcement.

Similarly, case #353 pertains to a public statement by the Global Shapers Centre in Ramallah regarding the Boycott of the 2025 World Economic Forum Annual Meeting (Figure 26b).

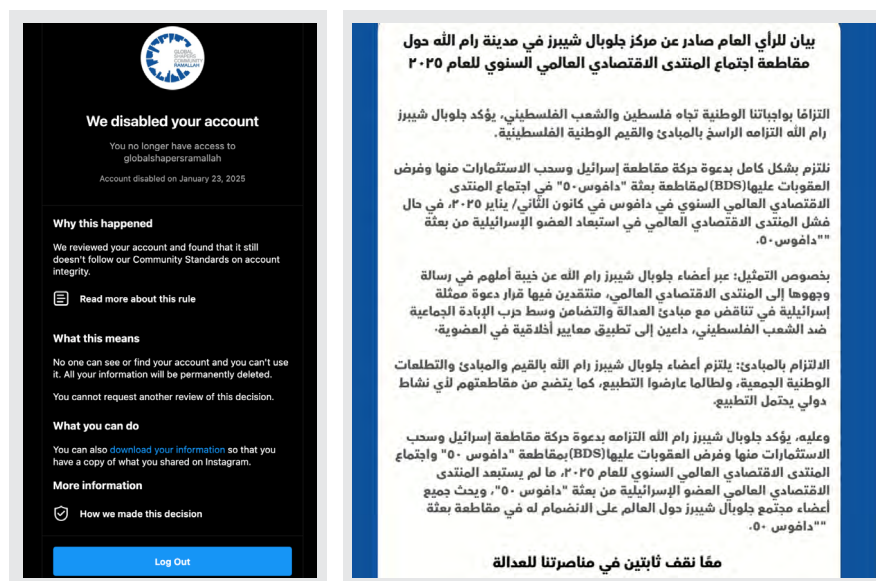


Figure 26b: Global Shapers' Ramallah Post and the Moderation Received.



Translation:

“In fulfilment of our national duty towards Palestine and the Palestinian people, Global Shapers Ramallah reaffirms its unwavering commitment to Palestinian national principles and values. We fully support the call by the Boycott, Divestment, and Sanctions (BDS) movement to boycott the Davos 50 delegation at the World Economic Forum Annual Meeting in Davos in January 2025, should the World Economic Forum fail to exclude the Israeli member from the Davos 50 delegation. Regarding Representation: In a letter to the World Economic Forum, Global Shapers Ramallah members expressed their disappointment, criticising the decision to invite an Israeli representative in contradiction to the principles of justice and solidarity amidst the genocidal war against the Palestinian people, and calling for the application of ethical standards in membership. Commitment to Principles: Global Shapers Ramallah members are committed to collective national values, principles, and aspirations, and have always opposed normalisation, as evidenced by their boycott of any international activity that could potentially lead to normalisation. Accordingly, Global Shapers Ramallah reaffirms its commitment to the Boycott, Divestment, and Sanctions (BDS) movement’s call to boycott Davos 50 and the 2025 World Economic Forum Annual Meeting, unless the World Economic Forum excludes the Israeli member from the Davos 50 delegation, and urges all members of the Global Shapers community worldwide to join in boycotting the Davos 50 delegation. Together we stand firm in our advocacy for justice”



As the post’s text shows, there is no apparent violation of Meta’s Community Standards by the organisation. The organisation is stating its position on the World Economic Forum and expressing support for the BDS campaign. Meta’s explanation for the suspension was limited to the claim that the user’s activity did not align with its Community Standards, without providing further details. Although the account was eventually restored, this case illustrates the reach and impact of Meta’s wrongful moderation through account suspension.

Besides activists, as shown in Figure 25, journalists and media outlets are also particularly targeted by suspensions. Suspending the account of a journalist or of a media outlet corresponds to silencing reporting, and should only be adopted in exceptional cases. In our qualitative assessment of suspension cases, we came across the interesting case of Motaz Azaiza, a journalist from Gaza. He is an extremely influential Palestinian journalist who has been regularly reporting from Gaza during the genocide. His account currently counts over 14 million followers, making him one of the journalists from Gaza with the biggest reach for their content. Motaz reported multiple cases to 7or. In Figure 26c, Cases #1162 and #1149 show the restrictions given to his account at the beginning of the genocide. In these cases,

Meta's combination of different enforcement tools effectively hindered the possibility for Motaz to report on the genocide. In these two cases, Meta removed some content of the user, suspended the account from going live and blocked access to non-followers to Motaz's content and account. As we can see, Meta applied a series of enforcement tools, including blocking algorithmic recommendations of the content, thus impacting the account's potential reach.

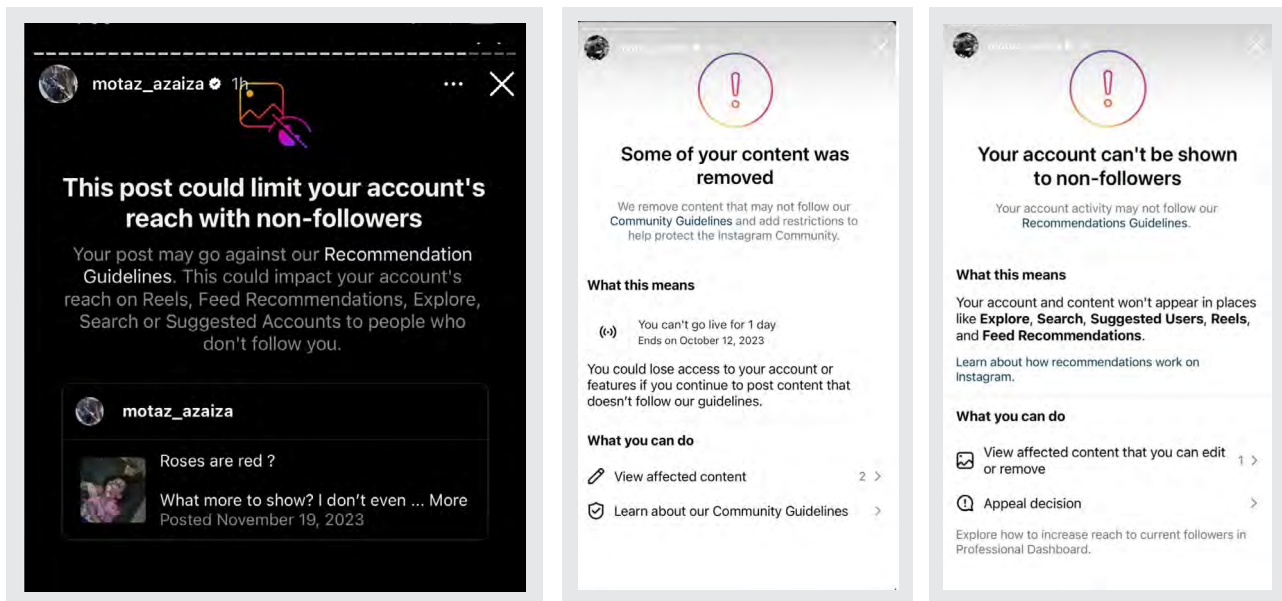


Figure 26c: Suspension of Journalist Motaz Azaiza on Instagram

Our analysis also reveals another dimension of Motaz Azaiza's suspension. In Figure 26d, case #675, the Gaza-based journalist claims that a Meta engineer was fired for discussing and resolving the moderation of one of his videos.⁶⁴ The engineer started a lawsuit against Meta over the firing, denouncing the specific instance that led to his termination as well as the general discrimination of the company towards Palestine-related content. This is an extremely interesting case because it indicates the extent to which Meta is willing to go to silence Palestinians, and Palestinian journalists specifically.

64. Tilo D., "Meta Faces Legal Action over Engineer's Wrongful Firing," Human Resources Director Canada, 6 June 2024, <https://www.hcamag.com/ca/specialization/employment-law/meta-faces-legal-action-over-engineers-wrongful-firing/492285>.

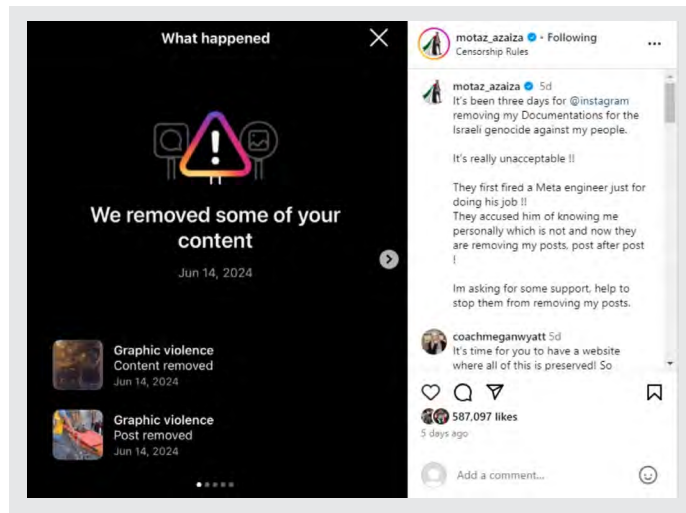
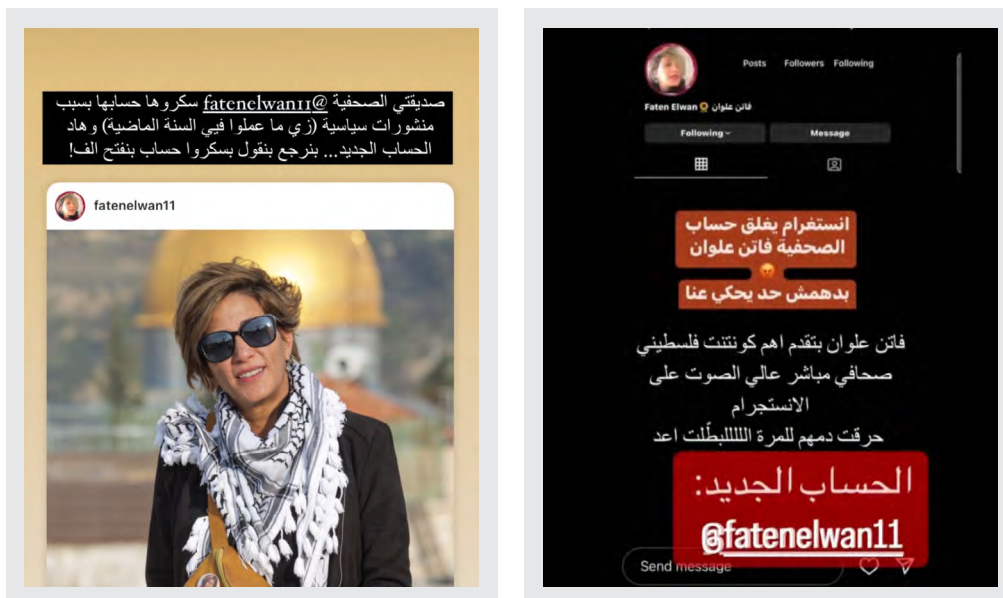


Figure 26d: Internal Meta Dispute over the Moderation of Motaz Azaiza's Account.

Suspending the accounts of Palestinian journalists also produces wider community effects. In Figure 26e, users respond to the suspension of journalist Faten Alwan by framing it as part of a repeated pattern of politically motivated enforcement against Palestinian voices. These reactions suggest that wrongful moderation does more than affect the individual account, but also disrupts broader networks of information circulation and witnessing.



“ Translation: “My journalist friend @fatanelwanii had her account shut down because of political posts (just like they did to me last year), and this is the new account... We say again, they shut down one account, we open a thousand!”

”

“ Translation: “Instagram shuts down journalist Faten Alwan’s account. They don’t want anyone talking about us. Faten Alwan presents the most important Palestinian journalistic content, live and loud, on Instagram. She drove them crazy for the umpteenth time, I’m done counting” ”

Figure 26e: Users Complaining About Faten Alwan’s Suspension

3.2.2. Focus Issue: Suspension of an Israeli activist

In our analysis of suspension cases in the dataset, we identified a notable pattern of suspensions affecting the same Israeli activist on four separate occasions. The two earliest entries, cases #788 and #1520, do not provide sufficient context to determine the specific reason for moderation. However, in case #357, the activist was suspended after posting content showing Israelis protesting against Zionism. As shown in Figure 27a, the protest content does not contain strong language or a hateful tone. The protesters are holding signs calling for an end to the forced starvation of Palestinians in Gaza, which they link to the IDF blockade of humanitarian aid and the continued siege of the area.

