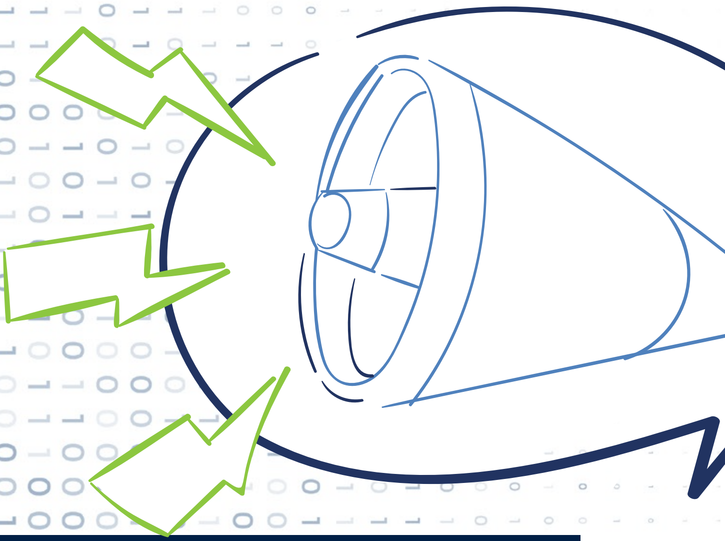
حملة – المركز العربي
لتطوير الإعلام الاجتماعي
7amleh - The Arab Center for
the Advancement of Social Media



The Impact of X Platform's Content Moderation Policies on Palestinian Digital Rights

7amleh –The Arab Center for Social Media Advancement

Position Paper on the Impact of X (formerly Twitter)'s Content Moderation Policies on Palestinian Digital Rights

Author: Jalal Abukhater

Copyeditor: Eric Sype

Contact Us:

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

Tel: +972 (0)774020670

Find us on Social Media: **7amleh**

Introduction

X, formerly known as Twitter, remains a leading online platform for information exchange. It is the preferred platform for journalists and those who seek live news updates and analysis in the micro-blogging format. It is also the preferred format to find public statements and commentary by policymakers, state officials, and other influential figures in society. **While X's 556 million active monthly users** pale in comparison with larger online platforms such as Facebook (2.96 billion), Instagram (2 billion), and TikTok (1.05 billion), X maintains a leading position in rapid dissemination of news and remains an important source for original content, much of which carries significance in social and political contexts within society. However, many issues continue to persist on X, putting into doubt the platform's commitment to user safety and to safeguarding digital rights.

There are many reasons for concern, including the unprecedented amount of hate speech and disinformation proliferating on X. However, there are also reasons for optimism, particularly within the Palestinian digital rights context. This can be characterised by: the contrast in the openness of X to a wide array of Palestinian voices, as opposed to the silencing of Palestinian voices observed on other online platforms; the promotion of grassroots Palestinian journalism and human rights advocacy in the response to mainstream media often misconstruing and decontextualizing the Palestinian narrative; as well as the provision of evidence regarding incitement and other human rights and international humanitarian law violations. This evidence potentially implicates Israeli officials in the International Court of Justice's proceedings on plausible violations of the Genocide Convention during the war on Gaza.

Ever since the take-over of Twitter by Elon Musk, there have been growing fears about what this platform could become. Musk wasted no time in implementing bold changes, from tweaking the platform's interface to overhauling its content moderation policies. Twitter users and many digital rights advocates were justifiably concerned. The take-over of the platform by Musk has been unequivocally troubling, however, the comparative absence of a crackdown on Palestinian speech deems it worthy of further investigation.

When Musk expressed his intention to acquire Twitter, it stemmed from his pronounced dissatisfaction with the platform's handling of free speech and content moderation matters. He expressed his intention to make significant changes, citing his belief in Twitter's **"potential to be the platform for free speech around the globe,"** and adding that he considers free speech to be a societal imperative for a functioning democracy. Musk acquired the company in October 2022.

