Racism and Incitement

Index 2024

March 2025

7amleh- the Arab Center for the Advancement of Social Media

Racism and Incitement Index 2024

March 2025

Author: Rama Salahat and Ahmad Qadi

Revised and edited by: Jalal Abukhater

Table of Contents

Introduction

As the genocidal violence in Gaza continued throughout 2024, a second layer of digital violence emerged in the form of relentless hateful and inflammatory online content targeting Palestinians. This digital violence mirrored the physical violence on the ground, with both forms feeding into one another. Calls for increased bombardment and killing circulated widely online, while Israeli soldiers frequently uploaded videos of their acts of violence against Palestinians to social media, further fueling the harm and dehumanization of Palestinians.

The surge in violent Hebrew content on social media has been significant, though systematic hate speech against Palestinians predates October 2023. Between January and September 2023, 7amleh detected approximately 7 million instances of violent Hebrew content targeting Palestinians across various platforms, with Facebook alone accounting for 24.57%.

A 2021 report by Business for Social Responsibility (BSR) found that Meta lacked functional classifiers for hate speech and violent content in the Hebrew language, leading to under-enforcement of policies. While Meta claimed to work on introducing classifiers for Hebrew language in 2022, it was revealed that those classifiers were not effective, as revealed through the company's internal documents in October 2023. Since October 2023, violent Hebrew content targeting Palestinians has surged as revealead by 7amleh's documentation, exposing the continued failure of Meta's moderation efforts.

Given this history, Meta recently announced significant changes to its content moderation approach, stating that its filtering systems will now focus only on what it classifies as "high-severity" violent content. This shift means that many forms of violent content previously moderated will now remain on its platforms. While the full impact of this change is unclear, it appears likely to allow even more unchecked proliferation of violent Hebrew content targeting Palestinians online.

While Meta's changes are concerning, X remains the primary challenge when it comes to violent content against Palestinians. This is particularly alarming because, despite a smaller Israeli presence on X compared to Meta's platforms, the volume of violent content is significantly higher. Moreover, this large amount of violent tweets originates from a relatively small number of users compared to Meta.

Meanwhile, at the local political level, Israeli authorities continue to detain and repress Palestinian citizens of Israel for merely posting or expressing solidarity with

Palestinians in Gaza. In contrast, Israeli Knesset members, government ministers, and politicians openly lead hate and incitement campaigns against Palestinians without any legal consequences, further highlighting the double standards in enforcement both online and offline.

Over the last year, 12,482,041 pieces of content in Hebrew were identified as violent or hateful, equating to 23.6 instances per minute, with interactions totaling 187,226,176 across platforms. This staggering volume reflects the systemic nature of online hate, exacerbated by significant real-world events, particularly the devastating genocide in Gaza. The tragedy not only intensified the conflict but also fueled an alarming surge in incitement and hate speech, creating fertile ground for inflammatory rhetoric targeting Palestinians in order to dehumanize them and legitimize the Israeli attacks.
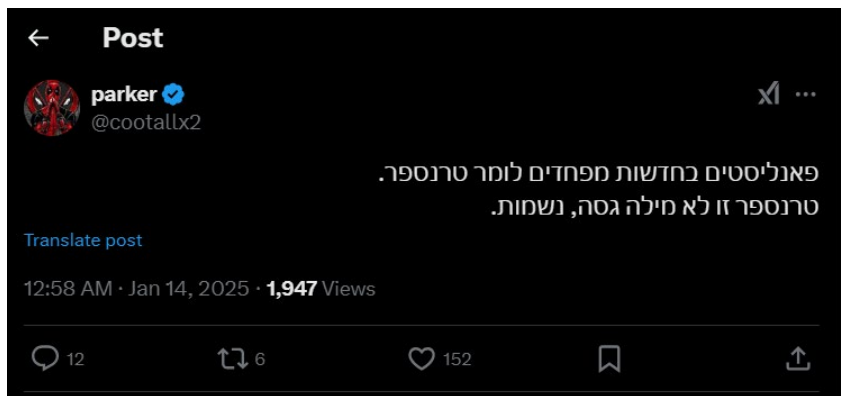

Death to all Gazans, Amen, God willing. How simple this is, and how true it is.

Also, there are no innocent people in Gaza

5

Lev Liberman
January 6 at 10:29 AM · 🌐

טרנספר זאת לא מילה גסה.
בואו נחזיר אותה לשיח הציבורי, צריך להתייחס לטרנספר כפיתרון לגיטימי לחלוטין.
הרי השמאל כבר עשרות שנים בעד טרנספר! פשוט הוא רוצה טרנספר של יהודים
ולכולם משום מה זה נראה סבבה.
אז בואו נדבר על טרנספר של ערביי יהודה ושומרון

👍 11   4 💬   1 ↪

Transfer is not a vulgar word. We must bring it back into public discourse, and we must refer to transfer as a completely legitimate solution.
So, the left in Israel has supported transfer for decades! They simply want the transfer of Jews and everyone else, and for some reason, that is considered acceptable.
So, let›s talk about the transfer of the Arabs of Judea and Samaria

← **Post**

parker ✓
@cootallx2   𝕏 ···

פאנליסטים בחדשות מפחדים לומר טרנספר.
טרנספר זו לא מילה גסה, נשמות.

Translate post

12:58 AM · Jan 14, 2025 · **1,947** Views

💬 12   🔁 6   ❤ 152   🔖   ↥

The Palestinians are afraid to say transfer in the news.
My dear ones, transfer is not a vulgar word.

The Racism & Incitement Index 2024 provides a detailed examination of hateful and violent Hebrew content trends against Palestinians on social media platforms, focusing on Facebook and X online platforms. Hateful and violent content could be statements or communications that create an imminent risk of discrimination, hostility, or violence by directly calling for action, expressing intent, advocating for violence, hoping for harm, aspiring violence or conveying approval, encouragement, glorification or identification with acts of violence. Considering that violence is broadly defined as actions that result in harm, injury, or damage to individuals, groups, or property. It also could be any type of hate speech without necessarily calling directly for violence, which is any form of discriminatory or pejorative verbal or written discourse, communication, or content that expresses, encourages, flares up, or incites hatred against people or groups based on inherent characteristics

or specific factors of their identity such as gender, race, color, nationality, religion, origin, or political opinion. In doing so, hate speech creates an environment of violence and social, political, and cultural rifts.

The report categorizes hate based on motives such as political, racial, religious, and gender-based biases. It reveals a disproportionate targeting of Palestinians, but also specific religious communities, particularly Muslims and Christians, alongside heightened hostility toward Jerusalemites and gloating over victims in Palestinian towns in Israel. Temporal analysis shows spikes in hateful content during key moments of escalation in Gaza, indicating a direct correlation between violence on the ground and its reflection in digital discourse.