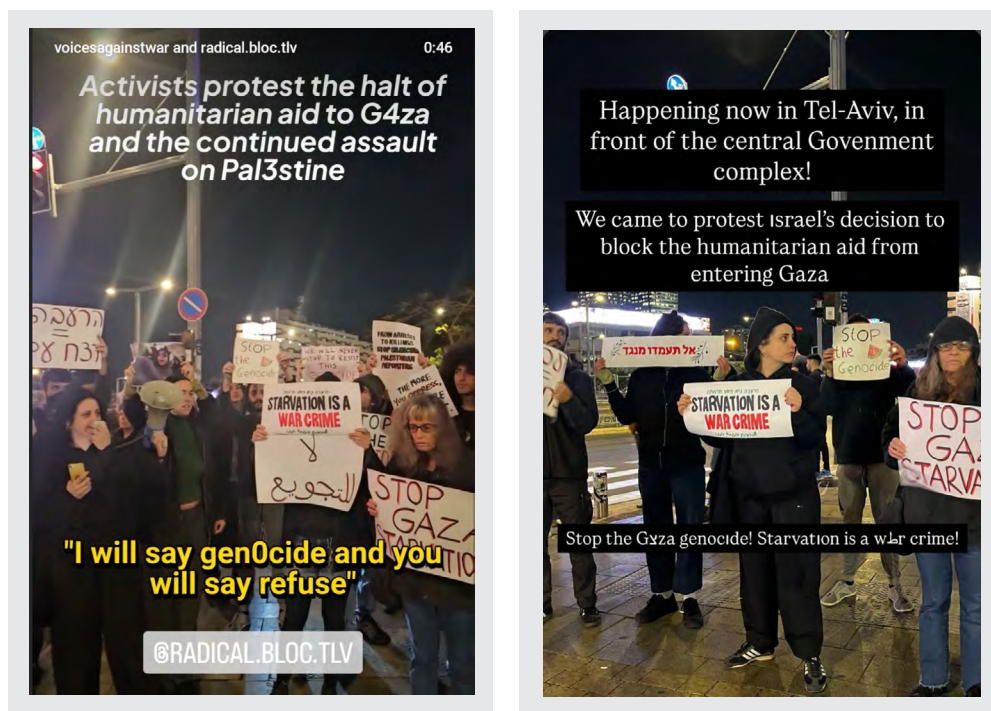
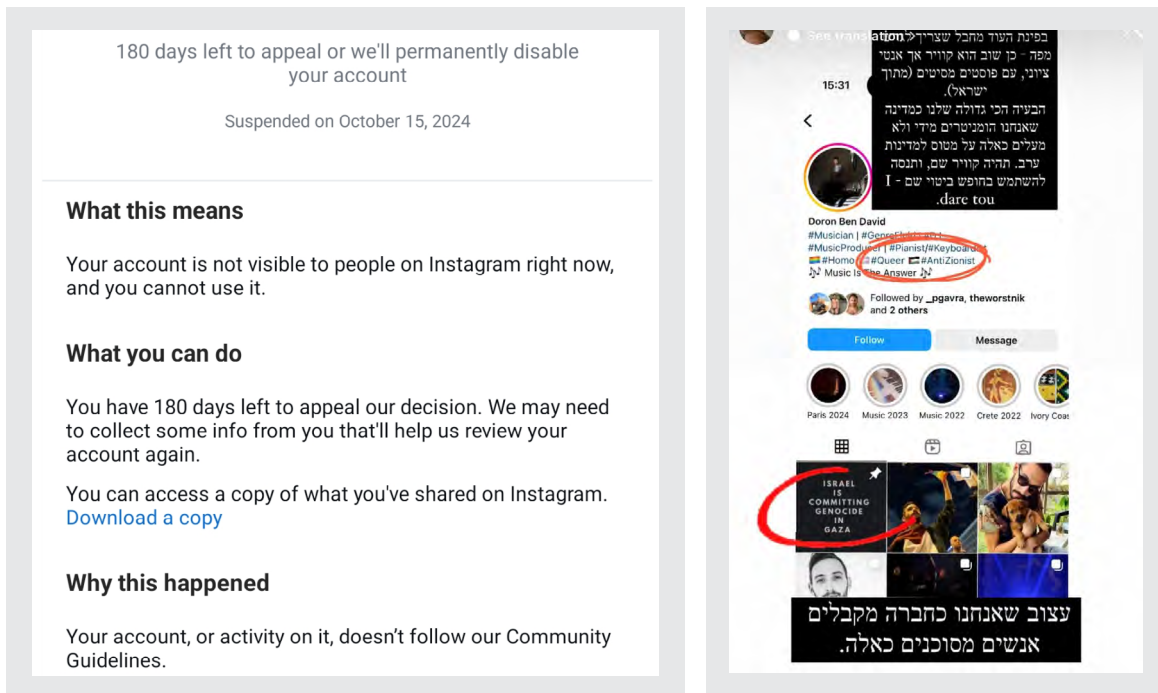


Figure 27a. Posts that Led to the Suspension of the User.

The account was suspended, and the appeal was denied. This case is informative because it shows how Meta’s enforcement of the DOI policy, when wrongfully applied through account suspension, can hinder civil society’s mobilisation efforts. The same activist appears again in case #516, where the suspension seems to have followed reports from other users. Based on the evidence submitted by this user through 7or, we could establish that other Israeli users targeted his account and flagged it for a DOI violation. As shown in Figure 27b, users shared the activist’s profile and encouraged others to report it. This case, therefore, highlights both Meta’s wrongful policy application and enforcement, but also the possibility that Meta’s reporting tools can be mobilised by coordinated users to suppress pro-Palestine-related mobilisation, also within Israeli civil society. This becomes even more problematic when considering Meta’s under-enforcement of Hebrew-language content, as established by a recent 7amleh’s report on the topic.⁶⁵



“ Translation: “In the corner of the Oud is a terrorist who needs to be removed - yes, again he is queer but anti-Zionist, with misleading posts (from Israel). Our biggest problem as a country is that we are too humanitarian and don’t put people on a plane to Arab countries. Be queer there, and try to use freedom of expression there” ”

Figure 27b. Suspension and Users Inciting to Flag the Activist’s Page.

65. 7amleh, Meta’s Role in Amplifying Harmful Content During the Genocide, 2 September 2025, available at: <https://7amleh.org/post/meta-s-role-during-genocide-en>

3.3. Restrictions

Our analysis indicates that restrictions are the second-most-used enforcement tool in the dataset. When compared to suspensions, restrictions are an enforcement tool of a different nature; they are tied to strikes received for community standard violations. Meta defines strikes for restriction as follows:

“For most violations on Facebook, strikes will lead to the following restrictions:

- » *One strike: You’ll get a warning, as this is your first strike.*
- » *Two to six strikes: You’ll be restricted from specific features, such as posting in groups, for a limited amount of time.*
- » *Seven strikes: You’ll get a one-day restriction from creating content, which includes posting, commenting, creating a Page and more.*
- » *Eight strikes: You’ll get a 3-day restriction from creating content.*
- » *Nine strikes: You’ll get a 7-day restriction from creating content.*
- » *Ten or more strikes: You’ll get a 30-day restriction from creating content.”*⁶⁶

Other types of restrictions include bans and restrictions on going live. The latter is not included in Meta’s dedicated page on restricted accounts. However, as of the 13th of October update for better moderation of content regarding the genocide, Meta introduced more stringent clauses on restrictions for going live. As Meta states:

“Facebook and Instagram Live: We recognise that the immediacy of Live brings unique challenges, so we have restrictions in place on the use of Live for people who have previously violated certain policies. We’re prioritizing livestream reports related to this crisis, above and beyond our existing prioritization of Live videos. “

This policy update is particularly relevant to restrictions on going live. Since the beginning of the genocide, such restrictions have increased and have been imposed for longer periods of time. As shown in Figure 28, restrictions on going live (n=235) disproportionately affected activists, who account for 67 cases (28.5%), and journalists, who account for 55 cases (23.4%).

66. Meta. “Restricting Accounts.” Enforcement. Updated November 12, 2024, <https://transparency.meta.com/enforcement/taking-action/restricting-accounts/>

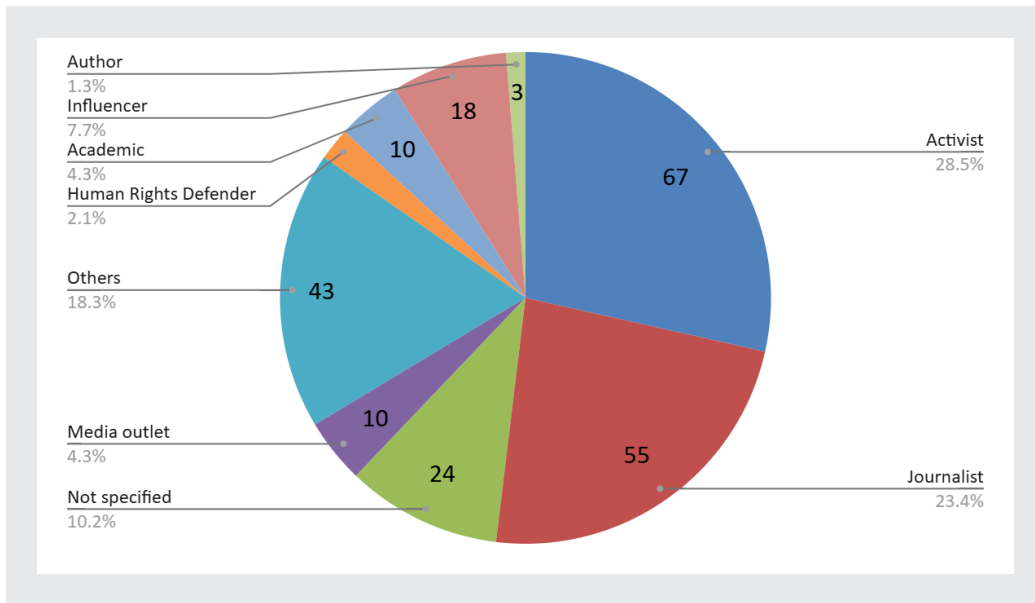


Figure 28. Distribution of Affected Classifications from Restrictions on Going Live.

Our analysis also indicates that restrictions on going live have a relatively high restoration rate, at 27.2% as shown in Figure 29. This indicates that Meta acknowledged its overenforcement of this measure. The ability to livestream is particularly important for the actor categories most affected in our dataset, as it enables users to report events in real time, document conditions on the ground, and raise awareness. Even when accounts or features are later restored, wrongful moderation can still have serious consequences. For journalists, in particular, the ability to communicate promptly is central to their work, especially when trying to report mass violence in real time.

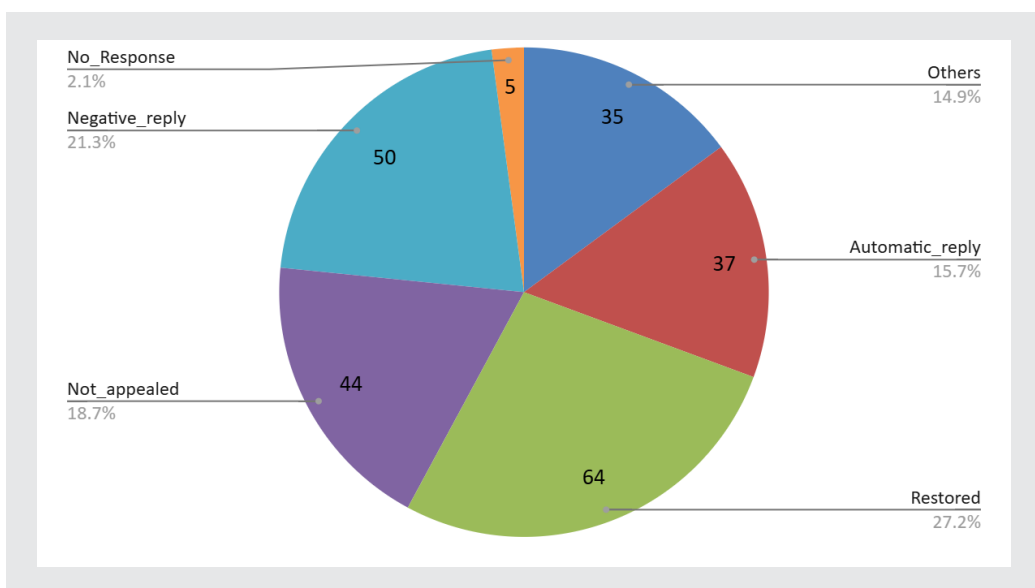
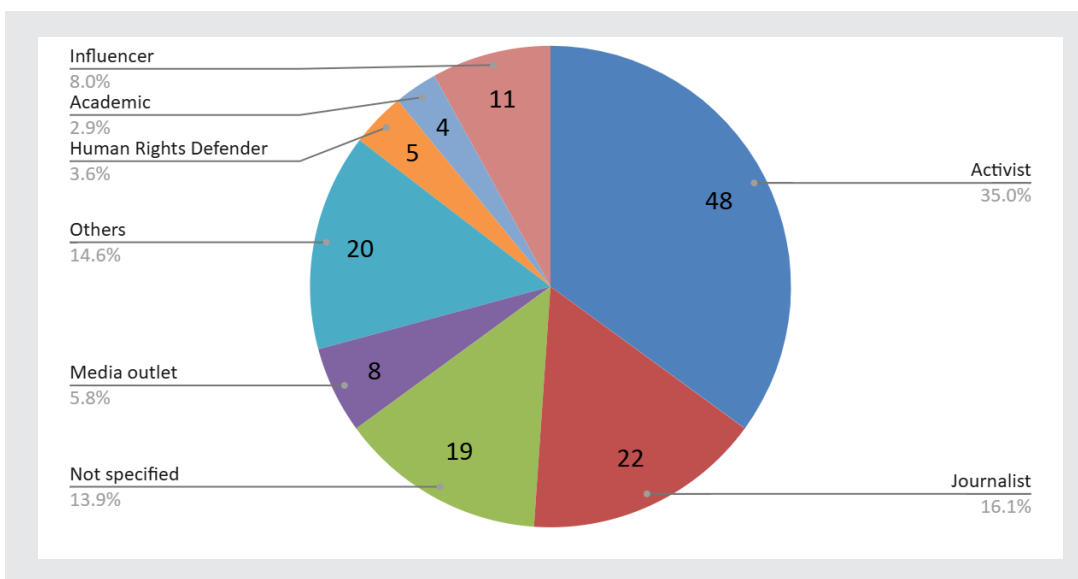
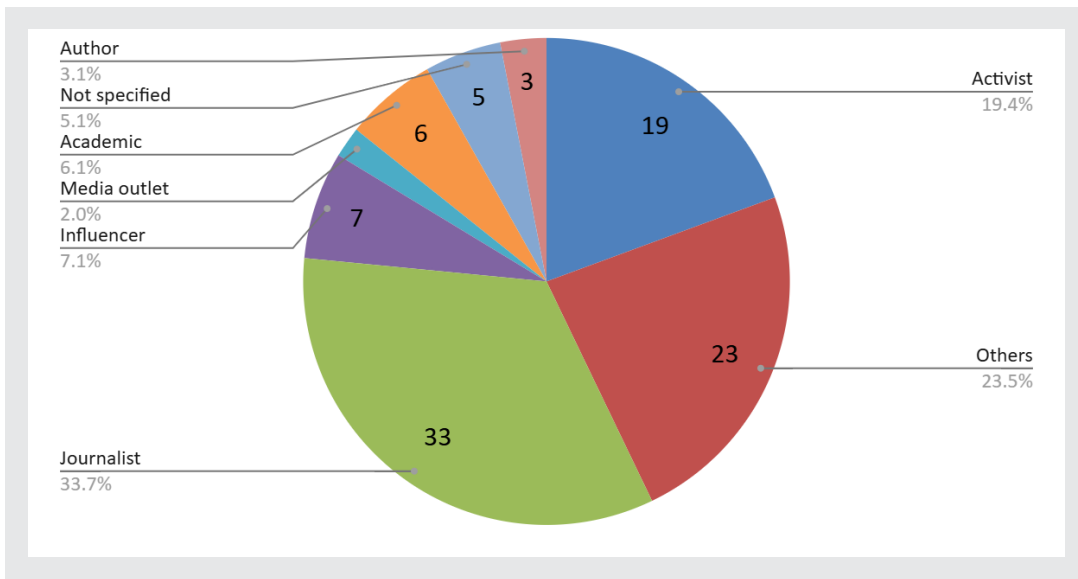