Musk's push to increase "freedom of expression" on X coincided with mass layoffs that affected many departments within the company. Since the takeover, **reports show that 1,213 trust and safety specialists have left X**, nearly eliminating the teams responsible for stopping abusive content online. **Human rights** and **communications** teams, as well as the **Trust & Safety Council**, were all dismantled. This created conditions for a "perfect storm" of disinformation, hate speech, incitement, and dehumanising content online.

When it comes to Palestine, the data collected by **The Palestinian Observatory for Digital Rights Violations (7or)** gave us a glimpse of where X stood on freedom of expression, compared to other major online platforms. In 2023, out of 1606 censorship-related digital rights violations, only 2% of those violations occurred on X. Meanwhile, Facebook, Instagram, and TikTok dominated the violations chart, respectively, boasting 93% of the total censorship-related violations recorded through 7or during the same period.

When it comes to hate speech, the situation is much more grim. 7or also monitored the growing proliferation of hate speech and incitement to violence on X, in comparison to other major platforms. Out of **2740 total instances of harmful content documented** in 2023, 33% of the violations occurred on X, which was only second to Facebook, a much larger online platform. Furthermore, an AI-powered Language Model that monitors the spread of hate speech and violence against Palestinians in Hebrew, the "**Violence Indicator**" monitored nearly 3 million instances of hateful and/or inciting content on X between October 6 and December 31st, 2023.

Musk also cast doubt on the platform's reliability as he pushed to dismantle the old blue-check verification system, accusing it of being an **elitist "lords" and "peasants"** system. He vowed to grant all users access to the blue "verified" checkmark if they paid for it. This move signalled a potential loss in the platform's credibility mechanisms and an increase in uncertainty and disinformation across the platform.

This paper concludes with a set of recommendations for X. The platform must commit to user safety and safeguard digital rights, including but not limited to the rights to privacy, freedom of expression, and access to information, as enshrined in the **Universal Declaration of Human Rights** and the **International Covenant on Civil and Political Rights**. It also needs to ensure ethical governance and accountability mechanisms are operational and effective. Furthermore, the recommendations also address policymakers and third-party actors, urging them to ensure that X complies with business and human rights standards and is aligned with the UN **Guiding Principles on Business and Human Rights** and other international regulatory frameworks.

Hate Speech persists on X

X's stated policy on **Hateful Conduct**, developed in April 2023, states: "You may not directly attack other people based on race, ethnicity, nationality origin [...]" prohibiting behaviour "that targets individuals or groups with abuse based on their perceived membership in a protected category." Similarly, its **Violent Speech Policy**, published in October 2023, clearly states: "You may not threaten, incite, glorify, or express a desire for violence or harm" on X platforms. Those aforementioned policies do not align with the reality of X, considering anti-Palestinian hate speech and incitement proliferated on the platform, reaching unprecedented numbers over the same time both of these policies were published.

Online Incitement Mirrors Real-World Harm

During the first months of 2023, 7amleh documented an unprecedented rise in hate speech and incitement to violence, particularly on X. This coincided with an intensification of violent attacks by Israeli settlers on Palestinian communities throughout the occupied West Bank. This concern climaxed on the night of the 27th of February during an attack on the village of Huwara. It was **documented that a total of 15,250 tweets published in Hebrew** during the first three months of the year contained direct incitement on Palestinians, and specifically on the village of Huwara. Hashtags such as #למחוק_את_חווארה (Wipe out Huwara) proliferated on X. 7amleh published a report on the matter, indicating how Israeli incitement to violence against Palestinians in the digital space directly correlated with violence inflicted upon Palestinians on the ground.

7amleh raised the alarm again with its groundbreaking AI-powered language model, which serves as a real-time indicator of the volume of hateful, inciting, and other forms of incendiary speech in Hebrew on X.