Through an in-depth analysis of temporal trends, content, and thematic focuses, the report unveils the pervasive impact of digital violence on marginalized communities. By exposing these patterns, the index aims to empower stakeholders, policymakers, and advocacy groups with actionable insights to address online hate effectively, promote accountability, and advocate for safer digital spaces.

Methodology

The data for this report was collected from multiple social media platforms, using tailored methods to identify and analyze hateful and violent content, the approach varied depending on the platform to ensure the most effective data gathering process.

For X, a carefully selected list of hateful and violent keywords and hashtags was used to track and collect relevant posts. For Facebook, the process began with compiling a list of hundreds of various pages. Posts from these pages were gathered, and the dataset was expanded to include the comments associated with these posts, offering a broader view of interactions and discussions.

Once collected, the data was analyzed using a combination of deep learning classifiers and large language models (LLMs). Our internal classifiers were used to determine whether a piece of text contained hate speech and to identify the motives of violence present. Additionally, advanced LLMs, similar to GPT models, were employed to classify content into more nuanced categories beyond hate speech, providing deeper insights into the nature of the discourse.

The numbers presented in the report are the result of automated predictions. While our system is generally reliable, it operates with an estimated accuracy rate of 84%, meaning some margin of error exists.

As of January 28, 2024, changes in X's data access policies required us to adjust our data collection methods. Any fluctuations in the number of recorded cases since this date should not be interpreted as a decrease in hate speech or violent content but rather as a result of these policy changes.

Findings

In 2024, a total of 12,482,041 pieces of content in Hebrew were identified by 7amleh's language model as violent and hateful on X and Facebook, equating to an average of 23.6 instances per minute.

Cases Timeline
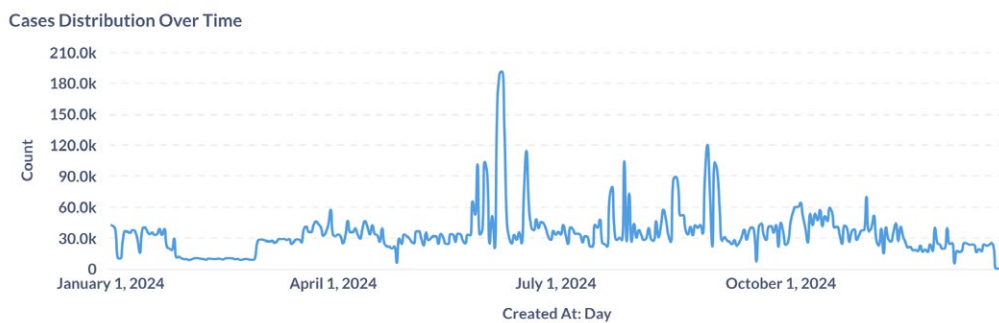
The data reveals distinct spikes in hateful and violent content at the beginning of June, mid-July, and towards the end of August, each correlating with significant socio-political events. The most pronounced spike occurred in June, largely driven by intense Israeli discourse surrounding the U.S.-proposed deal on Gaza and the subsequent UN Security Council resolution on a ceasefire. These events sparked widespread online discussions, which frequently escalated into incitement and hate speech. It is also worth noting that changes in data collection rates beginning in September, due to increased collection costs, may have affected the ability to detect additional spikes later in the year. Nonetheless, the data from June through August provides a compelling illustration of how major geopolitical events directly

influence the volume and nature of online hate, with users reacting to real-world developments in real-time.

Platform Distribution
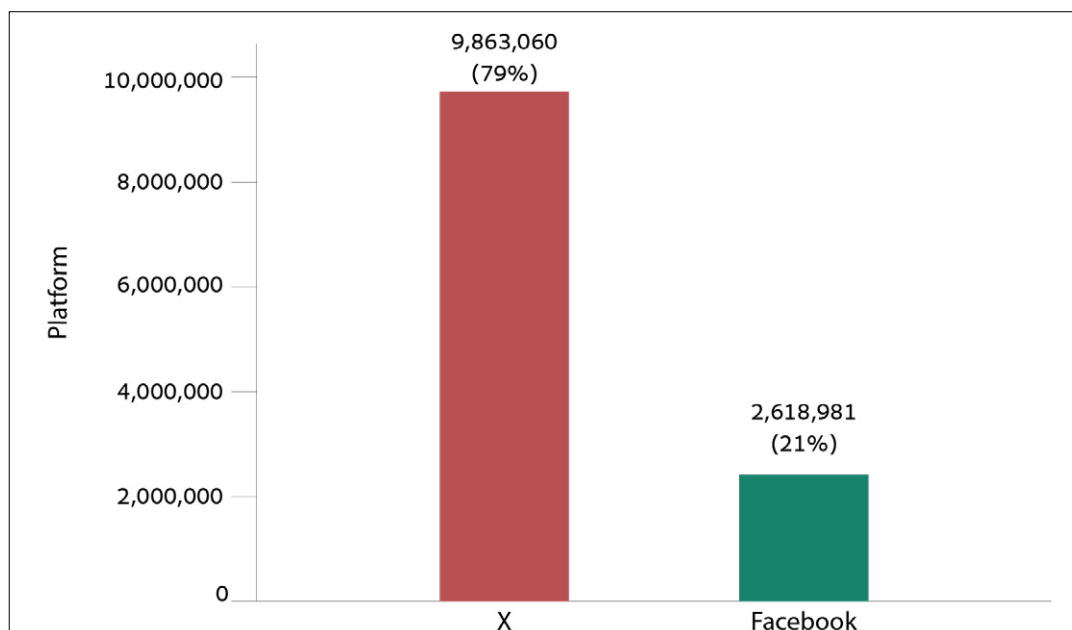
The data reveals a significant disparity in the distribution of hateful and violent content between X and Facebook. X accounted for 79% of the identified content, totaling 9,863,060 instances, while Facebook comprised 21%, or 2,618,981 instances. This imbalance suggests that X serves as a more active platform for the spread of such content. For context, despite Facebook being more widely used—87% of Israeli users are on Facebook compared to just 37% on X—violent content remains disproportionately concentrated on X, highlighting its loose moderation policies. This huge contrast proves that X platform doesn›t moderate Hebrew violent content at all.