Figure 29. Status of Cases on Restrictions on Going Live.

3.3.1. Focus Issue: Restrictions on Going Live during the Genocide

When comparing moderation before and after 7 October, we observe an increase in the proportion of activists among the targeted actor categories, as shown in Figures 30 and 31. In light of the policy changes discussed above, this increase is significant. Activists were already disproportionately affected by wrongful moderation under the DOI policy. Through livestream restrictions, Meta added another layer of limitation on their ability to communicate and mobilise.



Figures 30 & 31. Distribution of Affected classifications from Going Live Pre and Post Oct. 7th, respectively.

The case #308 (Figure 32a) is a good example of how Meta uses restrictions on livestreaming to target activists.

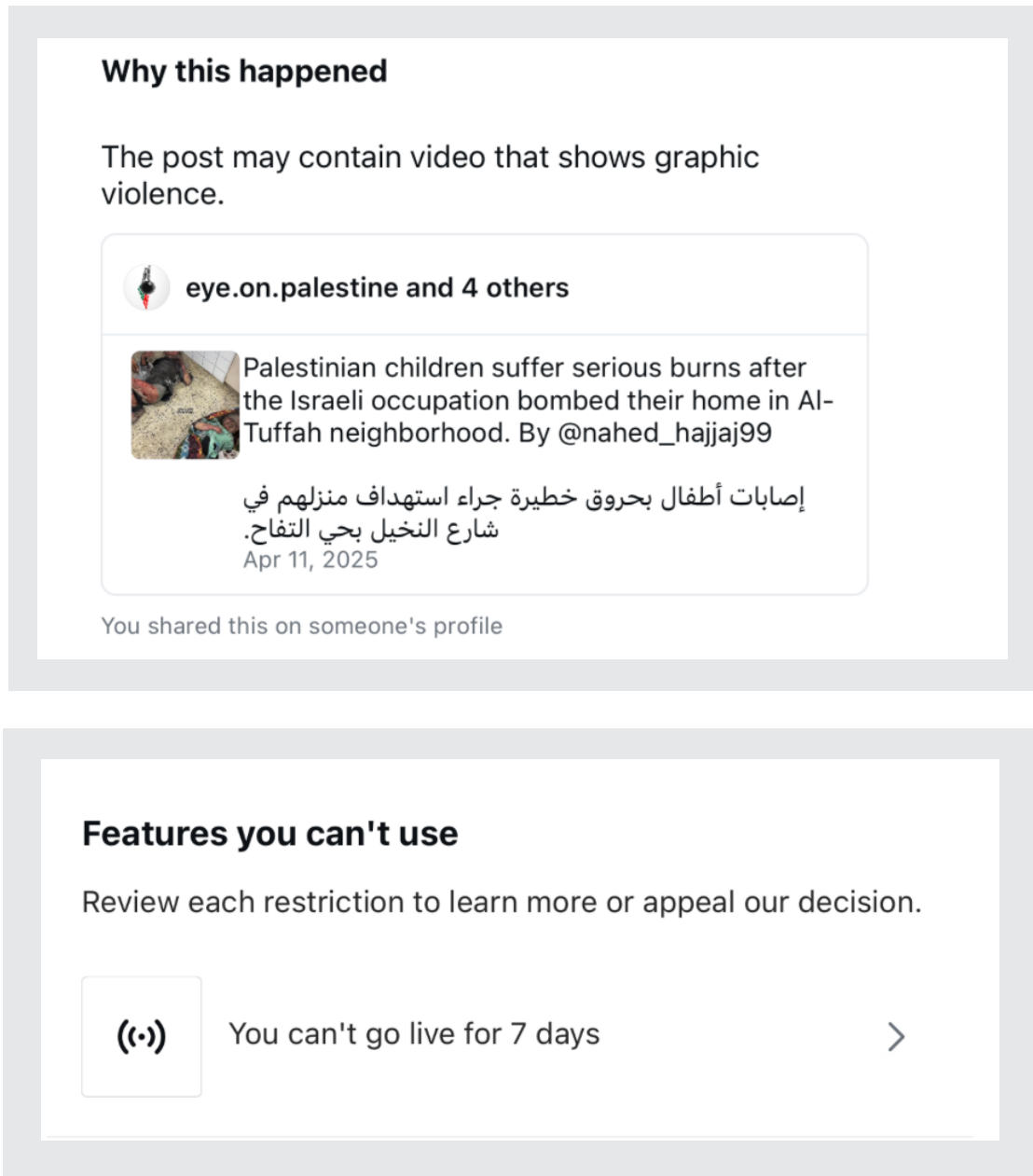
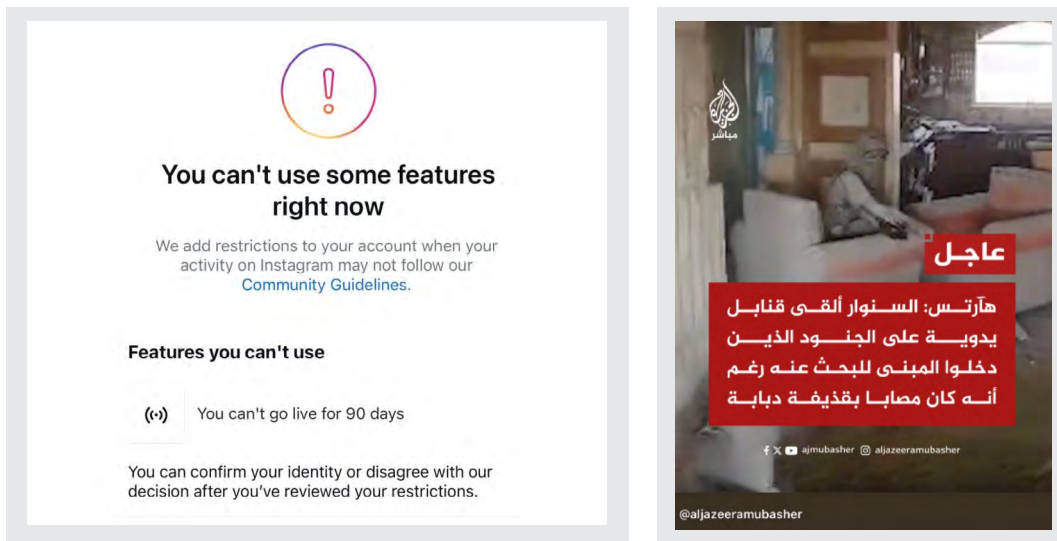


Figure 32a. An Activist Facing a Restriction On Going Live for Reporting on Israel's Bombing.

In the case above, the restriction on going live was later lifted after an appeal. However, even when access is eventually restored, the delay caused by a livestreaming restriction can seriously hinder a user's ability to engage audiences and report developments on the ground in Gaza in real time. Case #500, shown in Figure 32b, involves a journalist who was restricted from going live for an alleged violation of the DOI policy. As the screenshot reveals, we could not identify a clear basis for the specific policy application.



Translation:

“Haaretz: Sinwar threw hand grenades at soldiers who entered the building to search for him, even though he had been wounded by a tank shell”



Figure 32b. Another Example of a Restriction on Going Live for Reporting on Events Following Wrongful Moderation.

This restriction is particularly revealing of Meta's wrongful moderation, as the post in question simply cites reporting by the Israeli newspaper Haaretz. For a journalist in Palestine, a 90-day restriction on livestreaming can be especially harmful, as it limits their ability to reach audiences, maintain engagement, and report on Palestine-related developments in real time. Lastly, case #885, shown in Figure 32c, provides another example of Meta wrongly moderating an activist's account. The user received an 87-day restriction on going live, along with additional limits on content distribution and collaboration with other users. This combination of restrictions significantly reduced the account's capacity to engage with wider audiences.

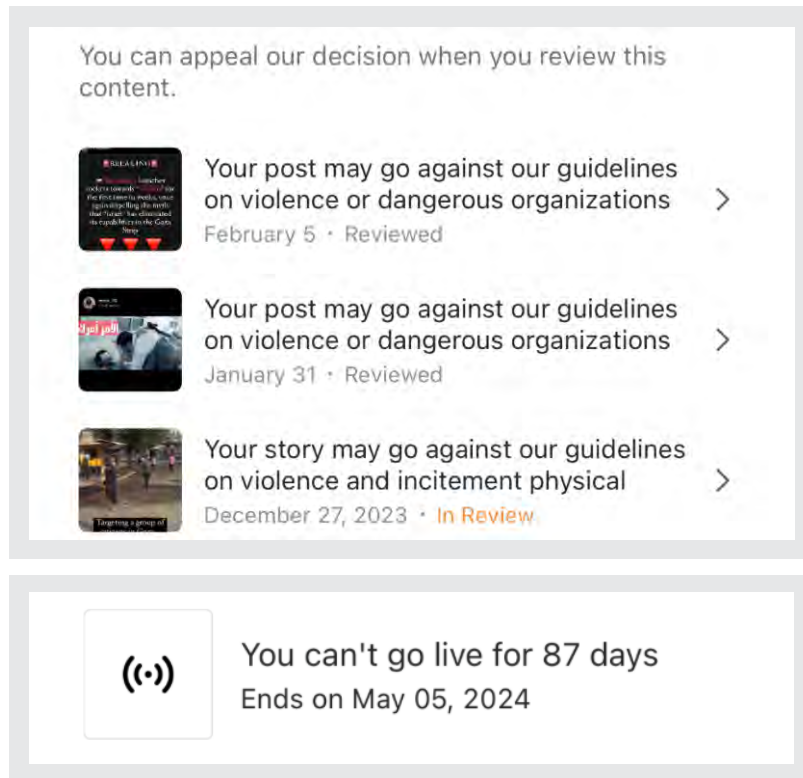


Figure 32c. Several Restrictions, Including on Going Live, Applied Following Wrongful Moderation of an Activist.

When examining the moderated posts more closely, both their content and Meta's communication are noteworthy. Meta states that the content may go against its Community Standards, rather than identifying a specific violation. These are highly intrusive moderation measures, as they directly affect the user's ability to communicate, reach audiences, and participate in public debate. For this reason, their application should require clear justification, transparent communication, and a meaningful explanation of the specific policy basis for the restriction.

3.4. Content Takedown

At 25%, content takedowns are Meta's third most used enforcement mechanism in the dataset. Content takedown refers to Meta's removal of posts that are deemed to violate its Community Standards. This enforcement tool may be applied through Meta's automated moderation systems or in response to user reports. Once removed, the post is no longer visible on the user's profile and cannot be accessed by other users. Compared with other enforcement mechanisms, Meta provides relatively limited information on how content takedowns are applied. It offers little detail on the criteria used, the types of content most likely to be affected, or the consequences for the user beyond the removal of the post itself. For these reasons, our analysis of how Meta used content takedowns is necessarily limited.

3.4.1 Focus Issue: The Palestinian flag

One of the patterns identified in relation to content takedowns pertains to the Palestinian flag, which triggers moderation enforcement in different ways: when posted as an image or emoji, both as content as well as part of a user's bio. Such variation of types of moderation enforcement on the Palestinian flag indicates that it is moderated in and of itself, even in the absence of other alleged violations of community policies. In case #63 below (Figure 33a), the user posted pictures of Palestinian women holding a Palestinian flag during their graduation ceremony, accompanied by the message

“We do not applaud ourselves while our people suffer. We ensure the world knows what Palestine is and what we are fighting for every chance we get.”