In the context of the war on Gaza, 7amleh detected **nearly 3 million instances** of violent content in Hebrew targeting Palestinians on X between October 6 and December 31. This kind of incendiary content by the Israeli public, along with statements by Israeli officials on X describing the Palestinian people as **"human animals"** and **"children of darkness"**, has translated into unlawful acts in Gaza.

These acts may be in violation of the Genocide Convention, including the killing of over **29,514 Palestinians**, the destruction or damage of over 70 percent of all housing units, and the forced displacement of over 1.9 million people, more than 80 percent of the population.

On X, it was evident through the **monitoring and documentation work conducted by 7amleh**, that hateful and violent content in Hebrew was not sufficiently moderated. According to the **first transparency report** published under the EU's Digital Services Act, X has only 12 human content reviewers for Arabic and 2 for Hebrew. For comparison, there are 2294 human content reviewers for English. There was serious concern about the unprecedented spread of hate speech and incitement on X, raising questions about the effectiveness of the platform's **'Hateful Conduct'** and **'Violent Speech'** content moderation policies.

X's Legal and Moral Responsibilities in Halting Incitement

On October 27th 2023, the United Nations Committee on the Elimination of Racial Discrimination expressed serious concern about "the sharp increase in racist hate speech and dehumanisation directed at Palestinians since 7 October, particularly on the Internet and in social media." For instance, a post from the Israeli Deputy Mayor of Jerusalem on December 8 described stripped and blindfolded Palestinian detainees in Gaza as "hundreds of ants" that he wanted to bury alive; "they are neither human beings nor human animals, they are sub-humans and that's how it should be." The post was removed after it was reported, but many others remain. Such rhetoric, allowed to persist unchecked, not only perpetuates the dehumanisation of Palestinians but also fuels an environment where violence is normalised and celebrated.

On January 26th 2024, the International Court of Justice (ICJ) ordered provisional measures in the case of South Africa v. Israel, determining the plausibility that Israel is carrying out genocide against the Palestinian people in Gaza and recognising the risk of irreparable harm. The Court observed several inciting statements made by Israeli leaders and specifically referred to a post shared by the Israeli Minister Israel Katz on X, which reads: "The line has been crossed. We will fight the terrorist organisation Hamas and destroy it. All the civilian population in Gaza is ordered to leave immediately. We will win. They will not receive a drop of water or a single battery until they leave the world." The Court adopted legally binding orders that include requiring Israel to prevent genocide against Palestinians in Gaza as well as to prevent and punish direct and public incitement to commit genocide, as outlined in Article III(e) of the Convention on the Prevention and Punishment of the Crime of Genocide.

Considering the documented use of online platforms to incite genocide against Palestinians in Gaza, including by the highest levels of the Israeli leadership, the ICJ order underscores the necessity to address the responsibility of online platforms such as X in respecting human rights and preventing the dissemination of harmful content.

This includes incitement to commit genocide, within their domains, both legally and morally. On the 7th of February 2024, the Palestinian Digital Rights

Coalition sent an urgent letter to the Chief Executive Officer of X, Ms Linda Yaccarino, calling for action to address X's legal and moral responsibilities in halting incitement against Palestinians in Gaza on its platform. This call was made in light of the ICJ provisional measures order in genocide case.

The failure to effectively address hate speech and incitement not only undermines the safety and well-being of the Palestinian people but also violates the platform's obligations under international law and human rights principles.

Combatting Disinformation

Following Musk's takeover, X continued to seek methods to combat the spread of mis & disinformation on the platform, especially since the company had laid off most of its Trust & Safety staff. The company's answer was "Community Notes," which was developed as an alternative method to combat disinformation. Community Notes is a crowd-sourced fact-checking tool where "contributors can leave notes on any post and if enough contributors from different points of view rate the note as helpful, the note will be publicly shown on a post." The new feature would attach a label to the offending post, highlighting inaccuracies or providing needed context in cases of alleged misinformation or disinformation. X today boasts of its commitment to accuracy through its "Community Notes" feature, which has more than 375,000 contributors across 65 countries. However, there was major doubt about the tool's effectiveness in combating disinformation on its own, in the absence of Trust and Safety staff at X.