Motives of Hate
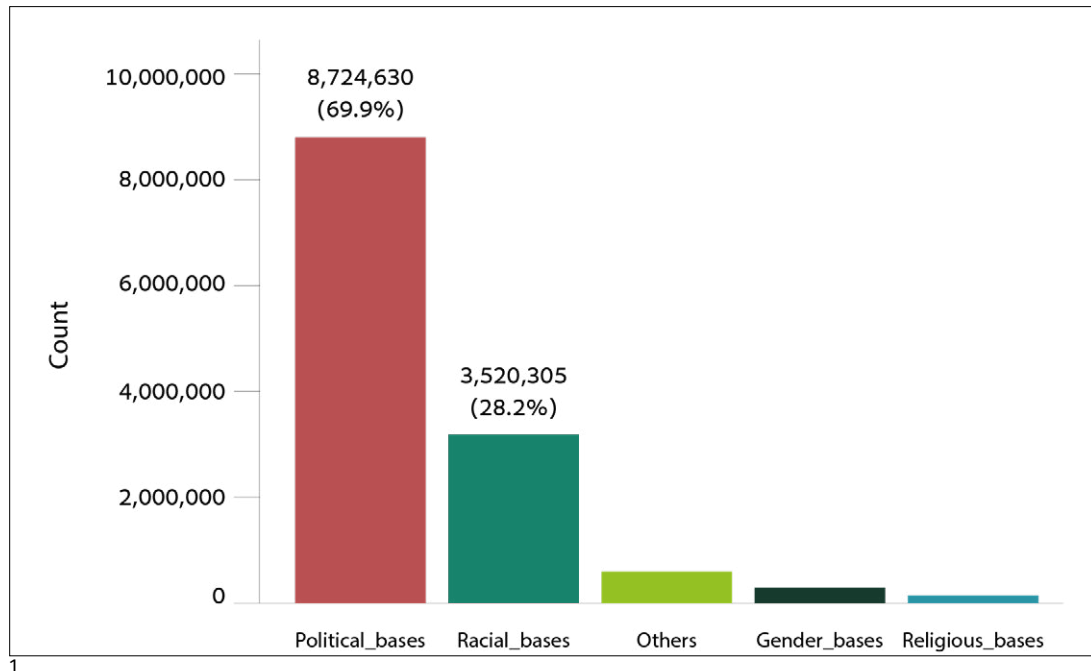
The data reveals that political motives dominate as the primary driver of hate speech, accounting for 8,724,630 instances, or the majority of the identified content. This suggests that hate and incitement are often rooted in political disputes, admist the context of occupation and genocide against Palestinians. Racial motives are the second-largest category, with 3,520,305 instances, highlighting the pervasive targeting of specific racial or ethnic groups, particularly Palestinians and other minorities. This underscores the entrenched racial biases present in online discourse. The significantly lower numbers for gender-based hate (83,553 instances) and other motives (110,026 instances) suggest that while these categories are less prominent, they remain critical areas of concern.

1   Religious bases: Motivated by hostility or bias towards a particular religion. Gender bases: Motivated by bias or hostility towards someone's gender identity or sexual orientation Racial bases: Motivated by bias or hostility towards a racial or ethnic group. Political bases: Motivated by opposition to or support for a political ideology, group, or figure. Other: Motivated by reasons not covered in the above categories.
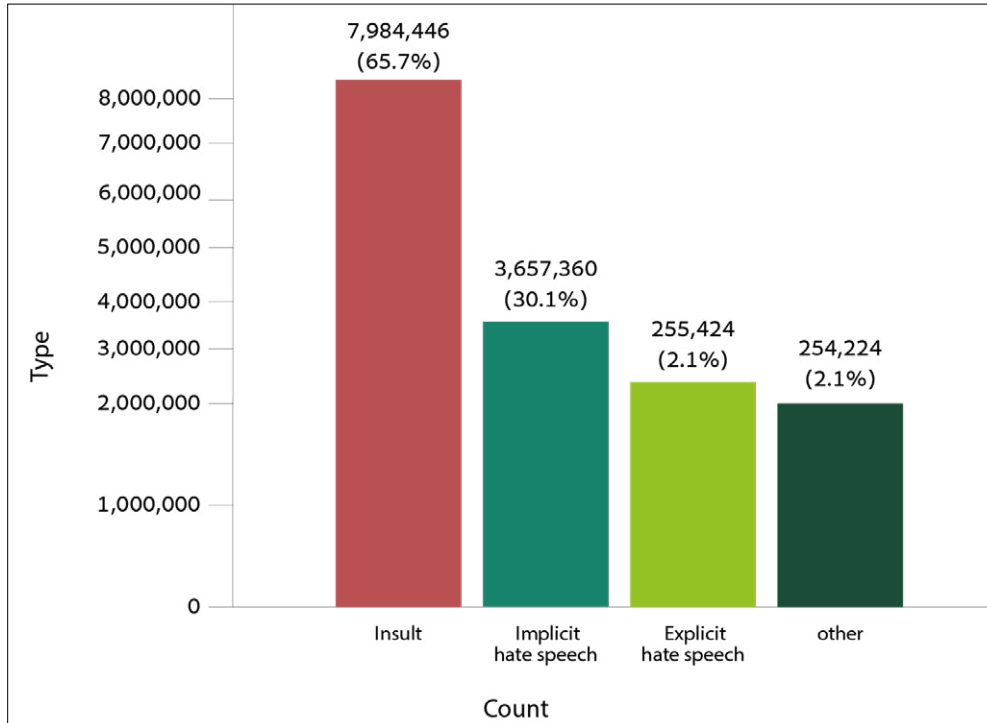
Religious Targeting Distribution
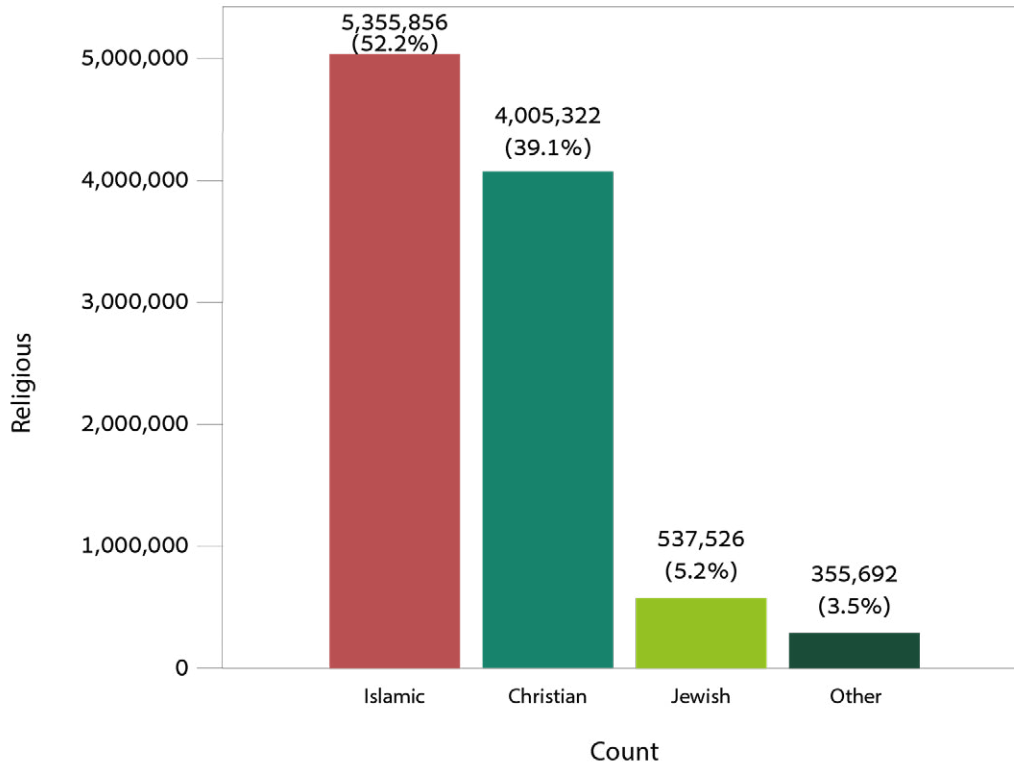
Distribution of Religion Across Speech Types
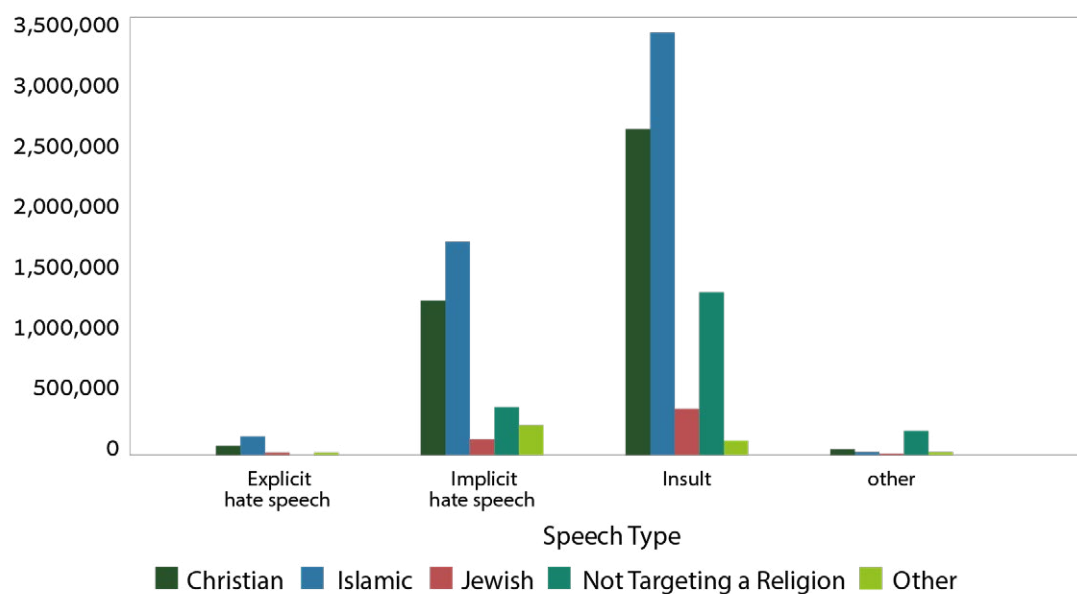
Hate Type Distribution

