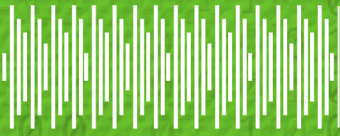




Briefing on

The Palestinian Digital Rights Situation



since
October 7th, 2023



حملة - المركز العربي
لتطوير الإعلام الاجتماعي
7amleh - The Arab Center for
the Advancement of Social Media



7amleh - The Arab Center for the Advancement of Social Media, has been following up on the rapidly escalating violations **of Palestinian Digital Rights**. Recent events have triggered a disturbing surge in the targeting of content posted by Palestinians and advocates for Palestinian human rights across various online platforms. The 7amleh team has [diligently documented these digital rights violations, and the data we have collected is deeply alarming](#).

Online hate speech and incitement to violence against Palestinians, predominantly in Hebrew, continue to proliferate across online platforms, notably on X and Telegram. We are acutely aware of the direct correlation between online content and real-world harm to Palestinians, with past instances, [such as the attack on the village of Huwara last February](#), serving as stark reminders of this perilous connection. Hate speech and calls for violence also contribute to the chilling effect.

Furthermore, we are actively identifying and documenting [instances of disinformation and misinformation](#), which significantly impact freedom of expression and access to information, as well as the right to security. False information is being weaponized to propagate hatred and incite violence against Palestinians. It is also used to manipulate the opinions of global citizens and divert their attention from the reality on the ground. This disinformation is employed to rationalise collective punishment against all Palestinians, and is often accompanied by incitement and calls for violence, escalating the danger for them both online and offline.

We are also gravely concerned about the significant and disproportionate censorship of Palestinian voices, which severely restricts freedom of expression and access to information, more direly so with journalists and human rights defenders. This is compounded by the one-sided and context-lacking portrayal of the region's situation by mainstream media. Residents of Gaza face additional hurdles in sharing and receiving vital information as [Israel, on October 27th, severed the last remaining, already limited and intermittent, lines of communication](#), cutting the people of Gaza off from each other, from emergency assistance, and from the world. 7amleh is closely monitoring the blackout in Gaza, a severe human rights violation in its own right. The lack of power and destruction of ICT infrastructures in the Strip not only disrupt the residents' communications but also deepen collective anxiety. It further hinders outsiders from understanding the situation within the besieged territory.

All these violations underscore the pressing need for immediate action from duty-bearers, specially tech companies, to protect both users and other rights-holders, especially in light of the current mainstream frameworks and political pressures that contribute to the dehumanisation of Palestinians. 7amleh, [together with 90 civil society organisations from around the world, launched a statement reminding online platforms of the responsibilities to uphold Palestinian digital rights during times of crisis](#).

Companies' approaches

Digital platforms, especially social media platforms, have been a site of repeated rights violations since the beginning of this crisis. 7amleh reiterates that based on the United Nations Guiding Principles on Business and Human Rights, companies have an obligation to ensure that their business practices do not contribute to adverse human rights impacts. When companies do contribute to adverse human rights impacts, as is this case with the current crisis in the region, companies also have an obligation to create appropriate remedies and make sure their platforms are safe for all.

Since October 7, the Palestinian Observatory of Digital Rights Violations ([7or](#)) had recorded 1009 verified violations of Palestinian digital rights. The number includes 411 cases of censorship, and 598 cases of hate speech, incitement to violence, or other forms of online harassment.

7amleh's newly developed AI-powered Language Model monitored the spread of hate speech and violence in Hebrew against Palestinians and their advocates on social media platforms since October 6. The real-time [Violence Indicator](#) documented over 590,000 model-classified cases of hate speech and violent content across platforms, majority of which were found on X.

With regard to companies role in the ongoing crisis, here is our assessment of some of the major firms' reactions:

1.Meta

Unfortunately, Meta's social media platforms have been a site for fairly widespread Palestinian digital rights violations. [7amleh and partners have documented](#) Meta's continual censoring and silencing of Palestinian voices and narratives for years, and even as the company has pledged to make improvements, there has still been a continuation of old harmful and discriminatory trends with Meta. Palestinian voices are censored, while hate speech and incitement to violence in Hebrew spread across the platforms. Since the beginning of the crisis, 7amleh documented manually through ([7or](#)) 627 violations across all Meta's platforms. This has included 344 cases of incitement to violence or hate speech and 283 cases of censorship.