While the Community Notes feature is being hailed by the company as a success, it has been deemed "not sufficient" by Community Notes contributors, as well as by the company's former head of Trust and Safety Yoel Roth, and former Trust and Safety staff members. Their primary concern is that 'Community Notes' should complement rather than replace Twitter's other misinformation methods, and trust and safety staff.

Additionally, an MIT study from April 2022 found that contributors to the Community Notes method are "much more likely to fact-check posts expressing political views that differ from their own," rather than have the knowledge and skills to debunk misinformation or offer alternative, correct information.

In the context of the War on Gaza, an isolated case showed the potential risk of the Community Notes feature. Amnesty International's Secretary General, Ms Agnes Callamard, expressed on X the organisation's deep alarm: "by the mounting civilian death tolls in Gaza, Israel and the occupied West Bank." A subsequent community note accused her of employing a "twisted logic" to blame civilians, attempting to inappropriately erode the author's credibility by generating the false impression of factual error. However, the note was promptly removed after several alerts from users. In this instance, the company's transparent guide was effective, which shows that the tool should not be abused with argumentative, biased, irrelevant, or rather speculative contributions, but the example highlights the problematic nature of outsourcing content moderation tasks to the general public.

Disinformation campaigns, disseminated through official channels, have also been a concerning issue, as they go hand in hand with calls for violence.

Following October 7th, X appeared to have had the worst outbreak of fake news related to the war of the major social media platforms. The intentional spread of false or misleading information exacerbated tensions and fostered an environment of mistrust.

Israeli officials using social media platforms such as X to endorse or perpetuate such disinformation undermine the right to information and hinder the prospects for cessation of hostilities.

X's owner, Elon Musk, received a rebuke from many EU capitals over the alleged disinformation about "the Hamas attack on Israel, including fake news and repurposed old images." This prompted the EU Commissioner for Internal Markets, Thierry Breton, to write to X requesting compliance with the Digital Services Act (DSA), writing: "We have indications your platform is being used to disseminate illegal content and disinformation in the EU." On December 18, 2023, the EU Commission opened formal proceedings against X under the

DSA. While it is significant for a regulating actor, such as the EU Commission, to address the issue of illegal content and disinformation, there were concerns about politicisation and discrimination by the EU Commission. Their initial letter gave no reference to the experiences of the Palestinian people during a critical time when illegal content, incitement and disinformation were heavily used online to **dehumanise Palestinians and justify their collective punishment**.

Verification, Identification, and User Privacy

One feature affected by Musk's overhaul of the platform was the 'Verified' blue check. Musk's **decision** to grant access to the blue check to all paying subscribers raised serious concerns about the risk of mis and disinformation spreading on the platform. The change raised doubts about the authenticity of information on X. However, one could interpret Musk's decision, unintentionally, as rendering the concept of the "blue check" useless, weakening what he described as "elitist" voices, including mainstream media giants. This change may have inadvertently amplified the voices of others eager to make an impact on the platform, for better or worse.

Many **critics raised questions** about impersonators and parody accounts who could deceive other users about their identity. To address this, X introduced mechanisms to prevent such deception by temporarily disabling the blue check for users who alter their profiles or violate the platform's policies under the subscription program.

In September 2023, X introduced the ID Verifications **Policy**, requiring X users who wish to obtain a verified badge to verify their government-issued ID. This was billed as an opportunity for X "to increase the overall integrity and trust on [the] platform." This method, proposed as a way to combat spam bots and trolls on X, raised concern about users' personal data privacy. X contracted a third-party company, Au10tix, as its "data processor" **to handle the user's personal information** and store the information for up to 30 days, as required by the verification process. This raised concerns about privacy rights in general, and specifically those of Palestinians and their allies.

The company Au10tix is located in Israel, and has a **well-documented history of military surveillance and intelligence gathering**.