Platform Account Distribution
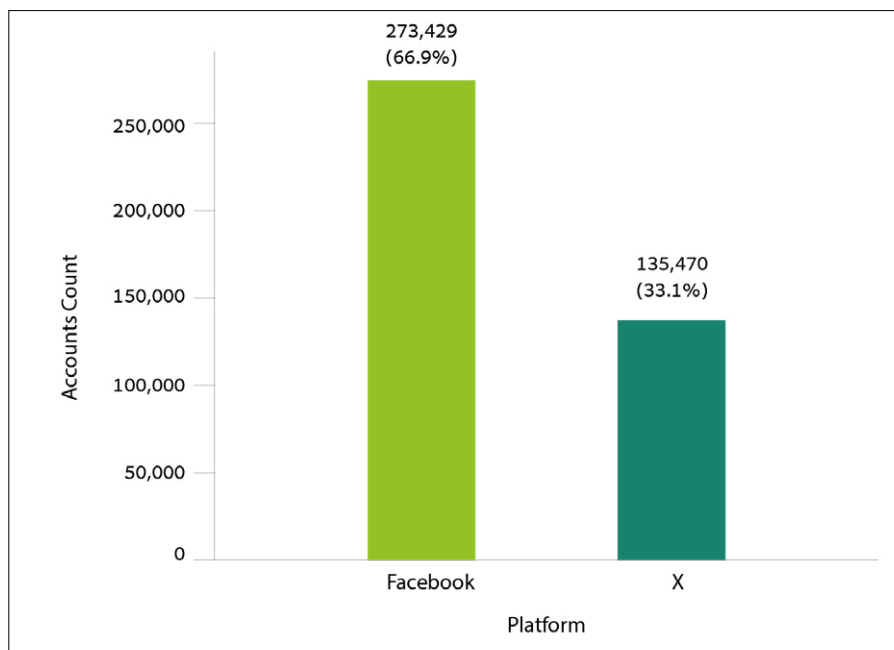
The chart highlights a striking contrast: although the number of users whose content was scraped on Facebook is twice as high as on X, the majority of violent content still comes from X. This suggests that, despite having fewer users in the dataset, X hosts a disproportionately high amount of violent content. This disparity indicates that X not only facilitates but also fails to curb the spread of such content, making it a more concentrated source of harmful material compared to Facebook. This is because Facebook is doing minimum content moderation in Hebrew, while X platform almost doesn't moderate Hebrew content.

Special Focus: Digital Violence against Jerusalemites

Throughout the year, violent speech and incitement specifically targeting Jerusalemites were prevalent across social media platforms. The analysis identified 8,484 posts, with the majority originating from X (7,480 posts), while Facebook accounted for 1,004 posts.