The 283 censorship violations on Meta's Instagram, Facebook, and whatsapp are mainly Restrictions and Content Takedown, however, 7amleh also notes there have been cases where pages were completely taken down and users were suspended and not allowed to appeal.

7amleh observes the silencing of Palestinian voices in several formats. Many users use the term Shadow-banning, however, 7amleh observes that different forms of restrictions by the company apply to different users, and are disproportionately affecting Palestinians. Some trend observations include:

- Low views on stories by influencers/their stories pushed to the end automatically because they share news content from other pages (shadow banning).
- Meta [changed the default settings](#) to all users in the region from public to [‘friends only’](#) to limit the amount of people who can read and comment on public posts.
- Comment settings change: only people who are [friends or followers](#) of the page [for more than 24 hours](#) can comment.
- [Inability to comment](#) due to unspecified activities “to protect our community”.
- The [translation](#) of [Palestinian to ‘Palestinian terrorist’ bug](#), which they [claim to have fixed](#) but without providing an explanation.
- [Palestinian flag emoji](#) is often considered a [negative/harmful symbol](#), and [is automatically hidden](#) (not deleted) on posts.
- The “bug” that removed content showing [Palestinian victims in the hospital bombing, under “nudity” violation](#).
- Similarly worded content in Arabic gets overly moderated and mistakenly considered a violation, while content in English is unaffected ([start from 1:22](#)).
- hashtags like [#למחוקאתעזה](#) (Erase Gaza) are [still active on instagram and are not censored](#) even though they are an obvious incitement to violence with real-world consequences.
- In contrast, Meta immediately censored the [Arabic hashtag #طوفان_الاقصى](#) on the first day of the escalation, and never censored the parallel hashtag in Hebrew [#חבובת_ברזל](#) because it was not seen to violate their policies.

In many cases, the reason Palestinian content gets over-moderated, and discrepancies are documented in implementation of company policies is because the company [“lowered the threshold”](#) from needing an 80% confidence to now only a 25% level of confidence on Palestinian content due to “larger volume of violating content” in Arabic. In other cases such as the translation ‘bug’, Meta used the word machine “hallucinations” as it apologized for the error, but did not address [bias in underlying training data](#) for this machine.

Meta’s newsroom updates also addressed several “bugs” or “glitches” which affected the reachability of Palestinian content. While the company ‘apologized’ for said glitches, we see this as a continuation of the systemic bias which was documented in the past. The repeated mistakes which disproportionately affect the same users over and over again puts into question the company’s seriousness in investing in and deploying effective models that take into account the Arabic, Palestinian, and contextual experiences necessary to prevent such glitches in the first place.

The disproportionate over-moderation leads to restrictions limiting the reach of Palestinian content, and in some cases, completely suspending users (including journalists, activists and human rights defenders) on the platforms. Notably, this happened to the Facebook account for [Quds News Network](#) (شبكة القدس الإخبارية), an independent news agency, with nearly 10 million followers. On the same day the page was unpublished, an account affiliated with the Israeli Foreign Ministry wrote [“Thank you, Meta”](#), responding to Mark Zuckerberg’s three days old statement. The removal of the page did not make sense within the parameters of [Meta’s Dangerous Organizations and Individuals Policy](#) (DOI), as the journalists running it had already been in contact with Meta’s teams and understood the implications of the DOI policy. According to journalists and observers, the content on that page was of news-worthy nature, similar to that published by Aljazeera and other outlets in Arabic, and saw that Quds News Network was unfairly singled out.

Similarly, there have been many cases where journalists' accounts were [restricted](#) or [content unpublished](#) for violating the DOI policy while reporting news in Arabic and/or from Gaza. The editor of 24FM Ihab Al-Jariri [described to Al-Jazeera](#) how they publish on Meta's platforms in three languages: Arabic, English, and Hebrew; however, only the Arabic content gets restricted, while the same news reports in English and Hebrew are not affected.