Instagram took down the post, moderated as a DOI with no justification. In our assessment, Meta's community policies were not violated, making the enforcement decision a clear case of wrongful moderation.

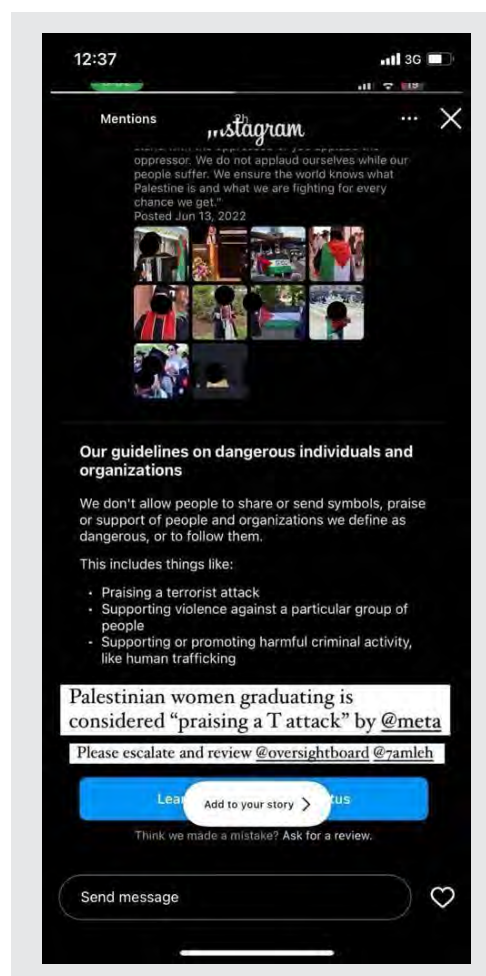


Figure 33a. A Post Displaying a Palestinian Flag During a Graduation Ceremony.

The presence of the Palestinian flag triggers moderation through content takedowns and leads to account suspensions and restrictions, also when users include it in their profile picture or bio. According to Meta's community standards, profile bios and other user-generated content must adhere to the platform's policies on hate speech, violence, and terrorism. However, flags are commonly included in bios on both Facebook and Instagram. In case #426 (Figure 33b) and #495 below (Figure 33c), the user's accounts were restricted because they included the Palestinian flag in their profile picture or bio, which Instagram assessed as a symbol of violence:

Figure 33b. A User Moderated Due to Their Profile Picture Being a Palestinian Flag

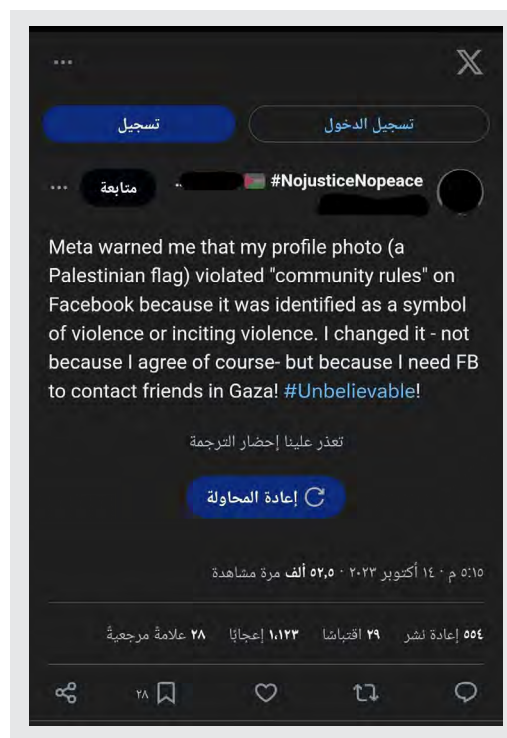


Figure 33c. A User Was Moderated Due to a Palestinian Flag in Their Bio.

Similarly, case #2322 shows a warning that leads to a restriction decision on generic grounds of noncompliance with the platform's standards. The post only included the emoji of the Palestinian flag and no text (Figure 33d):



Figure 33d. A Post That Contained Only a Palestinian Flag Was Moderated.

Meta's community standards do not explicitly ban the display of flags, whether they represent nations or political entities. These guidelines primarily focus on restricting content that promotes violence, hate speech, or terrorism. However, our analysis suggests a pattern of concern regarding the Palestinian flag. Moderation enforcement appears to be triggered not by the context in which the flag is displayed, but by the mere presence of the Palestinian flag itself. This indicates a potential bias in how the platform handles content related to Palestinian symbols, treating the flag as inherently problematic. Such practices raise important questions about fairness, consistency, and the exceptionalism of moderation that applies to Palestine.

3.5. Shadowbanning

Shadowbanning may be the most distinctive form of enforcement, both alongside and within other forms of moderation. Shadowbanning is an umbrella term for forms of enforcement mechanisms not made explicit to the user, for example, through reduced access or reach within a given platform.⁶⁷ Meta only publicly recognised the existence of this moderation mechanism in a September 2021 press release,⁶⁸ but stated that it fell under their wider remove, reduce and inform strategy they have been employing since 2016.⁶⁹ In practice, shadowbanning operates on the algorithmic distribution of specific content to followers and non-followers. Getting shadowbanned results in lower visibility, less engagement and being placed at the end of “for you pages”. According to their enforcement page, Meta uses shadowbanning for content that *“doesn’t violate the Community Standards, but might still be problematic or otherwise low quality.”*⁷⁰ Their process for determining what constitutes ‘problematic’ or ‘low quality’ is based on user activity and survey responses, with examples given as: “Low-quality content, such as clickbait and engagement bait, links to websites that are covered with ads, slow to load or broken or Low-quality comments that are repeatedly copied and pasted.” Furthermore, Meta states that, through shadowbanning, it aims to disincentivise publishers from creating content with limited originality, low-quality videos that abuse video formats, or misinformation.

The ambiguous application of shadowbanning is also reflected in our analysis of the dataset, particularly in the relatively small number of identifiable cases compared with the broad range of situations in which this enforcement tool could potentially be applied. For this reason, the fact that shadowbanning appears in only 58 cases, or 1.6% of the 3520 cases in the dataset, is more likely to reflect the difficulty of detecting this mechanism than its actual prevalence. Despite its low overall presence in the dataset, identifiable shadowbanning cases increased sharply after 7 October: from 2 cases, or 0.1% of the 1938 pre-7 October cases, to 56 cases, or 3.5% of the 1582 post-7 October cases. This represents a substantial increase in the cases where shadowbanning could be identified.

While the sharp increase in identifiable cases of shadowbanning is concerning in itself, it is made more troubling by the uneven distribution of these cases across different classification categories. As shown in Figure 34, activists constitute the largest group affected by

67. For a comprehensive overview of shadowbanning, see: Marten Risius and Kevin Marc Blasiak, “Shadowbanning,” Business & Information Systems Engineering, October 28, 2024, <https://doi.org/10.1007/s12599-024-00905-3>.

68. Meta. “Sharing Our Content Distribution Guidelines.” Newsroom. September 23, 2021, <https://about.fb.com/news/2021/09/content-distribution-guidelines/>.

69. Meta. “The Three-Part Recipe for Cleaning up Your News Feed.” Newsroom. May 22, 2018, <https://about.fb.com/2018/05/inside-feed-reduce-remove-inform/>.

70. Meta. “Reducing the distribution of problematic content.” Enforcement. Updated April 25, 2025, <https://transparency.meta.com/enforcement/taking-action/lowering-distribution-of-problematic-content/>.

shadowbanning, accounting for 25 of the 58 cases in the dataset, or 43.1%. This mirrors the disproportionate targeting of activists under the DOI policy discussed in Chapter 2.

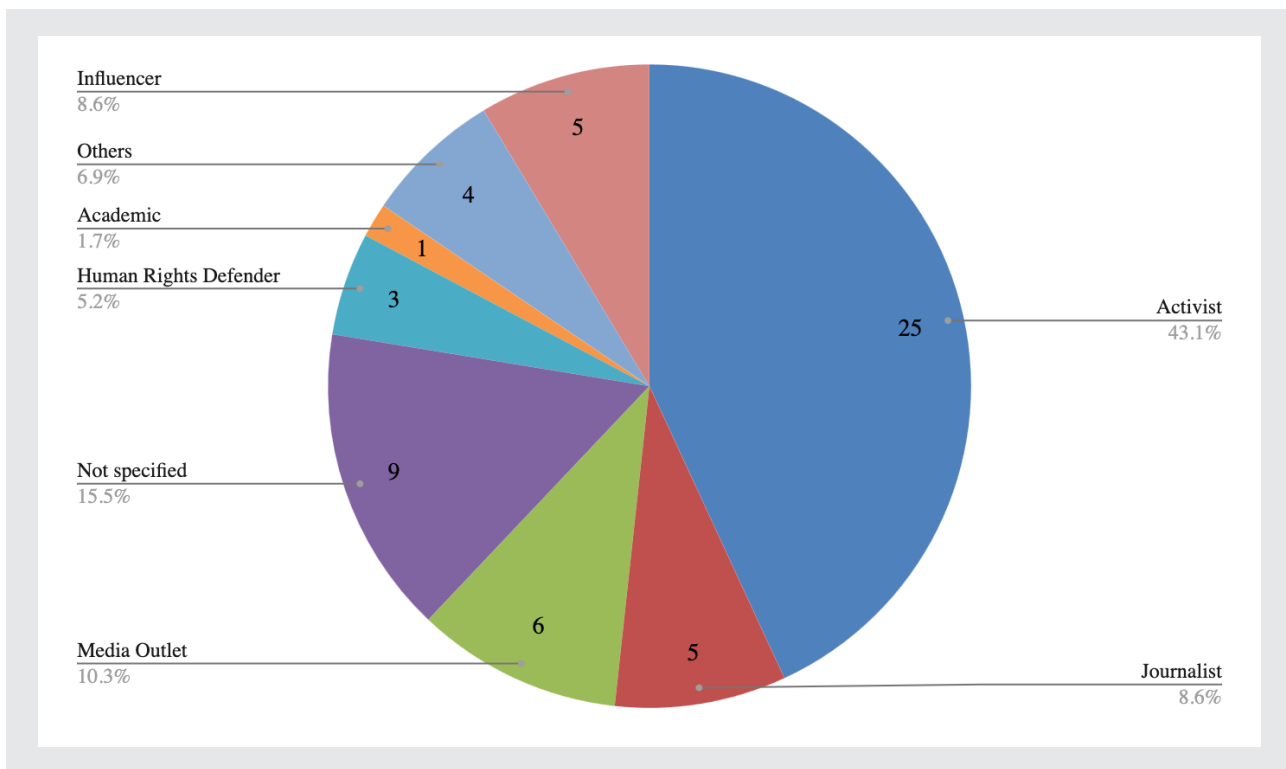


Figure 34. Distribution of Classifications in Shadowban Cases.

3.5.1. Focus Issue: Shadowbanning activists

As discussed above, shadowbanning is particularly problematic because it can be applied to content that does not meet the threshold for violating Meta's Community Standards. This creates a troubling combination: a broad, poorly defined enforcement mandate coupled with a difficult-to-detect form of enforcement. For these reasons, shadowbanning is best analysed through a qualitative assessment of the relevant cases in the database. Case #179, shown in Figure 35, provides a clear example. The user is a pro-Palestinian activist with a reasonably large following, yet this following is not reflected in the level of viewership their content receives.

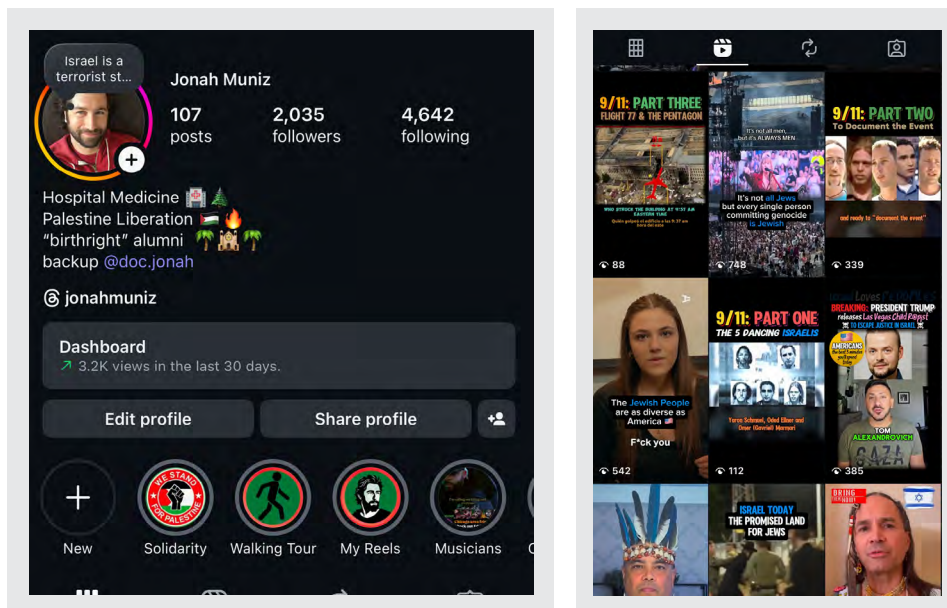


Figure 35. An Activist Critical Of Israel Is Being Shadowbanned.