To identify such content, keywords referring to Jerusalemites in various contexts were analyzed, including terms such as:

"ירושלמים" (Jerusalemites), "תושבי ירושלים" (Residents of Jerusalem), "אזרחי ירושלים" (Citizens of Jerusalem), "פלסטינים מירושלים" (Palestinians from Jerusalem), "ערבים מירושלים" (Arabs from Jerusalem), "מוסלמים מירושלים" (Muslims from Jerusalem), "נוצרים מירושלים" (Christians from Jerusalem), "יהודים מירושלים" (Jews from Jerusalem), "מזרח ירושלים" (East Jerusalem), "תושבי מזרח ירושלים" (Residents of East Jerusalem), "סילואן" (Silwan), "שועפאט" (Shuafat), "העיר העתיקה" (The Old City), "ראס אל-עמוד" (Ras al-Amud), "ואדי ג'וז" (Wadi al-Joz), "בית חנינא" (Beit Hanina), "ירושלמי" (Jerusalemite, feminine), and "ירושלמית" (Jerusalemite, masculine).

These findings reveal the systemic targeting of Jerusalemites through incitement and hate speech, reflecting a troubling pattern of hostility directed toward individuals and communities in Jerusalem.

Timeline

Distribution of Jerusalem Related Records
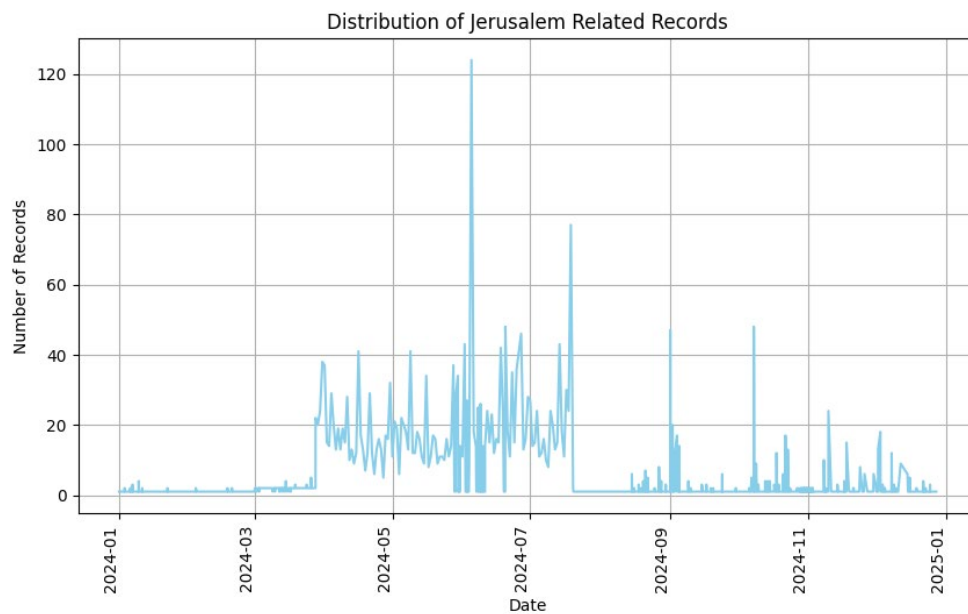
The timeline of spikes in hateful rhetoric against Palestinian Jerusalemites closely mirrors the patterns seen in broader trends of hateful rhetoric directed at Palestinians in general. This similarity is striking and suggests that Palestinians, regardless of their specific geographic location, are subjected to a uniform level of online hostility. Such parallel trends indicate that the targeting is not confined to a particular subgroup but is instead a part of a broader, systemic pattern of digital aggression aimed at Palestinians as a whole.

Content Type Distribution by Platform

■ Gender_base ■ Others ■ Political_bases ■ Racial_bases ■ Religious_bases
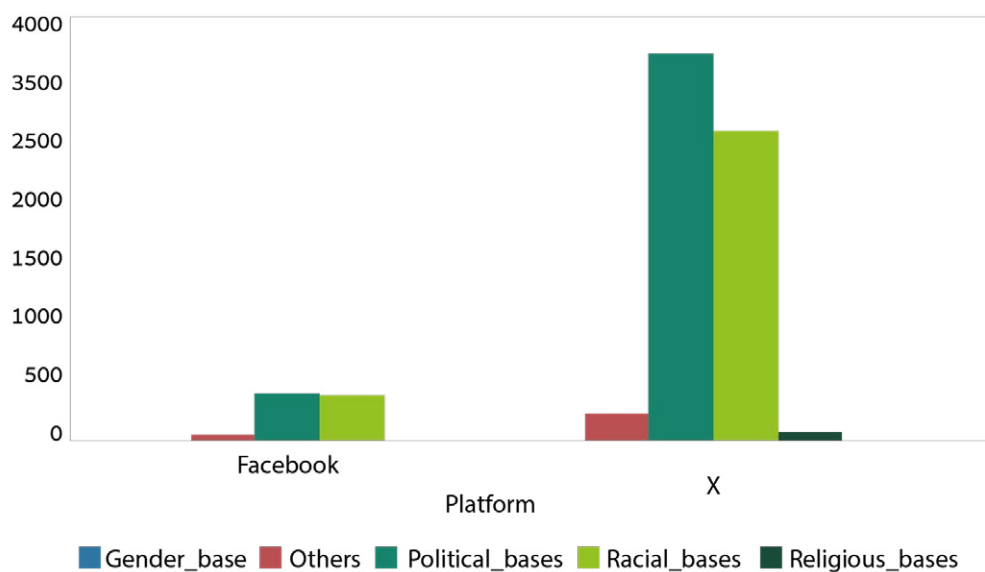
Content Type Distribution By Religion
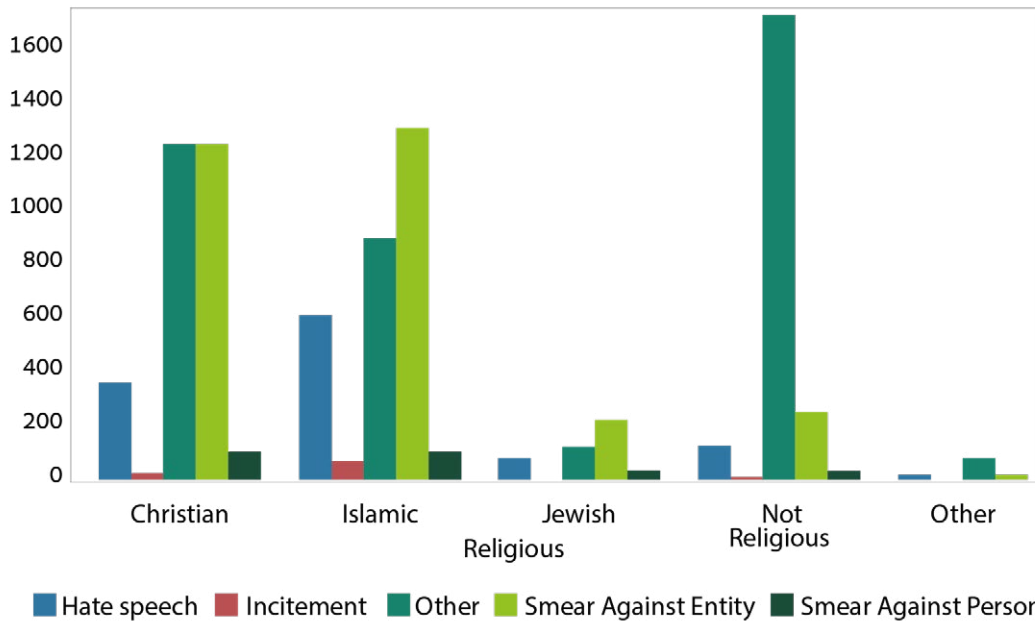
The chart highlights a significant volume of hateful content directed at Palestinian Jerusalemites, with a focus on attacks based on their nationality and religion. These figures not only reveal the extent of online hostility but also reflect the broader power dynamics in an occupied city. The prevalence of such content underscores the ongoing challenges faced by those living under occupation, where hate speech is often used as a tool to further marginalize and dehumanize Palestinian Jerusalemites.