While there is a well-documented issue with censorship of Palestinian and Arabic content on Meta's platforms right now, hate speech and incitement to violence are still largely going unmoderated. [Meta internal documents acknowledged](#) that the Hebrew hostile speech classifier on Instagram comments was not as effective as it should, because it didn't have enough data for the system to function adequately. Our teams will continue to actively report cases of hate speech and incitement to Meta, and will also call on the company to implement stronger mechanisms to counter the spread of violent speech.

Civil society has made great effort to document and escalate cases to Meta throughout the course of this crisis. This is being called the [whac-a-mole](#) process, where civil society organisations follow up with the company on a case-by-case basis in order to ensure Palestinian content is not disproportionately targeted by the automated applications of the classifiers in Arabic. These efforts have been hugely important and resulted in protection of major journalist and human rights defenders' accounts, but this effort should not absolve the company from its responsibility to take action and change their policies.

2.X

X has been a primary concern for 7amleh since the start of the crisis. Since the takeover of Twitter by Elon Musk, the platform (now called "X") has become a hotbed for hate speech and incitement to violence against Palestinians. [This was happening before the recent escalations](#), and [has only gotten worse](#).

To better understand and document the situation, 7amleh recently released the "[Violence Indicator](#)", which documents hate speech and incitement to violence against Palestinians in Hebrew across social media. The indicator is updated in real-time, and as of writing this, it has already documented over half a million individual posts of harmful content on. This extremely high prevalence of harmful content on the platform is likely due to the company's lack of a Hebrew classifier, which would greatly increase its capacity to appropriately moderate violent and hateful posts in Hebrew. Moreover, X has shown a lack of interest in cooperating with civil society, which further exacerbates this issue and others.

As a consequence of the platform's erratic approach to freedom of speech, censorship of Palestinian voices, as well as other voices advocating for Palestinian human rights, is not as prevalent on X as it is on Meta, but there have still been documented instances where this was put to question. The account of a US-based [pro-Palestinian activist group](#), having originated from the UK, saw their account face a "[glitch](#)" where they were [suppressed from receiving any](#)

[new followers](#). Due to their prominence among circles of Palestinians and allies, the issue was noticed immediately by many users and civil society actors who called out X for suppressing the group's account. Three days later, and after a lot of [public pressure](#), the issue was 'resolved' and the account moved from having nearly no followers to over 200,000 followers in a short period.

In both incitement and disinformation cases, 7amleh has encountered a significant challenge posed by two of the company's policies: the community notes and the limited visibility warning labels applied to accounts with a substantial following. These policies advocate for [keeping the content online](#), even in situations where there may be potentially life-threatening consequences.

3. Telegram

Telegram, with its notorious policy of not moderating any content, has been a top concern throughout this crisis. Founder and CEO Pavel Durov calls the platform a messaging application based on channels, not a social media platform. There are no algorithms governing content, and users can choose to enter channels that align with their interests. This approach has created a hotbed for hate speech and incitement to violence against Palestinians. There are several Telegram channels which are [openly calling for genocide](#), and collective punishment of all Palestinian people. In some of the groups, moderators post [addresses to Palestinian individuals](#), community gatherings, and marches, calling on followers to plan attacks. These groups spread hate and organize real-world violence, but Telegram refuses to take them down. 7amleh has appealed to the company, but to date, it remains firm in its decision to continue to host bigoted and violent groups.

4. TikTok

TikTok is inundated with discussions concerning the situation in the region. The danger of significant digital rights violations has grown exponentially. 7amleh has thus uncovered similar issues of violations against Palestinians on the platform, although not as widespread as across other online platforms. In terms of censorship, after receiving pressure from civil society, TikTok reinstated several prominent accounts including the journalism and media outlet [Mondoweiss](#).

Regarding incitement, it appears that Hebrew content on TikTok remains under-moderated, despite the removal of a significant portion of the content reported by civil society. While documented violations still exist, the situation is not as dire as on other platforms. TikTok has proactively collaborated with civil society actors to better protect Palestinian users and content, displaying a more effective response compared to other platforms.

