

BEYOND THE SLUR:

**How Online Misogyny in the Levant Targets Men,
Erases Women, and Escapes Platform Policy**



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



October 2026



October 2026

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

Beyond the Slur: How Online Misogyny in the Levant Targets Men, Erases Women, and Escapes Platform Policy

A data-driven analysis of technology-facilitated gender-based violence (TFGBV) in Levantine Arabic online spaces

Author: **Jalal Abukhater, Policy Manager, 7amleh**

Data Collection: **Rama Salahat, Technology Officer, 7amleh**

Designer: **Louise Dib, studio Chimbo**

This brief was produced under the Safety Voices project, with support from the Association for Progressive Communications (APC). All figures are drawn from 7amleh's Levantine Arabic TFGBV monitoring corpus (58,798 posts, 1 January - 22 April 2026, Facebook and X/Twitter). For methodology details, further breakdowns, or access to the underlying dataset, contact 7amleh's policy team.

This version is licensed under the following International License: Attribution-NonCommercial-NoDerivs 4.0 International

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

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

Contact us:

Email: info@7amleh.org

Website: www.7amleh.org

Telephone: +972 (0) 7740 20670

Find us on social media: 7amleh



Table of contents

Executive Summary	<u>4</u>
Introduction	<u>4</u>
Methodology	<u>5</u>
What forms does TFGBV take?	<u>6</u>
Findings	<u>7</u>
Misogyny is the register, not always the target	<u>7</u>
Who is targeted and how	<u>8</u>
LGBTQ+ targets	<u>9</u>
The Severity Paradox	<u>10</u>
Dispersed abusers and relentless targets	<u>11</u>
Platform patterns	<u>12</u>
Regional implications	<u>13</u>
Conclusion	<u>14</u>
Recommendations	<u>15</u>
For platforms	<u>15</u>
For civil society	<u>15</u>
For donors	<u>15</u>



Executive Summary

Most of the misogyny in this corpus is not aimed at women at all. Technology-facilitated gender-based violence (TFGBV) is usually measured by counting attacks on women. That measure misses most of the problem. In a 58,798-post corpus of severe abuse in Levantine Arabic, 7amleh finds that 55.2% of posts use language that degrades women, but only 21.3% are actually directed at a woman.

Nearly half of all misogynistic language in this corpus (46.3%) is aimed at men, using a woman who is not present (a mother, sister, or wife) as the instrument of the attack. Standard platform categories built around a target being harassed, or a protected group being attacked as a group, cannot see this pattern, because it is neither.

This brief sets out what the data can and cannot support. It documents the forms TFGBV takes, how misogyny functions as a weapon between men, the distinct exclusionary pattern faced by LGBTQ+ people, a structural blind spot we call the severity paradox, and how abuse accumulates against individual women even when no single account reaches a reporting threshold. It closes with policy recommendations addressed to platforms, regulators, and funders, grounded only in what this dataset demonstrates.

The brief also makes a capacity argument. Producing a regionally accurate, dialect-specific snapshot of this kind requires sustained scraping, Levantine dialect classification, and multi-level severity coding across a four-month window. 7amleh's AI-assisted monitoring tools can do this at scale, and this brief is offered as a demonstration of what that tool makes visible that generic, standard-Arabic or English-language moderation tools do not.

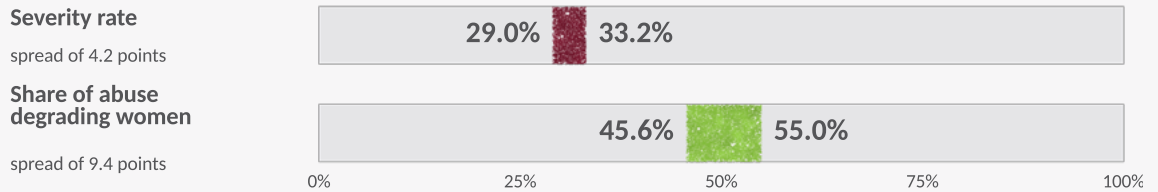
Introduction

TFGBV research on Arabic-language content is frequently conducted in Modern Standard Arabic or aggregated across the whole Arabic-speaking world, which flattens dialect-specific insult vocabularies, honour-based idioms, and region-specific political triggers into a single undifferentiated category. This corpus instead isolates Levantine Arabic specifically: of 252,837 Arabic-language GBV records monitored, 114,791 were classified as Levantine dialect, and of those, 58,798 scored above the mildest severity tier and form the corpus analysed here.

This matters for policy for a concrete reason: 7amleh's regional-arena analysis finds that the abuse repertoire is remarkably stable across the different political contexts Levantine speakers discuss, Syria, Lebanon, Palestine, Iran, and the Gulf all show severity rates within a few points of one another (29.0%–33.2%), and the share of abuse degrading women varies only between 45.6% and 55.0%. A tool, moderation ruleset, or policy ask built for one Levantine context is very likely to transfer to the others. That transferability is itself a policy-relevant finding, and it is only visible because the corpus was built at the dialect level rather than the national or platform level.