These findings underscore the problematic nature of violent speech and incitement targeting Jerusalemites across Facebook and X. The volume of hate speech directed at Palestinian Jerusalemites aligns with broader patterns of digital hostility toward Palestinians, indicating that online incitement is not isolated but part of a systemic trend.

Furthermore, the targeting of Palestinian Jerusalemites is not limited to general hate speech; rather, it reflects deep-rooted power dynamics within an occupied city, where digital aggression serves as a tool for marginalization, control and dehumanization.

Most used words in the detected violent content:

"טרור" (**Terror**), "מחבלים" (**Terrorists**), "מטומטם" (Dumb), "רצח" (**Murder**), "מוות"(**Death**), "האויב" (**The enemy**), "מחבל" (Terrorist), "טבח" (**Massacre**), "להשמיד" (**To destroy**), "פלסטינית" (**Female Palestinian - woman/girl**), "רוצחים" (**Murderers**), "אפס" (**You're nothing/loser/a zero**), "זבל" (**Scum**), "טיפש" (**Idiot**), "שנאה" (**Hate**), "נשק" (**Weapon**), "הפלסטינים" (**The Palestinians**),"לרצוח" (**To kill**), "להרוג" (**To kill**), "אלימות"(**Violence**), "שקר" (**Lies**).

Entities/Peoples mentioned:

"ערבים" (**Arabs**), "מוסלמים" (**Muslims**), "הפלסטינים" (**Palestinians**), "עזה" (**Gaza**), "חמאס" (**Hamas**), "איראן" (**Iran**).

Special Focus: Gloating over rockets falling on Palestinian towns

During the war on Gaza, many Palestinian citizens of Israel were killed & injured by rockets launched from Gaza, Lebanon, or Iran. Israeli users often gloated over the victims, posting comments that expressed pleasure and satisfaction simply because the victims were Palestinian Arabs or suggesting that Palestinians deserved such a fate. The following graph presents samples of this hateful content.
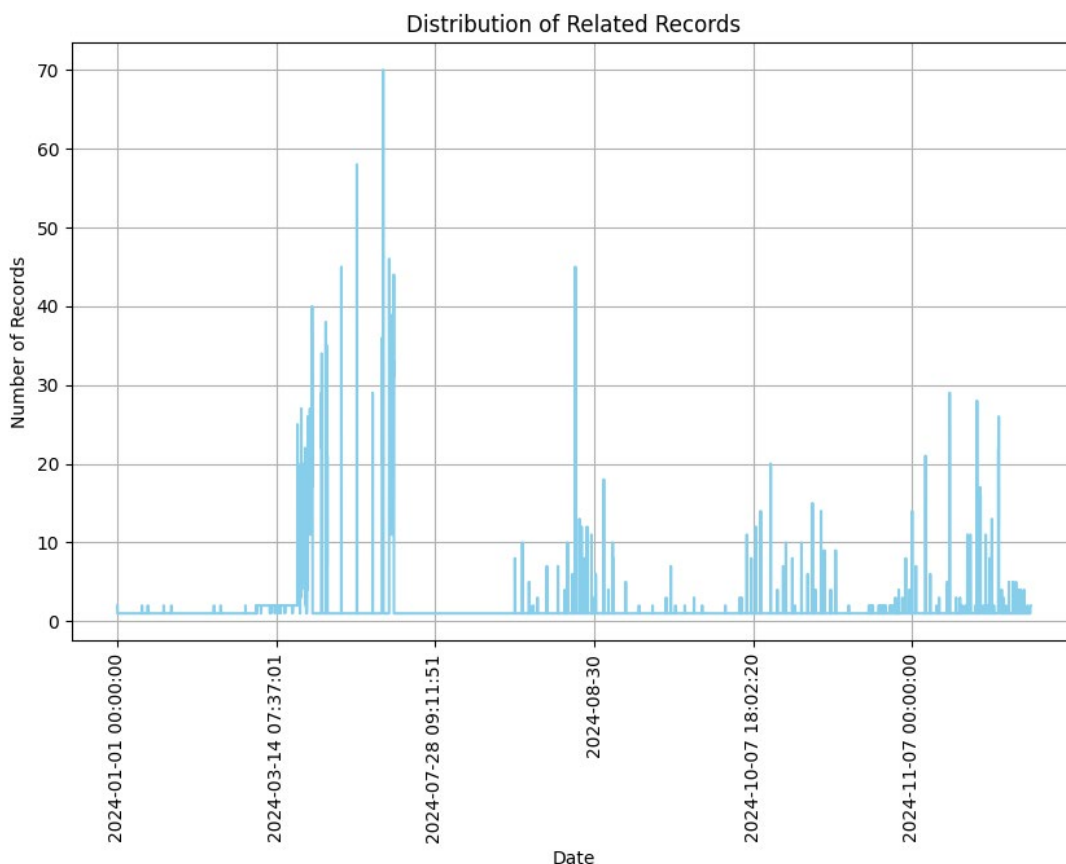
Post Examples
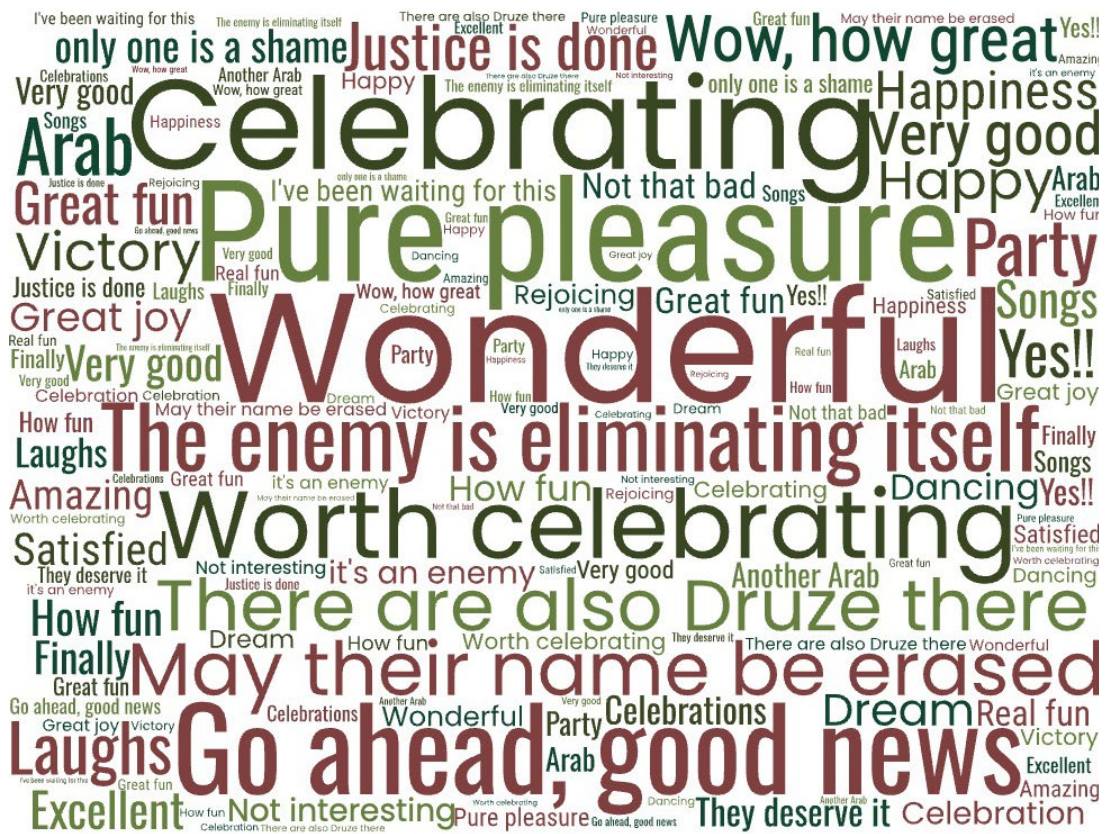
איזה כיף"- "כיף" (**Arab**), "יש שם גם דרוזים" (**There are also Druze there**), "ערבי"-"ערבייה" (**Arab**), "כיף אמיתי" - "גדול" (**How fun**)-(**Great fun**)-(**Real fun**), "היידה, בשורות טובות" (**Go ahead, good news**),"האויב מחסל את עצמו" (**The enemy is eliminating itself**) "מגיע להם"-"מגיע (**They deserve it**), "לא כל כך נורא בסך הכל אוייב רק חבל שרק אחד (**Not that bad, after all, it's an enemy, only one is a shame**) "לא מעניין" (**Not interesting**), "לו"- "מגיע לה ימח" (**Another Arab**) "עוד ערבי ערבייה" (**Another Arab**) "שמם" (**May their name be erased**), "ימח שמו"-"ימח שמה שמח"-"-", "חגיגה"-"חגיגות"-"חוגגים"-"חוגגות" (**Celebration-Celebrations-Celebrating**), "-"שמח (**Arab**)