Au10tix’s founder and current chairman, Ron Atzmon, served in the Israeli Military’s Intelligence **Unit 8200**.

That unit is responsible for surveilling Palestinians and gathering intelligence to use for “political persecution”, according to **a letter by 34 Israeli soldiers** who served in Unit 8200. Their letter further added that Unit 8200’s purpose is to “maintain the continued control over millions of people through thorough and intrusive supervision and invasion of most areas of life”.

In addition to its founder, several of Au10tix’s engineers served in Unit 8200 as well. The company’s **leadership** held previous positions in major Israeli cyber companies accused of controversial surveillance tactics, such as **Cellebrite**, **AnyVision**, and **Voyager Labs**. Au10tix’s association with X raises questions about the potential implications for user privacy and data security.

Challenging the narrative

In April 2023, X introduced changes to the verification process, which previously required the account holder to be **“authentic, notable, and active,”**. Individuals also had to meet the criteria of being a government figure or entity, entertainer, athlete, company, activist, content creator, news organization or journalist. The new change made obtaining a verification blue check available to paying subscribers on X, rather than solely relying on “notable” status and/or association with notable organisations.

This change opened up the space for independent journalists and researchers, including many Palestinian and pro-Palestinian rights advocates eager to create and publish content and relay a narrative without having to self-censor. This is in contrast to **other platforms** where disproportionate suppression of Palestinian voices and narratives was well-documented. This allowed many to challenge established mainstream media narratives that are often biased against Palestinians.

A quantitative analysis of major media outlets, conducted by the Intercept, showed that The New York Times, Washington Post, and Los Angeles Times' coverage of Israel's war on Gaza demonstrated a consistent bias against Palestinians. Meanwhile, by October 19th, pro-Palestinian posts were 3.9 times more common than pro-Israeli ones on X.

Mainstream media outlets bear the responsibility of providing accurate and unbiased coverage. In the new dynamics at X, independent and upstart journalists carry similar weight to mainstream media outlets in regard to reach and visibility. This is due to the organic nature of news on the platform, and the users' ability to utilise the community-based features to counter potentially misleading information.

In the quest for justice, the media becomes a force for change. When Palestinian narratives are not systematically suppressed but rather amplified, bearing in mind the heavily asymmetrical nature of the struggle that Palestinians face daily, one becomes aware of the critical role of a free and organic press. This includes journalists in Gaza who risk their lives and liberty to share their reality on platforms like X, illuminating the critical role of empowering such voices

Free Speech, Discrimination and Impulsive Censorship

Rhetoric vs Action

In light of the War on Gaza, there was a growing affinity towards X as an alternative to platforms known to censor Palestinian voices and narratives. X seemed to keep up with its promise to preserve the right to freedom of expression as a core tenet on the platform.Early on, Musk repeatedly expressed his intentions for the platform by describing X as "the global town square, from the people for the people." The words of X CEO Linda Yaccarino confirm the continuation of Musk's vision, as she recently said: "Faced with the option of a future dependent on gatekeepers and moderators that restrict the free flow of information or a world where information flows freely – I choose information independence." Information Independence is described by the CEO to be the free exchange of ideas, information and knowledge through freedom of speech.

While Musk promoted himself as a **"free speech absolutist"**, his words did not always match his actions. Inside the company, he would go after and dismiss company employees for making **"snarky comments"** about him, deeming them as "untrustworthy or disloyal". Additionally, Musk made it his mission to ban an account that used publicly available data to track the locations of his private jet. The account, called ElonJet, was banned despite Musk **having said** "my commitment to free speech extends even to not banning the account following my plane," **one month before banning** the account, reneging on his supposed commitment.