The abuse repertoire is stable across political contexts

range across Syria, Lebanon, Palestine, Iran and the Gulf



Methodology

This policy brief is built entirely from the 7amleh-monitored corpus of 58,798 abusive Levantine Arabic posts (37,673 accounts; Facebook and X/Twitter; 1 January – 22 April 2026), filtered by dialect and severity from a larger set of 252,837 Arabic-language gender-based-violence records. Every figure below is drawn directly from that dataset.

A representative, dialect-accurate picture of TFGBV cannot be assembled only from keyword lists, it requires continuous collection, dialect classification, and multi-level human/automated severity coding, sustained over time. 7amleh's monitoring pipeline documented 252,837 Arabic-language GBV records before dialect filtering was even applied.

Severity coding refers to how posts are scored on a three-level scale: Level 1 covers mockery, appearance or age-based insults, and general belittlement. Level 2 covers explicit sexual insults, sexual humiliation, honour or virginity-based defamation, accusations of immorality, moral guardianship, homophobic slurs, and accusations of treason or collaboration. Level 3 covers threats of physical or sexual harm, death threats or incitement, extortion, threats to expose private information or images, threats against family, and calls for others to cause harm. This brief analyses only posts scoring above Level 1, the most obviously severe content, which is what makes the 58,798-post corpus a high-confidence sample rather than an inflated one.

252,837

Arabic-language gender-based-violence records
documented by 7amleh's pipeline, before any dialect filtering



▼ filtered by dialect and severity

58,798

abusive Levantine Arabic posts
the corpus analysed throughout this brief



23.3% of the records

37,673

accounts

2 platforms

Facebook and X/Twitter

1 Jan – 22 Apr 2026

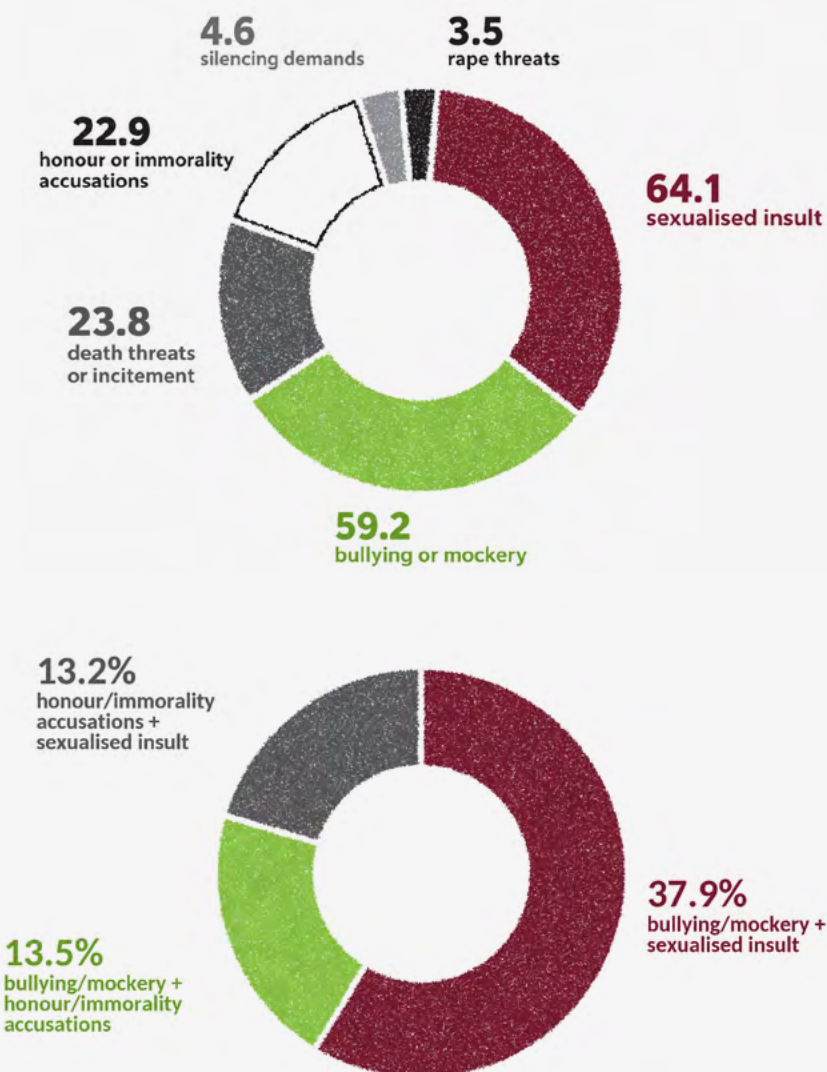
monitoring period



What forms does TFGBV take?