שמחה מאוד"-"שמח מאוד"-"שמחה"" (**Happy**),"כיף אמיתי" "כיף גדול", "איזה כיף", **-How fun**
Great fun - Real fun (מצוינת" -"מצוין", "סוף סוף", (**Finally**), (**Excellent**) "טוב" -"הגיע הזמן
מאוד מרוצה"-"מרוצה"-"מדהים"-מאוד""(It's time), (**Very good**), (**Amazing**), (**Satisfied**), "צדק
סיפוק גדול" - "סיפוק" -"נעשה", (**Justice is done**), (**Satisfaction**), (**Great satisfaction**),
"מסיבה" - "שירים" - "ריקודים" (**Dancing**, **Songs**), (**Party**), "צוהלות"-"צוהלים"-"שמחה גדולה
(**Yes!!**), (**Great joy**), "על הכיפאק"(**Rejoicing**), (**Worth celebrating**) כדאי לשמוח", "יששש"
(**Great**), "צחוקים"-"אושר", "ניצחון" (**Victory**) - (**Happiness**) - (**Laughs**),"-איזה יופי" בואנה
(**Pure**), (**Dream**), (**For real**), (**Wow, how great**)"חלום" -"אשכרה"-"פשוט תענוג - "צרוף
for waiting been I've),נהדר", "איזה יופי" מעולה", "טוב מאוד", "כמה חיכיתי לזה", (**pleasure**
this), (**Very good**), (**Great**), (**Wonderful**).

The number of posts identified as gloating is 9,289. However, this figure is not necessarily representative—it reflects only what 7amleh was able to detect, while the actual scale of the phenomenon may be much larger.