Unchallenged compliance with government-issued take-down requests was another area of major concern. Before Musk's takeover, Twitter's legal team was described as an **"aggressive defender of free speech"** when it came to companies defending the privacy and free speech rights of its users. Among many things, Musk eviscerated **the company's legal talent**, firing the company's chief legal officer, its general counsel and deputy general counsel. As a consequence, the company's self-reported data showed that X's **compliance with government orders for censorship or surveillance**, especially in Turkey, Germany, and India, was over 80%. In the previous year, the full compliance rate hovered around 50%.

Palestinian Voices and the Free Speech Dilemma

In a Palestinian context, the spiel by X leadership on freedom of speech was encouraging. However, the shifting landscape post-October 7, coupled with political dynamics and external pressures, unveils an alarming trajectory on the horizon.

Early in October 2023, in response to a letter by the EU Commissioner Thierry Breton, **X CEO Linda Yaccarino wrote** back stating X had "removed hundreds of Hamas-affiliated accounts from the platform" under its **Violent and Hateful Entities Policy**. It is worth pointing out that X's Violent and Hateful Entities Policy clearly states what are not considered violations of this policy. The exceptions include discussions of violent and hateful entities for clearly educational, documentary, and/or newsworthy purposes. This is particularly significant because, when it comes to Meta's larger online platforms, the **demands for clarity regarding** the "**Dangerous Organisations** and Individuals Policy" and the demands for newsworthiness allowance are made repeatedly by Palestinian and international digital rights and other civil society organisations who document widespread and disproportionate censorship of Palestinian narratives.

There was careful optimism about X among Palestinian users. Incidents of Palestinian users' content being restricted or de-platformed on X remained low when compared to other large online platforms.

However, it is important to keep in mind X's monthly active user base is less than one-fifth of the largest platform, Facebook. Still, some issues persisted. Recently, an incident where accounts of prominent journalists and leftist voices were suspended, prompted users to describe the event as a 'purge'. The only thing all the suspended accounts had in common was they had been critical of Israel's war in Gaza. Hours later, and following an outcry, all accounts were mysteriously restored. On X, Musk went on to claim that the bans were due to a "spam/scam bot" accidentally flagging many legitimate accounts, and that it "is being fixed."

The concerns about X moving to censor speech related to Palestine only grew. A prominent UK-based pro-Palestine activist group, PalAction, had a US chapter initiated. Their US account @pal_actionUS started with 12k followers. However, it experienced an unusual throttling of followers, resulting in the number dropping to nearly zero despite many attempts to follow the popular account. The issue was an active topic of discussion that day, thousands attempted to follow, only to see with their own eyes the "follow" icon removed as soon as they refreshed the page. After some noise and follow-up by civil society, the account was "restored" and could accept new followers normally. It gained over 200,000 followers within its first few days. However, a few months later, the @pal_actionUS account was suspended for alleged rules violation, without further justification.

X's Abuse and Harassment policy, updated in January 2024, added further cause for concern. Under the policy, content that includes "Violent Event Denial" would be considered "Abuse," and X would enforce the policy by a range of options, including limiting post visibility. In February 2024, users saw the label "this post may violate X's rules against Abuse" appear when sharing commentary about the violence in Gaza and pointing to double-standard in reactions the global community showed to debunked claims vs verified documentation. The application of this policy could pose problems; it must be based on factual evidence rather than anecdotal accounts. Additionally, policy enforcement should be fair and consistent across the user base, particularly as denial of the ongoing human suffering in Gaza, or even Nakba Denial as a form of historical denialism, persists to this day, perpetuating violence and dehumanisation against the Palestinian people.

Elon Musk's Political Posturing and the Threat to Free Speech

While Musk's pledge to respect freedom of expression on X might have seemed encouraging initially, fears were growing among human rights advocates and defenders of freedom of speech, particularly regarding the suppression of discussions critical of Israel.

Moreover, Musk made multiple eyebrow-raising pro-Israel statements, in what seemed to be a way for him to avoid scrutiny after controversy surrounding the boosting anti-semitic content on X. On November 18, 2023, two days after said controversy, Musk went on X, saying "decolonization" and "from the river to the sea" necessarily imply genocide, and using those phrases would "result in suspension" on X. With the suspicious timing, many interpreted these statements as a way for Musk to deflect from criticism of actual anti-semitic behaviour.