Furthermore, case #305 (Figure 36) provides a clearer example of shadowbanning. With almost 70k followers, this Palestinian activist account shows a clear limitation of reach, given the staggeringly low view count on their stories. In this instance, the case was restored after appeal, thereby recognising the existence of the reduced distribution and reinforcing the wrongfulness of its application.

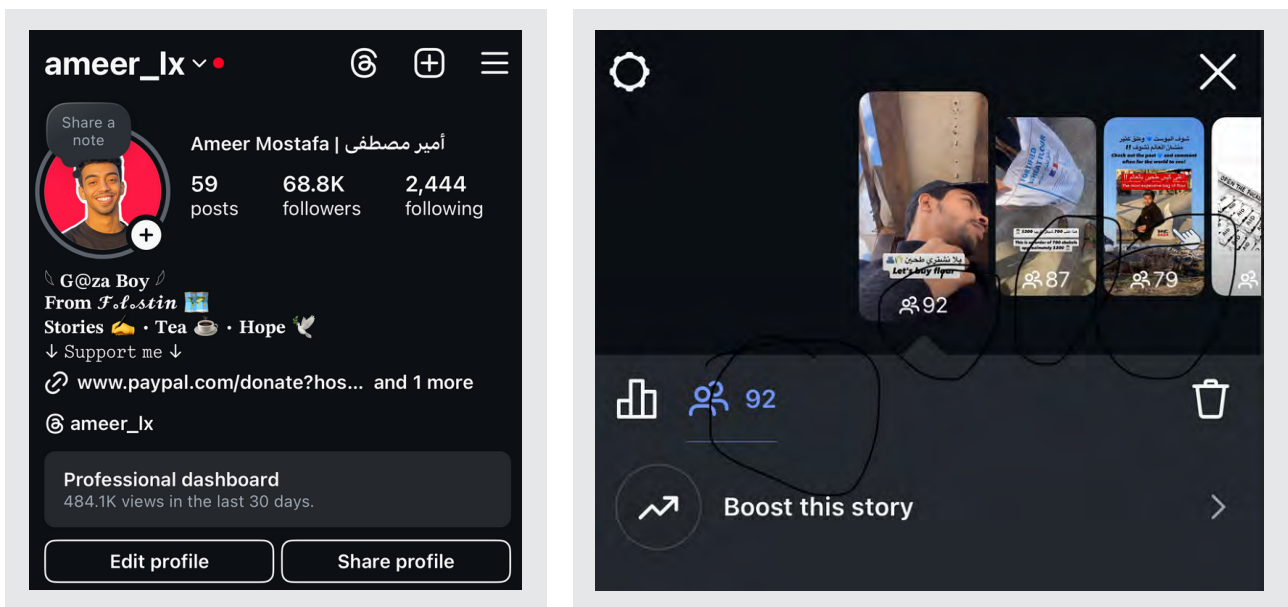


Figure 36. A Palestinian Activist With 68k Followers Receiving Very Low Story View Counts.

Similarly, case #331 (Figure 37) concerns another large Instagram account, with 214,000 followers and video content that has reached 12 million views. As the user interface indicates, Meta does not notify the user of any restriction or moderation action. However, the account's videos about Palestine show a sudden decline in both views and engagement, suggesting a visibility reduction that is not transparently communicated to the account holder.

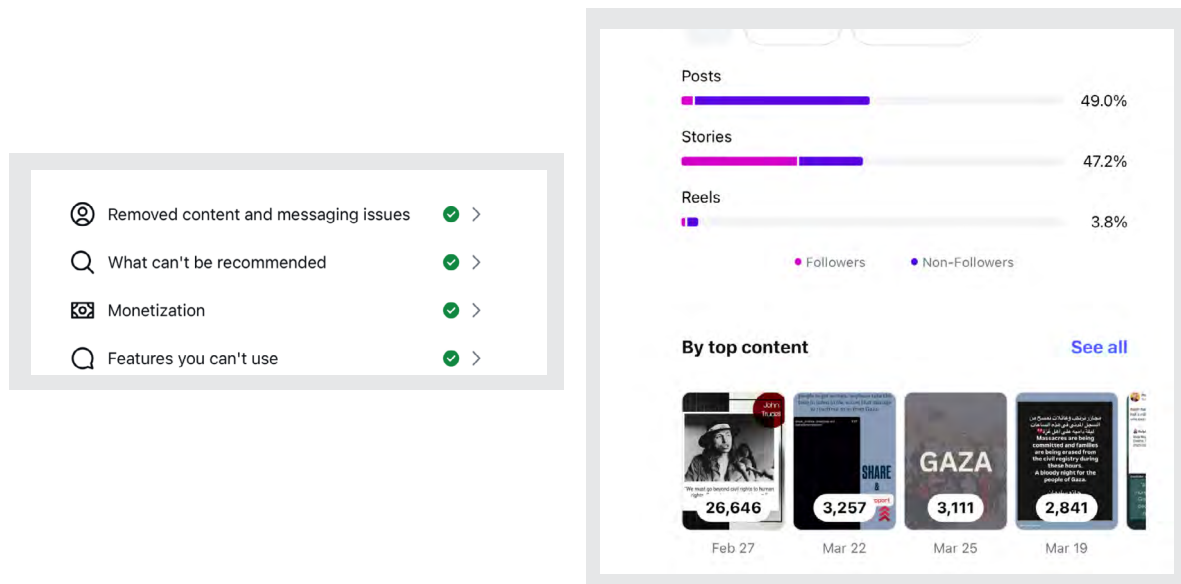


Figure 37. An Activist Receiving Unacknowledged Restrictions.

Our qualitative analysis of shadowbanning cases reveals a multi-layered system of policy application and algorithmic enforcement affecting Palestinian and Palestine-related activists. Users with substantial followings and potential reach appear to have their ability to communicate with both followers and sympathetic non-followers restricted, even when there is no clear indication of a violation. When considered alongside the previously discussed policies on reduced distribution of political content and the increased moderation of Palestine-related debates after 7 October, these cases help illustrate the broader scale of Meta's wrongful moderation. They suggest that Meta relies on a range of formal and informal enforcement tools that can limit the visibility of Palestinian content without necessarily removing their content or issuing clear violation notices.

4. Meta's Communication

This chapter focuses on Meta's responses and communication practices, scrutinising the clarity and informativeness of its explanations regarding the application of its policies and moderation enforcement, also in relation to appeals submitted by 7amleh following users' submissions. As mentioned in the introduction, 7amleh is part of Meta's Trusted Partner Programme, a preferential channel for appeals designed to assess the impact of Meta's platforms in local contexts. While this might sound promising, in the following section, we highlight how Meta either fails or inadequately engages with 7amleh's appeals, indicating an overall lack of communication between affected stakeholders and the platform in contradiction with the company's mission to promote stakeholder engagement.

4.1. (Not) Responding to 7amleh

Our analysis shows that 7amleh submitted 2828 appeals to Meta between 2021 and 2025. In 1355 cases, representing 48.1% of all submissions, Meta did not provide a response (see Table 9). This high rate of non-response points to a significant gap in Meta's communication with 7amleh. This lack of responsiveness is particularly concerning in light of Meta's stated commitment to stakeholder engagement as part of its efforts to improve content moderation. It also sits uneasily with Meta's broader commitment to "voice" and its stated ambition to work with external organisations in developing more transparent and inclusive content policies, including through its Content Policy Stakeholder Engagement processes.⁷¹

71. Meta. "Guide for conducting inclusive stakeholder engagement", Updated September, 2024.

Table 9. Distribution of the type of responses provided by Meta following the appeal

Status	Number of Entries	Proportion of Total Entries (%)
No Communication From Meta	1355	48.1
Negative Reply (Policy Mentioned)	390	13.8
Negative Reply (No Policy Mention)	274	9.7
Active and Unrestricted	273	9.6
Incorrectly Removed	262	9.2
Content Reported to the Wrong Channel	100	3.5
Incorrectly restricted	52	1.8
Meta Requested More Information	49	1.7
Delays in responding (Covid-19)	17	0.6
Prioritising cases(limited resources)	15	0.5
Suspicion of Accounts Being Hacked	14	0.5
Other	27	1.0
Total	2828¹	100

1 Does not include cases not appealed by Tamleh (n=692)

In 664 cases (23.5%), Meta responded and upheld the original moderation decision. However, a specific policy was referenced in only 390 of these cases, accounting for just 13.8%. In the remaining instances where the decision was upheld, Meta’s responses lacked any informative details or explanation of the basis for the decision, leaving significant gaps in transparency and clarity. In these cases, the response by Meta would read as the one in the following example (case #135):

“Hi, Thank you for bringing this content to our attention. We have checked the account in question and confirmed that it indeed goes against our policies. Therefore, we are taking no action against it. You can learn more about our Community Standards.”

This lack of detail creates a twofold problem. First, the user receives no meaningful feedback that would allow them to understand the reason for the restriction or eventually adjust their behaviour on the platform to avoid future moderation. Second, Meta retains broad discretion to impose restrictions without demonstrating consistency with its own policy standards. This absence of transparency therefore contributes to an uneven moderation

environment in which users affected by enforcement decisions have limited ability to understand enforcement, and eventually appeal.

4.1.1. Focus Issue: 'Active and Unrestricted': Targeting Influencers

Our analysis of Meta's communication practices also revealed a significant lack of accuracy and transparency regarding account restrictions. Specifically, in 9.6% of cases (273 instances), Meta indicated that the user's account was active and unrestricted. Upon reviewing the cases where content was available, we found that users had posted that their accounts remained restricted. This discrepancy primarily involves instances in which 7amleh submitted an appeal to Meta and received a response stating that the account in question was fully active and that any restrictions had been lifted. As shown in case #21, the notifications of 'active and unrestricted' are similarly generic and lack specific context about the enforcement decision:

“Hi, Thank you for your report. The account appears to be active and unrestricted. Do not hesitate to let us know if you have any questions. Kind regards, Meta Team”

Despite this communication from Meta, users often report ongoing restrictions, prompting them to approach 7amleh to document their experiences of wrongful moderation and communication. In many of these cases, users actively seek support from others to regain full access to their accounts, highlighting the frustration and confusion caused by Meta's inconsistent communication. Examples include (# case):

690: Support my account pleaseee!! My account is restricted. I will not forget this favour of supporting me

691: Support me pleaseee!! my account is restricted

692: My account was restricted on facebook. I'm not allowed to post or comment, I can only use messenger

694: AISalamuAlikum, please support the account, it has multiple reports against it and only now I was able to retrieve the account back

Our analysis further indicates that inaccurate notifications about restrictions appear to affect influencers the most. This would indicate that the lack of informativeness disproportionately affects Palestinian users and content with high reach, regardless of the content targeted by moderation. In a case of shadowbanning already discussed in Chapter 3 (case #331, Figure 37), the activist user had received numerous restrictions, prompting them to initiate a review of

the moderation with Meta's assistance chat. The screenshots of these conversations were uploaded under case #627. In these conversations, the user lamented the takedown of their posts, the reduced distribution of their content and the inability to collaborate on posts. At the beginning of the exchange, shown in Figure 38a, the user reports the takedown of their content and states that they could not identify the basis for Meta's moderation decision. They describe repeated wrongful moderation as obstructing their advocacy work. Although the support assistant acknowledges the importance of this advocacy, it reframes the decision as part of a broader effort to reduce offensive material, without assessing whether the specific content was fairly moderated. When the user asks how long the profile restrictions will remain active, Meta's response remains vague, offering neither a clear timeframe nor actionable information about the basis or duration of the enforcement measure.

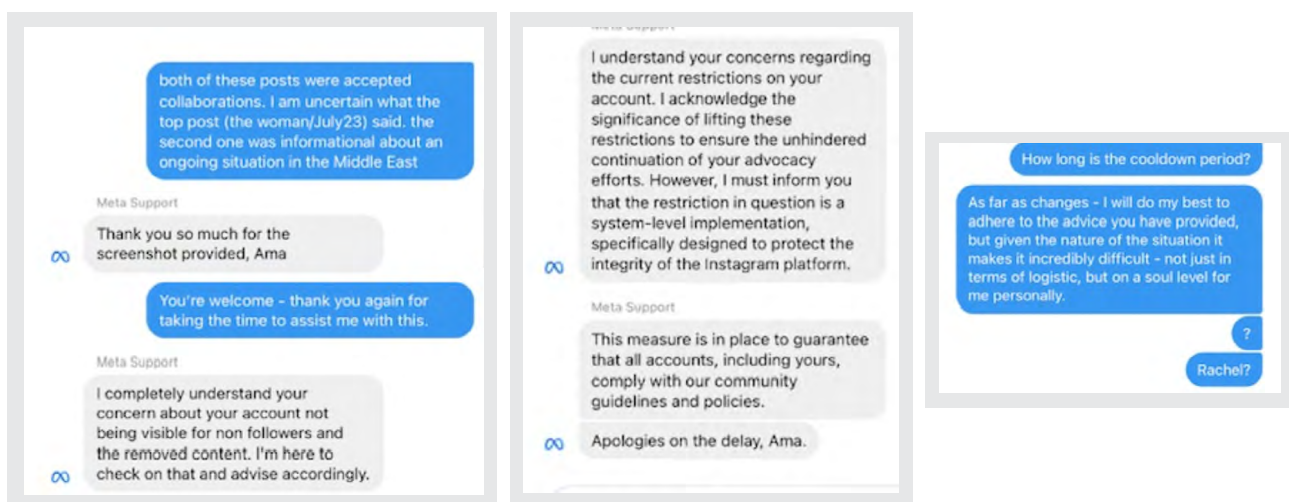


Figure 38a. Activists' Conversation with Meta's Assistance Chat (1).