Timeline

Distribution of Related Records

Content Type Distribution by Platform

Legend: Gender_base, Others, Political_bases, Racial_bases, Religious_bases

Content Type Distribution by Religion
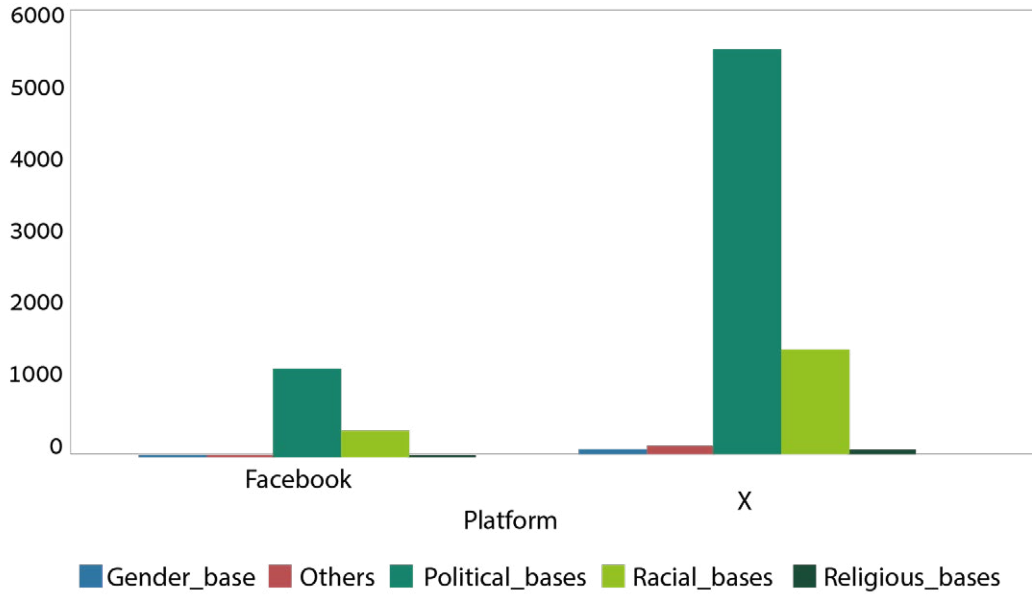
Legend: Anti-Palestinian hate, Hate speech, Incitement, Other, Smear Against Entity, Smear Against Person
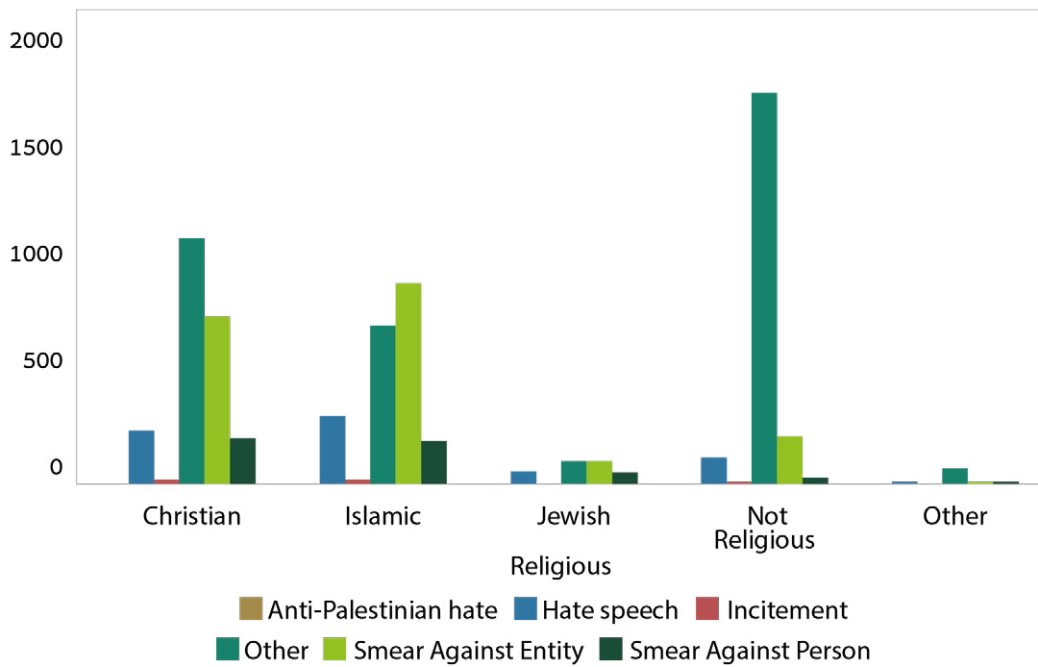
It is noteworthy that, while most of the content consists of gloating, many posts also included direct hate speech, incitement, or smearing against Palestinian victims. The majority of this content was rooted in political or racial motives, with some instances also reflecting religious hostility.

The findings reveal a pattern of online hate speech and incitement, where Israeli users frequently gloated over the deaths of Palestinian citizens of Israel during the war on Gaza. The scale of this phenomenon, with at least 9,289 identified posts, underscores the widespread normalization of digital hostility toward Palestinians. However, this number likely represents only a fraction of the actual volume of such content, given the limitations of detection. Beyond expressions of gloating, many posts also contained direct hate speech, incitement, and defamatory rhetoric, often driven by political, racial, or religious animosity. This demonstrates how digital platforms are being used to dehumanize Palestinian victims, reinforcing societal divisions and legitimizing violence against them.

Conclusion

The Racism & Incitement Index 2024 paints a picture of the systemic spread of online hate speech and incitement, particularly targeting Palestinians. With 12,482,041 instances of violent and hateful content recorded in Hebrew over the past year—an average of 23.6 instances per minute—the findings reveal the alarming scale of digital hostility, which mirrors and exacerbates real-world violence. The data makes clear that the bulk of this content is directly related to the war on Gaza, demonstrating how real-world violence is reflected and amplified in the digital sphere. Online platforms have become battlegrounds where hate speech flourishes, intensifying hostility and reinforcing narratives of dehumanization.

The data demonstrates that political, racial, and religious motives are the primary drivers of online incitement, with spikes in hate speech closely correlating with major geopolitical events, especially the war on Gaza. The report highlights two critical areas of concern: the targeting of Palestinian Jerusalemites and the gloating over the deaths of Palestinian citizens of Israel. These patterns underscore how digital platforms are being weaponized to dehumanize Palestinian communities, reinforce societal divisions, and legitimize acts of violence.

The disproportionate concentration of violent content on X compared to Facebook highlights the ongoing failure of social media companies to effectively moderate hate speech. While Meta›s policies and enforcement mechanisms remain inadequate in curbing incitement—allowing harmful narratives to spread unchecked—X makes little to no effort to flag violent Hebrew content. Additionally, growing concerns surround Meta's recent changes in content moderation policies, particularly its plan to rely on AI language models to flag only high-severity content. This shift raises fears that the spread of violent content, especially in Hebrew, will continue unchecked at alarming levels.

The findings of this report stress the urgent need for intervention from governments, civil society, and tech companies. Without decisive action, the unchecked spread of digital hate will continue to fuel violence, discrimination, and marginalization, both online and offline. Addressing this crisis requires stronger content moderation policies, enhanced accountability mechanisms, and meaningful engagement with affected communities to ensure that online spaces do not serve as amplifiers of real-world harm.

Recommendations

- Online platforms must enhance content moderation systems to accurately detect and remove violent Hebrew content, ensuring consistent enforcement across all languages.

- Platforms should urgently address the proliferation of violent Hebrew content, implementing targeted measures to curb its spread and prevent further harm.

Meta and X must:

- Conduct independent, publicly available human rights impact assessments to evaluate and mitigate platform-related harms.

- Allocate sufficient resources, including Hebrew-language expertise, to strengthen moderation and enforcement mechanisms.

- Establish clear, transparent, and timely mechanisms for addressing digital rights violations reported by civil society organizations.

- Engage in regular, structured dialogue with Palestinian civil society and affected communities to assess and mitigate platform-related harms.

- Protect user data privacy, ensuring that personal data is not weaponized against vulnerable populations, including Palestinians.

- Prevent the use of platform technologies in ways that facilitate war crimes, crimes against humanity, or genocide.

Contact 7amleh:

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

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