The use of "From the River to the Sea" is interpreted differently by different people, however, in the Palestinian context it often reflects peaceful aspirations and should not be considered obscene or inappropriate. For many, it embodies the desire for equality, freedom, and dignity for all inhabitants of the land, regardless of race, ethnicity, or religious background. It also addresses the ongoing denial of Palestinians' right to self-determination and fosters solidarity in a global context.

Musk's statement on "decolonisation" and "from the river to the sea" prompted the Anti-Defamation League's (ADL) CEO, Jonathan Greenblatt, to thank Elon Musk, adding that he "appreciates this leadership in fighting hate." Greenblatt himself is an advocate for conflating discussion of the root causes of the conflict and Israeli policies to support for terrorism. During recent months, the ADL itself was accused of defaming Palestinian students in the US as "terrorist supporters," causing the ADL's own staff to decry the organisation's "dishonest" campaign to silence Israel critics.

Banning the use of "decolonisation" or the term "from the river to the sea" is wrong. Such a drastic policy suggestion is not only baseless, it would set a dangerous precedent, harming Palestinian users and their right to free expression online. While the policy suggestion brought up by Musk did not materialise into policy enforcement as of the time of writing, the cause for vigilance remains high. The owner of X is known to be impulsive in his decision-making process. His words, even if they have not materialised into policy, have to be taken seriously.

X tries to promote itself as a free space where everyone can have their say. While causes for concern remain high, there were reasons for Palestinian users to choose X over other platforms, mainly due to the low bar set by other major social media platforms, particularly in regards to the proclaimed guarantee of freedom of expression. Maintaining this guarantee seems to be an uphill battle for Palestinian rights advocates in light of rising influences and external pressures.

Preserving Evidence of Human Rights Violations

Following October 7th, and the ensuing violence unleashed by Israel, graphic content showing victims of violence proliferated across social media platforms. This violence prompted UN experts to state **Israel is "very likely" committing genocide** in Gaza. On October 10th, X issued an **update to their Public Interest Policy**, stating that while "it's sometimes incredibly difficult to see certain content, especially in moments like the one unfolding, X believes that it's in the public interest to understand what's happening in real-time."

Counter to their stated policy, 7amleh documented several instances where X forced users to delete media posts "of a graphic/sensitive nature", to regain access to their account. The affected content consisted of videos or images depicting civilians, often children, as a direct consequence of Israeli strikes on structures and densely populated areas in Gaza.

X must find ways to preserve access to the content of a graphic nature when there are plausible connections to rights violations. Documentation of this nature is critical to the general public's awareness of the issue, as well as future pursuits for justice.

In January 2024, X updated its **policy on Sensitive Media,** stating that, because people use X to show what's happening in the world: "You may post graphic content that falls under our definitions of Graphic Content [...] with a content warning, but you may not share this media in a live video, or in your profile picture or header, List banner, or Community cover photo." The policy noted that exceptions to the Sensitive Media / Graphic Content policy become applicable when the content is deemed to be of "documentary or educational" nature. The updated policy also stated that X would not force the deletion of unlabeled graphic content, but rather would add a content warning to the content in case users fail to add a content warning themselves.

A careful optimism with X's content moderation policies continues to stand on a thin pedestal. On one hand, activists and human rights defenders managed to collect evidence of Israeli human rights violations and war crimes, in addition to statements by Israeli leaders inciting violence. For example, a statement made by the Deputy Knesset Speaker that originated on X on **November 17**, saying "Burn Gaza Now," was used **in South Africa's case against Israel** at the International Court of Justice on January 11. On the other hand, actual incitement and hate speech targeting Palestinians ran rampant on the platform. The unprecedented amount of Hebrew-language inciteful content was a condemnation of Israeli society as well as a condemnation of X's ineffective content moderation policies.