In the other screenshots available, the user contacts support regarding a separate restriction. In this instance, the user explains that the collaboration restriction was not listed as active on their profile, already pointing to a gap in Meta's communication about the enforcement measures applied to the account. Meta's support response offers little substantive clarification. The user is told "to be patient", on the basis that they have already appealed other cases, and is directed to the Community Standards pages for further information. When the user asks again how long the restriction will remain in place, Meta's support does not provide a timeframe or any case-specific explanation. This exchange illustrates the limited usefulness of Meta's support process for users subject to restrictions. Rather than clarifying the basis, scope, and duration of the enforcement measure, the response leaves the user without meaningful information about how the restriction operates or when it may be lifted.

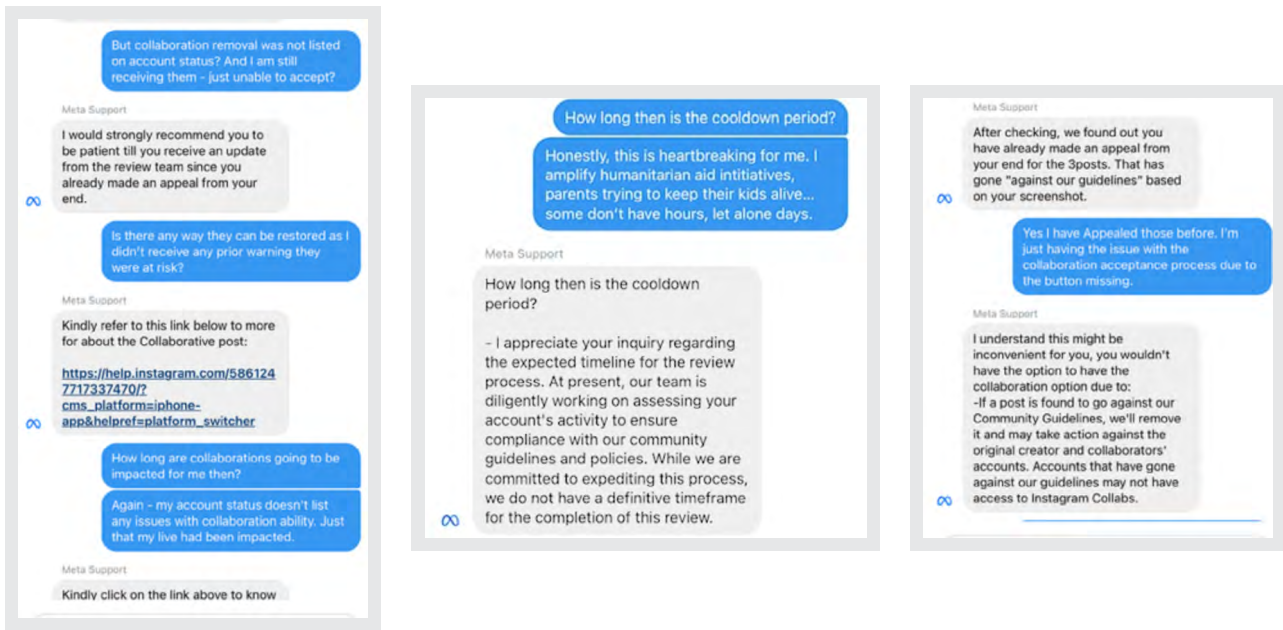


Figure 38b. Activists' Conversation with Meta's Assistance Chat (2).

Overall, the screenshots in Figure 38 point to two recurring problems in Meta's communication with users. In Figure 38a, the user requests an explanation and the lifting of restrictions, while also explaining that they had already reviewed their own content and could not identify a violation. Meta's response does not provide a clear account of the reason for the moderation decision or the basis for maintaining the restriction. This lack of communication limits the user's ability to understand the enforcement action and avoid similar moderation in the future. In Figure 38b, the problem is even more pronounced. The user's requests and appeals appear to receive no substantive response. This is particularly significant because the user reports receiving no prior communication about the moderation action and no meaningful explanation after appealing it. Taken together, these cases show the weakness of Meta's communication process in instances of alleged wrongful moderation.

Case #1331, shown in Figure 39, is another significant instance of poor communication, in this case related to a content takedown. The case concerns a video that had reached twelve million views before being removed. According to the user, Meta provided no communication explaining the decision, and the post was simply restored after 24 hours. This case is important because it shows how even temporary takedowns can affect high-reach content. The absence of an explanation also prevents the user from understanding the basis for the moderation decision or assessing whether the removal was consistent with Meta's Community Standards.

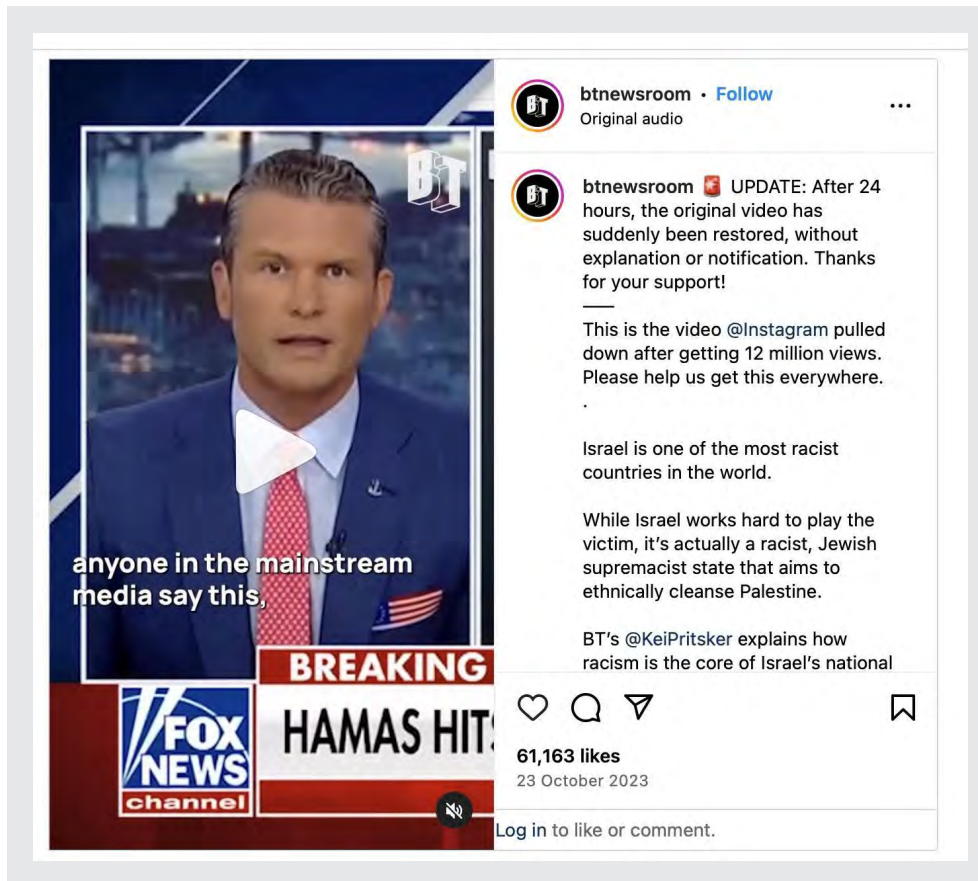


Figure 39. An Account Received A 24-Hour Video Takedown

In other cases, Meta indicates the policy applied but does not specify which content it pertains to. This is especially problematic for influencers, who are frequently targeted by coordinated mass reporting from anti-Palestinian users, sometimes covering content from years past. This makes it difficult for users to substantiate their appeals to Meta. Case #399 is one such example:

“Facebook says that I had a post which falls under dangerous individuals and organizations and that it was taken down, but it does not even show me the content of it. Feels like interfering with my personal business for nothing”

4.1.2. Focus Issue: The US Campaign for Palestinian Rights

In our analysis, we identified one account that reported perceived wrongful moderation by Meta on twelve separate occasions. This case is particularly useful because it brings together several dimensions of Meta's wrongful moderation and poor communication practices. The user in question is the Facebook page of the US Campaign for Palestinian Rights (USCPR), an NGO based in the United States that advocates for Palestinian rights. The organisation works to raise awareness of human rights violations in Palestine, supports the BDS campaign, and contributes to the organisation and mobilisation of pro-Palestinian movements in the US. It includes both Palestinian American and Jewish American members, and has collaborated with organisations such as Jewish Voice for Peace and other Palestine solidarity groups. USCPR has also been subject to political pressure beyond Meta's platforms, including restrictions imposed by Israel on the organisation and its activists.⁷² This broader context is important for understanding the impact of Meta's repeated moderation of the page. Across the dataset, the USCPR page appears to have been affected by several different enforcement tools and policy categories. In case #912, USCPR was restricted from posting and was found to be logged out of their Facebook account. The only explanation given on the user's interface, and after the appeal, is shown in Figure 40a. As we can read, there is no reference to the policy applied nor to the type of restriction to which the user is subjected.

72. US Campaign for Palestinian Rights, "USCPR Statement on Israeli Blacklist of Palestinian Rights Organizations," 7 January 2018, <https://uscpr.org/uscpr-statement-israeli-blacklist-palestinian-rights-organizations/>. The statement notes that USCPR was included on an Israeli Strategic Affairs Ministry blacklist of organisations whose members could be denied entry into Palestine/Israel because of their support for BDS.

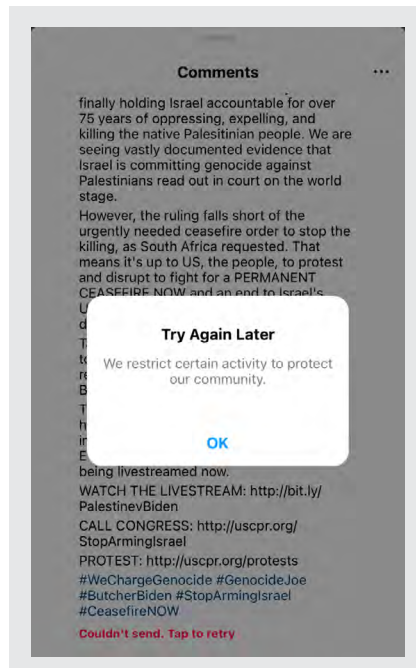


Figure 40a. Communication from Meta Regarding the User's Restriction.

As the USCPR appealed, Meta restored the account and responded:

“Hi, Thank you for your report. Meta’s automated systems identify and block suspicious behaviour, temporarily suspend one or more features on your account, and place limits and caps on certain activity. We do this to protect the site from abuse.”

Meta’s response provides no clarity about which policy was applied, which account features were restricted, or why the restriction was imposed. This lack of clear communication makes it more difficult for the user to understand the basis of the moderation decision and to avoid similar restrictions in the future. Similarly, in cases #913 and #914, USCPR reported being logged out of its Instagram profile without receiving any explanation. In case #994, USCPR uploaded a series of screenshots showing posts that had been taken down for allegedly violating Meta’s Community Standards on spam. This enforcement action was most likely triggered by the presence of external links in the posts. As shown in Figure 40b, the organisation included a set of links in each moderated post, directing users to USCPR initiatives. These included a link to email Congress calling for an end to US support for Israel, a call-in campaign to Congress, and a list of protests taking place across the United States.



Figure 40b: Meta’s Communication and Example of Moderated USCPR Post.

All of these posts were later restored, indicating that the external links they contained were wrongfully moderated as spam. Although the posts were eventually restored, the temporary removals still had meaningful consequences as USCPR’s work depends on timely publication for its advocacy and mobilisation objectives. Meta’s poor communication becomes even more concerning when considering another case submitted by USCPR, case #1001, shown in Figure 40c. In this case, the organisation posted about a week of peaceful protests it had organised in Chicago. These posts were also removed without a clear explanation.

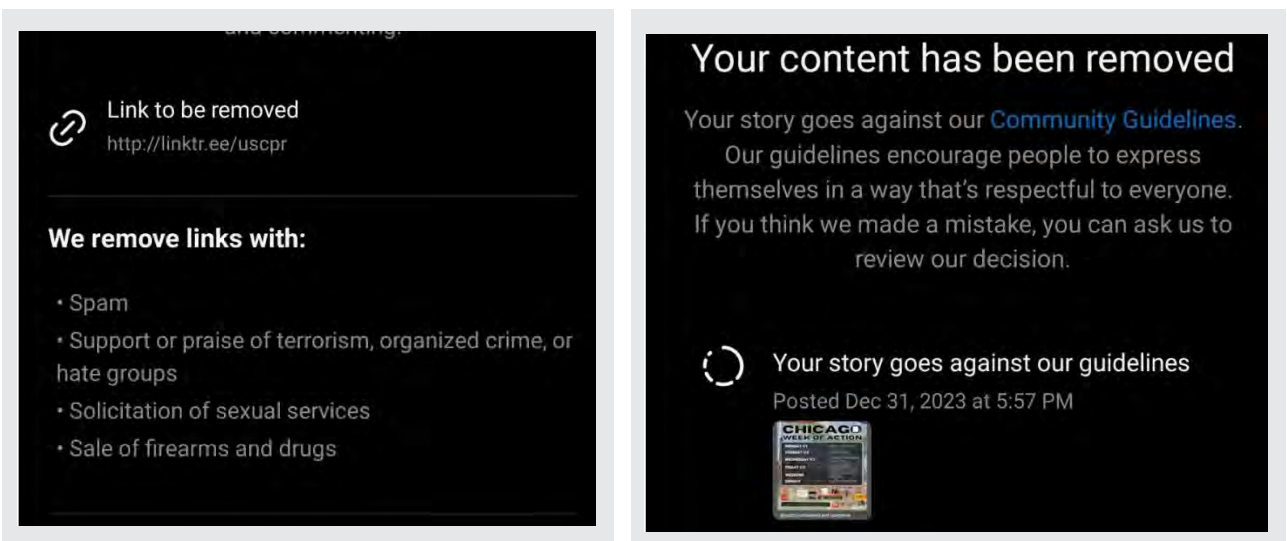


Figure 40c. Confusing Communication Regarding Takedown of USCPR’s Post.