5. Google

Since October 7, 7amleh has received reports of YouTube ads that could be potentially considered incitement to violence. The pro-war propaganda was sponsored by the Israeli Foreign Ministry. At some point, there were [75 different ads released](#), many of which contained particularly graphic material, or were directed at children. One ad was titled “babies can’t read the text in this video but their parents can”, as a lullaby played against a backdrop of rainbows and unicorns. Google said that the company did not allow ads containing violent language, gruesome or disgusting imagery, graphic images or accounts of physical trauma. It was [reported that \\$7.1 million](#) had been spent on the potentially violating ads, which were heavily targeted [towards audiences](#) in France, Germany, UK, Belgium, US, Switzerland, and elsewhere. Civil society actors continue to engage with Google/YouTube in order to ensure that the company maintains its responsibility to deny add privileges to violating materials, such as hate speech, incitement to violence or disinformation.

6. Other platforms

- 6.1 LinkedIn

7amleh has gathered reports of shadowbanning and content deletions on LinkedIn, related to users who referenced the crisis while advocating for the importance of adhering to international law. In the cases of deletions, users have been told they were violating the platform’s ‘violent and graphic content policies’.

- 6.2 MailChimp

7amleh documented at least one case where a Gaza-based human rights organization had their [account suspended](#) moments after having added its address in Gaza. The suspension, which cited “regulatory requirements and/or corporate policies” is discriminatory in essence as it does not apply to organizations elsewhere.

- 6.3 Economic deplatforming

Similarly, Palestinian users of the Financial service eToro were notified that the platform would [“no longer be able to provide services to accounts which appear to be / are operating from the Palestinian Territory.”](#)

Israeli government

During the recent escalation, the Israeli government's actions have raised serious concerns about a crackdown on digital rights that some [observers have likened to a witch-hunt](#). This escalation in repression has particularly impacted [Palestinian citizens of Israel](#) and [residents of East Jerusalem](#), with a primary focus on digital rights. This section aims to comprehensively address the multifaceted issues arising from this situation, taking into account key aspects of arrests and interrogations, violations of digital rights, and broader human rights implications.

Arrests and Interrogations for Social Media Activity

One of the most alarming trends documented in this context is the arrest and interrogation of Palestinians for their social media activities. Numerous cases have emerged where [individuals have been apprehended](#) simply for expressing their views or opinions on various digital platforms. Adalah, the Legal Center for Arab Minority Rights in Israel, documented [at least 161 cases](#) (not including East Jerusalem) where Palestinian citizens of Israel were arrested, interrogated or indicted based on social media posts on Facebook, Instagram, and other platforms.

These actions have included posting content critical of the Israeli government or expressing support for Palestinian rights. Such arrests not only suppress the fundamental right to freedom of expression but also cast a shadow over the right to a fair trial and due process, as individuals often face prolonged detention without clear charges.

Harassment and Doxxing

In addition to the actions of Israeli authorities, there has been a disturbing trend of Israeli citizens [harassing and inciting against Palestinians on digital platforms](#). This harassment includes doxxing, where personal information of Palestinians is exposed in Telegram groups and other online platforms, often with calls for the Israeli authorities or [Israeli employers](#) to take action. Such practices have led to severe consequences for Palestinians and allies, including Jewish Israelis. Cases of [loss of employment](#) and [expulsion from academic institutions](#) are aplenty. This digital harassment campaign not only violates individual privacy rights but also exacerbates tensions and contributes to an atmosphere of fear and intimidation.

A significant concern is the apparent collaboration between Israeli authorities and those who engage in harassment and doxxing campaigns. It is evident that the Israeli government works in tandem with individuals who harass and target Palestinians, creating an environment where digital rights violations are not only tolerated but may even be facilitated. Some [analysts described this as a “totalitarian shift”](#), where large sections of Jewish Israeli society is on board with their government in not tolerating any voices expressing “Palestinianness”. This close cooperation between state authorities and individuals engaging in harassment further erodes the trust in the legal system and the protection of human rights.

Forced Phone Inspections in East Jerusalem

Additionally, Israeli forces have been [reported to forcefully inspect the mobile phones](#) of Palestinians in East Jerusalem. When any content is found on these devices that express solidarity with Gaza or documents rights violations by Israeli officials in the region, it often leads to arbitrary arrests or physical violence. Such actions not only infringe upon the right to privacy but also compromise the safety and well-being of Palestinian youth who are subjected to these inspections.