Reiterating the International Court of Justice's (ICJ) **order on provisional measures** in the case of South Africa v. Israel, the fourth provisional measure prohibiting public and direct incitement to genocide was not the only item requiring attention by online platforms like X. The fifth provisional measure instructs Israel to "ensure the preservation of evidence" on potential violations of the Genocide Convention. Likewise, online platforms have a responsibility to ensure they **do not impede the preservation of evidence** related to potential breaches of the Geneva Convention and other violations of international human rights and humanitarian law.

X's Sensitive Media policy is a good indication that the platform takes seriously the concern that disproportionate content moderation could result in denying experts access to content posted on X with evidentiary value documenting human rights violations. This is consistent with **Musk's statement** early in October that "disturbing content on X should be seen, even if it is difficult." Nevertheless, the platform must take all effective measures to ensure the preservation of content of evidentiary value, specifically content that could be used in the genocide case, showing ethical and moral compliance with the International Court of Justice's provisional measures.

Recommendations

For X (formerly Twitter):

Commit to User Safety:

Reinstate the Trust and Safety Council and commit to fighting Islamophobia, anti-Semitism, and other forms of hate speech and incitement to violence in all languages to safeguard users globally. X must Invest in effective hostile language classifiers, particularly for the Hebrew language, and ensure having staff with adequate language ability as well as knowledge of socio-political contexts, to manage content fairly and equally amidst the sharp rise of incitement in Israeli society against Palestinians.

Develop effective mechanisms to combat the spread of mis and disinformation, as the current community-based tool is ineffective as a standalone mechanism. Last but not least, X should establish clear and accessible mechanisms for users to appeal decisions and seek remedy in cases of content takedowns and provide detailed and timely information on the reasoning behind decisions on content moderation.

Safeguard Digital Rights:

- **Right to Privacy:**

  Safeguard user privacy as enshrined in Article 12 of the Universal Declaration of Human Rights, by prioritising the protection of user data integrity and privacy rights. X must have a vetting policy to approve all third parties that X shares data with and avoid companies known for espionage or negative surveillance practices, to protect the right to privacy of all users. Lastly, X should restore and reinforce its legal teams to bolster defence against government surveillance and arbitrary content takedown, safeguarding user privacy and free expression.

- **Freedom of Speech:**

  Uphold the right to Freedom of Speech as enshrined in Article 19 of the Universal Declaration of Human Rights, ensuring Palestinian voices are not silenced for political reasons, upholding the principles of free expression regardless of political affiliations.

Establish a transparent and inclusive process for evaluating and implementing speech-related policies by involving representatives from diverse backgrounds, including Palestinian communities, to ensure their perspectives are considered.

- **Access to Information:**

  Ensure free or affordable API access for academics, researchers, and civil society groups to promote research and understanding. Additionally, commit to full transparency regarding both legal and voluntary takedown requests from governments and Internet regulatory authorities (IRUs). Preserve content of evidentiary value documenting human rights violations, including potential war crimes, despite removal requests or automated removal processes. Lastly, respect freedom of the press, recognize the newsworthiness of content created by citizen journalists, and permit its presence on the platform, even when it contains references to illicit organisations or graphic content to ensure access to information as enshrined in the International Covenant on Civil and Political Rights (ICCPR).

Ensure Ethical Governance & Accountability:

Regularly conduct thorough human rights due diligence impact assessments to mitigate risks of platform policies contributing to violations of human rights and international humanitarian law. Commit to ongoing co-design with civil society to enhance policies and processes related to Palestinian content, improve support for users, provide transparent reasons for content policies, and establish clear appeal methods with timely responses.

For Policy Makers:

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.

Call on social media platforms to take all means necessary to prevent atrocities against humanity, including the crime of genocide, and combat the proliferation of hate speech, incitement to violence, and disinformation that leads to dehumanization, discrimination, and violence.

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.

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.

Contact 7amleh:

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

Find us on social media : **7amleh**