In none of the communications to USCPR does Meta clearly state which policy the link allegedly violates. The link directed users to the organisation’s official website. USCPR is a registered NGO with legal status in the United States and is a widely recognised civil society organisation. USCPR’s posts were repeatedly removed, and the organisation’s account was restricted several times, apparently for sharing links to its official website. Because Meta did not clearly explain why the links were treated as problematic, USCPR could not adjust its posting practices or understand how to avoid future moderation. This resulted in significant

restrictions on the page's content and platform activity and, ultimately, its mobilisation efforts. The repeated nature of these removals suggests that Meta failed to communicate its enforcement process to the organisation, even after multiple previous instances of wrongful moderation.

4.2. Restored with No Explanation

The extent of Meta's lack of communication becomes even more concerning when considering that, following 7amleh's escalation, moderation enforcement was reversed in 926 cases, representing 26.3% of the entire dataset (see Figure 41 & Table 10).

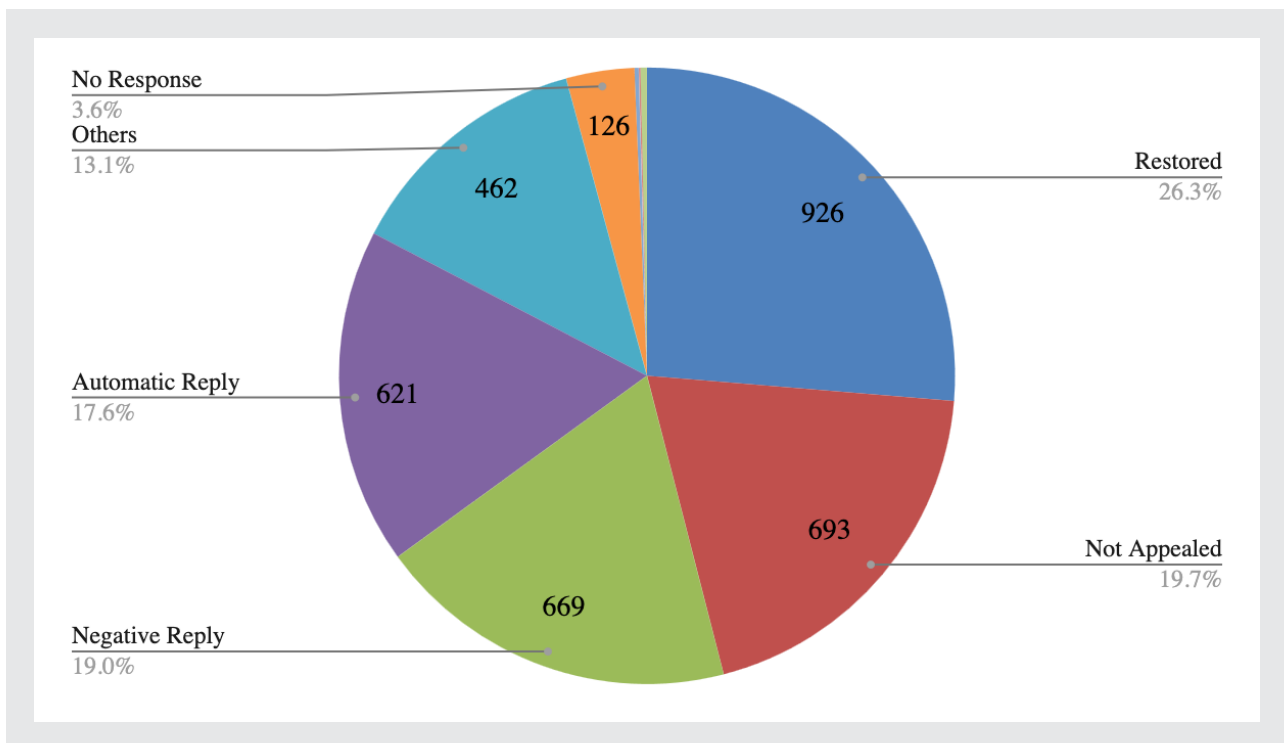


Figure 41. Distribution of Cases' Statuses after Follow-Up from 7amleh.

Table 10. Status of cases after follow-up from 7amleh

Status	Number of entries	Proportion of Total Entries (%)
Restored	927	32.8
Negative Reply	673	23.8
Automatic Reply	621	22.0
Others	469	16.6
No Response	126	4.5
Removed	8	0.3
Secured	3	0
In Review	1	0
Total	2828	100

Does not include cases not appealed by 7amleh (n=692)

Of these 927 restored cases, our analysis identified 619 (66.8% of the total restored) instances in which Meta provided vague or unclear communication about why the decision was reversed, including statements such as *“We have checked the post in question and confirmed that it indeed goes against our Community Standards.”* These responses contain no mention of which policy or guidelines were breached. In comparison, only 308 (33.2%) included a response from Meta explaining why the account or content was restored (see Figure 42). These included cases of incorrectly removed or restricted content/accounts, cases where hacking was suspected, and cases where Meta contacted a user or secured their account. This means that the vast majority of restored cases lacked any explanation, leaving the reasoning behind the initial takedown and subsequent reversal unclear.

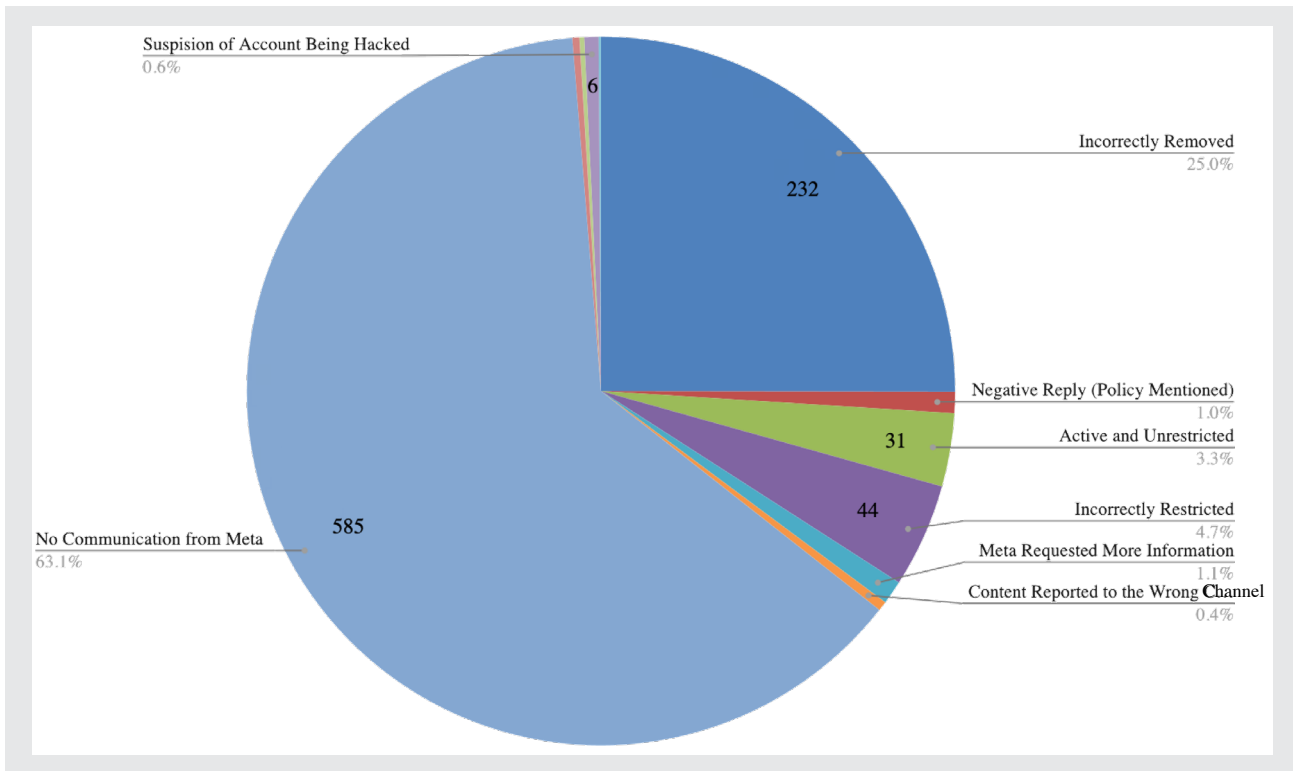


Figure 42. Distribution of Meta's Communication in Restored Cases.

This indicates that when Meta restores content or profiles after an appeal, it fails to provide any justification for the reversal. With no clear information provided either when moderation actions are enforced or when appeals are successful, our analysis indicates that Palestinian users are left in the dark about the reasons behind decisions that affect their presence on Facebook and Instagram. This lack of transparency in communication leaves users with minimal understanding of the platform's moderation processes and the criteria applied to their cases.

4.3. Meta's Responses After October 7th

Moving into an examination of how Meta's responses and communication evolved after October 7, our analysis reveals that, while the overall substantial lack of responsiveness and clear communication persisted, there was an increase in cases wrongfully marked as active and unrestricted (see Table 11).

Table 11. Status of cases after follow-up from 7amleh before vs after October 7th

Response	Before Oct 07th		After Oct 07th	
	Number of Entries	Proportion of Total Entries (%)	Number of Entries	Proportion of Total Entries (%)
No Communication from Meta	1013	58.8	342	31.8
Negative Reply (Policy Mentioned)	247	14.3	143	13.3
Negative reply (No Policy Mentioned)	199	11.5	75	7
Active and Unrestricted	172	10.0	101	9.4
Meta Requested More Information	32	1.9	17	1.6
Delays in responding (Covid19)	17	1.0	-	-
Incorrectly Removed	14	0.8	248	23.0
Prioritising cases (limited resources)	14	0.8	1	0.1
Content Reported to the Wrong Channel	13	0.8	87	8.1
Incorrectly Restricted	3	0.2	49	4.5
Suspicion of Account Being Hacked	-	-	14	1.3
Total	1724	100	1077	100

Does not include cases not appealed by 7amleh (n=692) and other cases (n=27), which were categorised under 9 other labels.

One way in which this shift can be explained is shadowbanning: users' accounts appear active and unrestricted, yet face hidden restrictions that limit their visibility and engagement without explicit notification. Similarly interesting, our analysis points to an increase in cases that Meta recognised as 'incorrectly removed', from 0.8% to 23.0%. This clearly indicates less accurate and more indiscriminate moderation enforcement after October 7, also because of an increased reliance on algorithms (see Section 2.3.2). Turning to restored cases, our analysis shows a sharp increase in Meta's reverted moderation decisions after October 7. Before October 7, only 17 cases, or 2.2% of cases involving clear communication from Meta

following an appeal, resulted in restored content or accounts. After October 7, this figure rose to 330 cases, or 44.2% of comparable cases (Table 12).

Table 12. Meta's communication in response to appeals before vs after October 7th

Status	Before Oct 07th		After Oct 07th	
	Number of Entries	Proportion of Total Entries (%)	Number of Entries	Proportion of Total Entries (%)
Negative Reply	446	59.2	209	28.0
Others	255	33.8	183	24.5
Automatic Reply	24	3.2	23	3.1
Restored	17	2.2	330	44.2
No Response	12	1.6	-	-
Removed	-	-	2	0.3
Total	754	100	747	100

In tandem with this, the proportion of cases in which the user received communication from Meta, which our qualitative assessment deemed wrongful, was significantly larger after October 7 than the equivalent before this date (Table 13 & Figure 43). These two results indicate that Meta has become increasingly inaccurate in their moderation of Palestinian content during the genocide, and this has coincided with an increase in restored content, which indicates Meta's acknowledgement that its initial moderation decision was wrongful.