This underscores the urgency of addressing these issues from both a digital rights and a broader human rights perspective. It is essential to examine the violations of digital rights, freedom of expression, right to privacy, and the right to a fair trial, within the larger context of a situation that is causing profound harm to individuals and communities. The cases highlighted by Zameh and other human rights organizations provide a stark illustration of the pressing need to ensure the protection of human rights in the digital age and to counter the repressive climate that has emerged as a result of this crackdown on digital rights.

Incitement to Violence and Disinformation by Israeli Politicians

In the context of the recent escalation in repression of digital rights, it is crucial to address the disturbing phenomenon of incitement to violence and the dissemination of disinformation by Israeli officials, including Members of Knesset, government ministers, and military leaders, on social media platforms. These actions, which are often overlooked, play a significant role in perpetuating a climate of hostility and aggression.

Incitement to Violence

Numerous Israeli officials have [made inflammatory statements on social media](#) platforms, [inciting to violence](#) and hatred. Such rhetoric, when coming from individuals in positions of power, carries significant weight and does have real-world consequences. It not only fuels animosity but also legitimizes violence, making it particularly alarming in the context of digital rights violations and the broader human rights crisis.

Disinformation Campaigns

[Disinformation](#) campaigns, disseminated through official channels, have also been a concerning feature of the digital landscape in this context. The [intentional spread of false](#) or [misleading information](#) exacerbates tensions and fosters an environment of mistrust. Israeli officials using social media platforms to endorse or perpetuate such disinformation not only undermine the right to information but also hinder the prospects for peace and resolution.

Cyber Unit Activity

A proportion of the content removed by online platforms results from periodic requests made by Israel's Internet Referral Unit, commonly known as the "Cyber Unit." In total, they submitted 2,150 requests to Facebook, 1,240 to TikTok, 680 to Twitter, 245 to Instagram, and 143 to Telegram. [According to a representative of the Israeli cyber unit, Meta and TikTok responded positively to 90% and 85% of these takedown requests, respectively.](#) Notably, the Cyber Unit also takes credit for requesting the censorship of critical hashtags, including the Arabic hashtag #طوفان_الاقصى across Meta's platforms mentioned earlier.

International duty-bearers

Given the serious nature of the digital rights violations mentioned, it is important to highlight that various international duty-bearers hold responsibilities to ensure full compliance with international law, especially in the digital domain, particularly considering the impact of these violations on the physical world. This responsibility falls on entities like the United Nations, notably its Special Procedures and mechanisms, tasked with minimizing harm to affected individuals and communities. Countries, especially when their law enforcement authorities request online platforms to remove content related to the Palestinian context, also bear duties aligned with the principle of non-discrimination.

As a hallmark of global regulatory efforts in the digital sphere, understanding the impact of the EU's Digital Services Act (DSA) within the context of ongoing infringements on Palestinian digital rights becomes crucial. In light of mounting pressure exerted on online platforms by various stakeholders, most notably the European Commission and select Members of the European Parliament, influenced by a mainstream narrative that often overlooks the Palestinian perspective and the ongoing human rights violations by Israel, the current approach of the DSA in addressing illegal content and disinformation inadvertently contributes to the further erosion of Palestinian digital rights. This manifests in the suppression of Palestinian voices and advocates for Palestinian human rights, while simultaneously allowing for the proliferation of hate speech and incitement to mass violence against Palestinians.

The EU Commissioner for Internal Market, the DSA's chief enforcer", Commissioner Thierry Breton, [made statements addressing social media platforms'](#) obligations under the Digital Services Act, concerning content moderation in crisis situations. The statements, to our dismay, appeared to omit Palestinian experiences during the escalation of violence. This was seen as an extension of the structural human rights violations pervasive in the occupied territories. This kind of framing contributes to discrimination against Palestinians' online content, perpetuating their dehumanization and overlooking their pain and suffering. Non-discrimination among users is a fundamental principle of the DSA, aligning with the law's regulatory objectives and the EU Charter of Fundamental Rights.

In the United States, the House Energy and Commerce Congressional Committee [called on online platforms](#) to brief the committee on their moderation policies regarding the spread of illegal content, again omitting Palestinian experience and strictly addressing Israeli concerns. Similarly, US Senator Ted Cruz led the Senate's Commerce Committee [in addressing online platforms](#) X, Meta, TikTok, and Google, wholly omitting the Palestinian experience and focusing on how the platforms are complying with US laws and sanctions that require proactive censorship of Palestinian content.

At this critical moment, we reiterate our expectations of international duty-bearers to uphold a balanced and non-discriminatory approach to content moderation in the context of global crisis, such as the one we are currently witnessing. We also seek assurances that companies must adopt measures to address and prevent discrimination against the online content and lives of Palestinians and other vulnerable communities within the framework of the appropriate national and international laws and legislations.

Recommendations

Above all else, there needs to be an immediate end to the continued violence. For there to be substantial progress made in ensuring the protection of digital rights in the region, a ceasefire leading to the prompt end of the war is a necessary prerequisite.

Subsequently, all duty-bearers, whether in the public or private sector, must actively collaborate to unequivocally put an end to the systemic and deliberate discriminatory online policies and practices against Palestinians and individuals advocating for Palestinian human rights on a global scale.

- **Online platforms**

Online platforms need to prioritise a comprehensive approach that genuinely mainstreams and safeguards Palestinian human rights and addresses the root causes of discrimination against the community and narratives, in full transparency and in line with the United Nations Guiding Principles on Business and Human Rights. Obligations of due diligence, transparency and reevaluation of content moderation policies and practices must be met to create a safe space for all users and prevent violent speech in both Hebrew and Arabic languages equally.

In order to do this, 7amleh calls for the following:

- Commit to:

1. Streamlining the process for addressing digital violations reported by civil society actors, ensuring that every report is met with timely and well-founded responses.
2. Investing additional efforts and resources to prevent further harm and protect the rights and safety of both users of the platforms and other rights-holders, including in staffing with regional, linguistic, and socio-political knowledge to ensure localized content moderation is done in a rights-respecting manner.
3. Meaningful and regularly scheduled engagement with Palestinian civil society and stakeholders, not only during escalation in hostilities.
4. Recurrent, comprehensive, and transparent due diligence exercises to evaluate the impact of platforms' content moderation and curation decisions on individual and collective digital rights in Israel/Palestine.
5. Fully implementing, within a transparent, clear and detailed timeframe, all recommendations made by independent due diligence reports.

- Ensure:

1. Proportionality and accountability through due process as an additional safeguard to protect Palestinian digital rights.
2. Users are provided the chance to respond to content takedowns and are provided detailed and timely information of the reasoning behind decisions on content moderation.
3. Deleted content is kept for a proportional amount of time.

- Guarantee:

1. Non-discrimination and avoid "one-size-fits-all" policies and practices.
2. Respecting freedom of the press and recognition of the newsworthiness of content created by citizen journalists, and permit its presence on the platforms, even when it contains references to illicit organizations or graphic content, in order to ensure access to information

- Provide full transparency on:

1. The implementation of notice and takedown mechanisms.
2. Governmental (both legal and voluntary) requests for content removal.
3. Transparency about where automation and machine learning algorithms are being used in the content moderation process, including complete data on whether the content was correctly moderated, as well as keywords and hate speech lexicon used for both Arabic and Hebrew languages.

- **International duty-bearers**

- Call on the international community and the United Nations to take immediate and effective measures to halt ongoing systematic infringements on Palestinian digital rights, as well as other fundamental human rights.
- Urge the European Commission to ensure that online platform content moderation obligations, as set by the DSA, are evaluated in a non-discriminatory manner and systematically taking into consideration all the specifics of the context, in full compliance with the DSA requirements and spirit.
- Advocate for the enforcement of regulations and alternative measures to compel online platforms to cease discriminatory practices against Palestinians.
- Urge online platforms to adhere to business and human rights principles, as well as international humanitarian law, during the development and implementation of their policies, with a specific emphasis on due diligence responsibilities, especially in times of crises.
- Actively engage with civil society and various stakeholders to effectively address their concerns regarding the safeguarding and advancement of Palestinian digital rights.