Table 13. Assessment of moderation before and after October 7th, in case of communication from Meta

Reason for Assessment	Before Oct 07th		After Oct 07th	
	Number of Entries	Proportion of Total Entries (%)	Number of Entries	Proportion of Total Entries (%)
Fair Moderation	44	5.8	43	5.8
Unfair moderation	155	20.6	385	51.5
Cannot Assess	555	73.6	319	42.7
Total	754	100	747	100

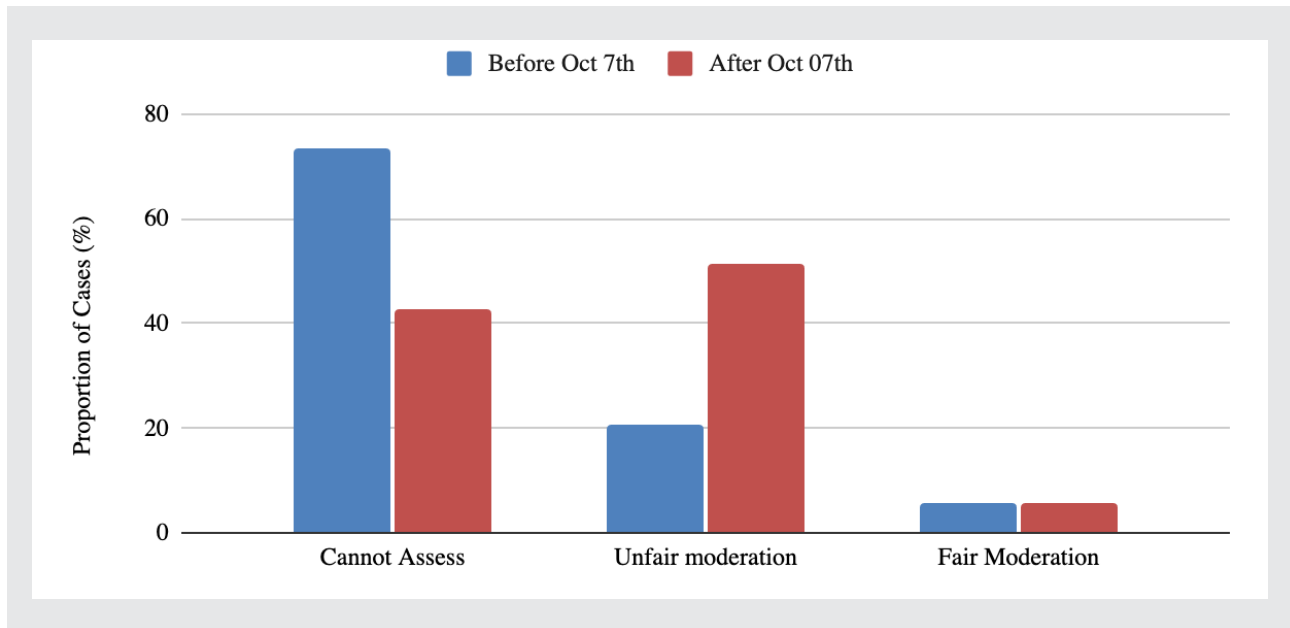


Figure 43. Distribution of Assessments of Cases with Communication from Meta Pre & Post Oct 7th.

5. Conclusions

This report has examined Meta’s moderation of Palestinian and Palestine-related content on Facebook and Instagram between January 2021 and December 2025. Based on 3520 cases submitted through 7amleh’s Palestinian Observatory for Digital Rights Violations, the analysis has focused on three dimensions of platform governance: the application of Meta’s policies, the enforcement mechanisms used against users and content, and Meta’s communication with affected users and 7amleh. Across these areas, the findings point to a recurring pattern: Palestinian content is frequently governed through opaque policy application, severe account-level enforcement, limited explanation, delayed remedy, and reduced visibility. Taken together, these practices contribute to what this report has conceptualised as the platformicide of Palestine: the narrowing and erasure of Palestinian digital presence, participation, and visibility on Meta’s platforms.

5.1 Summary of findings

The first major finding concerns the opacity of Meta’s policy application. Of the 3520 cases analysed, only 1144, or 32.5%, contained an identifiable policy that was allegedly violated. In some cases, this information was provided by Meta. In others, the research team had to infer the relevant policy from the content, the user submission, and Meta’s Community Standards. This means that in most cases, users were not given sufficient information to understand which rule they had allegedly breached, which content had triggered enforcement, or how the decision had been reached. Among cases with an identifiable policy, the DOI policy was by far the most frequently applied. DOI accounted for 654 cases, or 57.2 %, of all identifiable policy applications. This is significant because the DOI is a policy designed to address terrorism, organised violence, and serious offline harm, yet in this dataset, it was repeatedly applied to all users’ classifications, including journalists, media outlets, and ordinary users. This means that the policy appears to operate as a broad security lens through which many forms of Palestinian expression are assessed. In particular, reporting and documentation, or even mourning, are treated by Meta as if they were indistinguishable from praise or support for dangerous organisations.

This problem is especially clear in relation to Palestinian journalism and media work. The report found that 128 out of 180 journalist cases with an identifiable policy, or 71.1%, were moderated under DOI. Among media outlets, 102 out of 172 identifiable cases, or 59.3%, were moderated under the same policy. These figures are particularly concerning because Meta’s own standards recognise the importance of news reporting through the so-called “newsworthiness allowance”. The qualitative analysis shows that many of these cases

involved ordinary forms of journalistic practice, including the quotation of public statements, reporting on events, and documentation of unfolding violence.

The second major set of findings concerns enforcement. Across the dataset, suspensions were the most common enforcement mechanism, accounting for 1,187 cases, or 33.7%. This is significant because it shows how Meta's moderation enforcement primarily relied on account-level penalties through suspensions, restrictions, reduced visibility/shadowbanning, and blocking live broadcasting. In particular, reduced visibility and shadowbanning can make content difficult to find while leaving users without a clear explanation of what has happened. These enforcement practices had particular consequences for public-interest users and accounts, such as journalists and activists. The report also found that less visible forms of enforcement became more prominent after 7 October 2023. Shadowbanning cases increased from 2 cases before 7 October to 56 cases after that date. Because shadowbanning and reduced distribution are difficult to detect, these numbers should be interpreted cautiously, as the scale of the problem might be larger. This is especially important because Meta lowered thresholds for algorithmic intervention after 7 October in order to reduce the recommendation of potentially violating or borderline content. In a context where Palestinian content is already vulnerable to wrongful enforcement, lower thresholds increase the risk that lawful and public-interest content will be made less visible.

The third major finding concerns Meta's communication with users and with 7amleh. Out of 2800 appeals submitted by 7amleh between 2021 and 2025, Meta provided no response in 1363 cases, or 48.7%. Where Meta did respond, its explanations were often generic or incomplete. In 664 cases, Meta upheld the original moderation decision, yet only 390 of these cases included a specific policy reference. In a further 274 cases, Meta upheld the decision without identifying the relevant policy. This leaves affected users in a difficult position: they are told that they violated the rules, but not which rule, how, or why. The report also highlighted a problematic pattern related to restorations. Restoration practices also reveal a significant accountability gap. Following escalation by 7amleh, Meta reversed enforcement decisions in 926 cases, representing 26.3% of the dataset. Yet in 642 of these restored cases, or 69.5%, Meta provided vague or unclear communication about why the decision was reversed. Only 284 restored cases included an explanation. Restoration without explanation limits the value of the remedy. It may return content or accounts, but it does not tell users why the original decision was wrong, whether the error was technical or policy-based, or what will be done to prevent similar errors in the future. In time-sensitive contexts such as a genocide, restoration can also come too late.

The post-7 October period is particularly important across all three dimensions of the report.

Meta's communication during this period shows a sharp increase in recognised moderation errors. Before 7 October, Meta recognised 14 cases, or 0.8%, as incorrectly removed. After 7 October, this increased to 248 cases, or 23%. Incorrectly restricted accounts also increased from 3 cases before 7 October to 49 cases after that date. Restored cases increased from 17 cases before 7 October to 330 cases after it. These figures suggest that Meta's moderation became more error-prone at the very moment when Palestinian content was most urgent.

5.2 Policy Recommendations

Taken together, these findings show that the platformicide of Palestine is the result of a systematic approach aimed at making Palestinian content and users less visible across platforms. Meta's own governance architecture contains safeguards that should protect freedom of expression. The evidence presented in this report shows that these safeguards have often failed Palestinian users and Palestine-related content. On the basis of this evidence, and in consultation with the research team, 7amleh calls on Meta to take the following actions:

- 1) Commit to conducting Heightened Human Rights Due Diligence of Meta's content moderation practices in Israel and Palestine throughout the genocide, including during and following periods of escalation, with meaningful participation of right-holders, affected communities, and regional civil society organisations. In light of Meta's well documented failures in this context, the company should also undertake an independent, transparent, and public audit of its human rights due diligence procedures and content moderation systems relating to Israel and Palestine since October 7th, 2023.
- 2) Overhaul the Dangerous Organisations and Individuals (DOI) policy to include proportionality safeguards for journalists, civil society groups, researchers, and human rights defenders, in order to prevent discriminatory impacts on political expression and documentation of human rights violations. Furthermore, publish the DOI list in full to ensure transparency and public accountability.
- 3) Strengthen and operationalise newsworthiness protections, especially for content concerning mass violence, atrocity crimes, and urgent public interest.
- 4) Restrict the use of severe enforcement measures against journalists, media outlets, and documentation-focused accounts unless clear, necessary, and proportionate justification thresholds are met, based on transparent and publicly available criteria that respects freedom of the press and the public interest role of documenting human rights violations. This requires a more consistent application of the newsworthiness allowance and related

public-interest exceptions. Palestinian journalists and media outlets should not be penalised for reporting on events involving designated actors. Where content is removed and later restored, Meta should provide a clear explanation and assess whether the delay has caused additional harm. In urgent reporting contexts, the remedy should be rapid enough to preserve the value of the content.

5) Continue to increase resourcing to improve Arabic-language and context-sensitive moderation to reduce wrongful enforcement against Palestinian content and other Arabic-speaking users across its platforms. The moderation of terms such as “shaheed” shows the risk of treating Arabic words or Palestinian cultural expressions as fixed indicators of dangerous content. Meta should develop reviewer guidance that reflects Arabic linguistic diversity, Palestinian political and cultural context, and the difference between mourning, documentation, and endorsement of violence. This guidance should be developed with independent experts and affected communities, and its effectiveness should be tested through regular audits.

6) Review and reverse the decision to extend hate speech policy protections to the term “Zionist” in ways that shield political ideology from legitimate criticism. Political ideologies are not considered protected characteristics under international law or normative hate speech frameworks, and affected communities must retain the ability to speak out against violent political movements attacking their community.

7) Provide full transparency and detailed explanations around content takedowns, reduced distribution, shadowbanning, recommendation suppression, demonetization measures, and account suspension, including clear disclosure of what enforcement measures were applied and why. Furthermore, provide meaningful explanations when content or accounts are restored, including acknowledgement of wrongful enforcement where applicable.

8) Increase resourcing dedicated to crisis-related and time-sensitive escalation mechanisms relating to journalism, human rights documentation, and civil society reporting.

9) Provide increased disclosure and transparency regarding all external takedown requests, including detailed reporting on voluntary government requests and state-linked escalation channels seeking content removals related to Israel and Palestine.

10) Improve the quality and timeliness of appeal communication. Affected users and trusted partners should receive a response that identifies the relevant policy, the content at issue, the reason for the decision, and the outcome of any review. Where Meta reverses a decision,

it should explain why the original decision was wrong or excessive. This is also essential for trusted partners such as 7amleh, whose ability to monitor and contest wrongful moderation depends on access to meaningful information.

11) The burden of accountability should not fall primarily on affected users and trusted partners; Meta must take responsibility for harms produced and amplified by its systems.

5.3 Future research

This report has focused primarily on Meta's own moderation architecture. This focus has been important because the findings show that Meta often fails to apply its own rules consistently, transparently, and with sufficient contextual awareness. Future work should build on this analysis by placing these findings within broader accountability frameworks. Further research should examine Meta's responsibilities under the UN Guiding Principles on Business and Human Rights, especially the requirements of heightened human rights due diligence in conflict-affected settings, meaningful stakeholder engagement, prevention and mitigation of adverse human rights impacts, and effective remedy. The BSR human rights due diligence report on Meta's impacts in Israel and Palestine after May 2021 provides an important reference point because it shows that many of the relevant risks were already known to Meta before the period examined here. Future research should therefore assess whether Meta responded to those risks in a sustained, measurable, and rights-respecting manner.

Further work should also assess Meta's practices against international freedom of expression standards, including ICCPR Article 19 and the principles of legality, legitimacy, necessity, and proportionality. These standards are especially relevant where content moderation affects journalism and public-interest reporting during armed conflict and mass violence. When relevant, there is also a need to connect Meta's platformicide of Palestine to emerging regulatory frameworks, including the EU's Digital Services Act, data protection law, and algorithmic accountability debates. More research is needed to also compare Meta's treatment of Palestinian content with its treatment of other conflict-affected contexts. Such a comparison would help determine which patterns are specific to Palestine and which reflect wider weaknesses in platform governance during mass violence situations. However, comparison should not dilute the specificity of the Palestinian case.

Future research should also examine the implications of content moderation for accountability processes. Content shared on Meta's platforms may constitute important evidence of potential human rights violations and international crimes, including material

relevant to ongoing investigations by national and international authorities. The removal, suppression, or loss of such content may therefore have consequences that extend beyond freedom of expression, potentially undermining efforts to document violations, establish responsibility, and preserve evidence for future accountability mechanisms. Finally, the evidence presented in this report shows that Meta's moderation of Palestinian content should be understood as a structural accountability problem. By framing Meta's systemic silencing of Palestinian content and users as platformicide, this report also ultimately points to the importance and urgency of holding Meta accountable.

Contact Us:

info@7amleh.org | www.7amleh.org

Our Social Media Pages: 7amleh