Abuse in this corpus is compound rather than singular: the average post carries 2.7 distinct abusive forms. The most prevalent forms are sexualised insult (64.1%; 37,671 posts) and bullying or mockery (59.2%; 34,826 posts), followed by death threats or incitement (23.8%; 13,978 posts) and honour or immorality accusations (22.9%; 13,486 posts). Rape threats (3.5%; 2,087 posts) and silencing demands (4.6%; 2,698 posts) are comparatively rare in isolation but concentrate heavily on specific groups, discussed below.

The most common two-form combination is bullying/mockery paired with sexualised insult (37.9% of the corpus), followed by bullying combined with honour/immorality accusations (13.5%) and honour/immorality accusations combined with sexualised insult (13.2%). This compounding matters for moderation design: a system that screens for a single flagged term per post will systematically under-detect content that only becomes severe in combination.



Findings

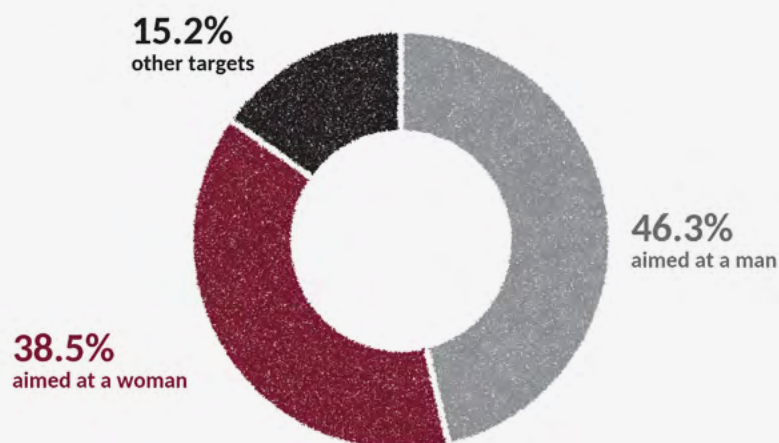
Misogyny is the register, not always the target

32,448 posts, constituting 55.2% of the corpus, use language that degrades women. Only 12,505 of them (21.3% of the corpus) are actually aimed at a woman. 15,017 posts, constituting 46.3% of all misogynistic content, are aimed at a man. That gap is the finding this policy brief is built around.

In these posts, a woman who is not present, not named, and not the addressee (the target's mother, sister, or wife) is invoked as the instrument of an attack whose intended injury lands on him. Where women are the actual target, the pattern reverses sharply: 87.7% of posts aimed at women degrade women, confirming that when women are addressed directly, gendered insult is close to the default register used against them.

This is a structural, not incidental, gap in platform policy. Harassment enforcement requires an identifiable target being harassed; hate-speech enforcement requires a protected group attacked as a group. Content that degrades women in order to wound a man satisfies neither test cleanly, even though it is nearly half of all misogynistic content in the corpus.

The mechanism runs in both directions through a shared honour lexicon. 9,770 posts, 27.8% of everything aimed at men, use honour-system terms (ديوث, عرص, قواد, دياثة, بلا شرف) that all denote a man who has failed to control the sexual conduct of the women around him. 3,279 posts combine this with misogynistic language in the same post. The abuse of men here is not simply borrowing women's bodies as ammunition, it is enforcement of the same honour system from the other side.





Who is targeted and how

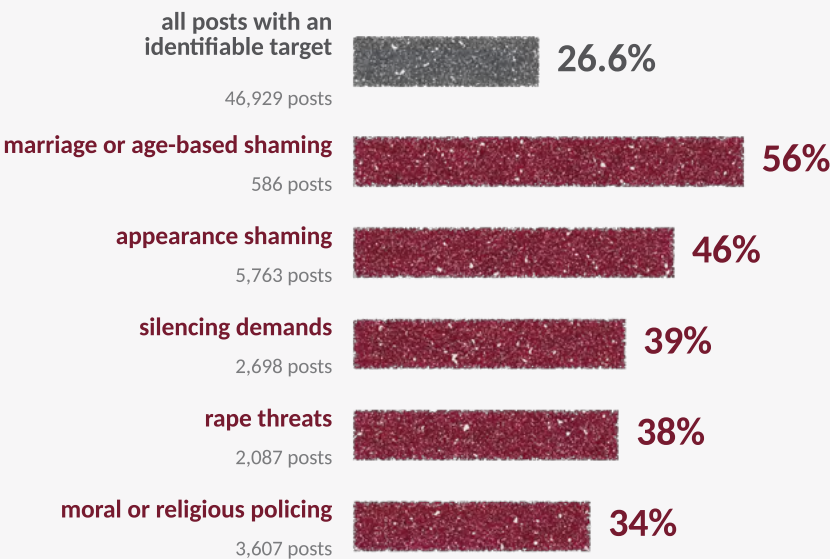
Among the 46,929 posts in the corpus with an identifiable target, 26.6% are aimed at a woman. This is the baseline against which specific forms should be read. Some forms concentrate on women far above that baseline: marriage or age-based shaming (56.0% of its 586 posts aimed at women), appearance shaming (46%; 5,763 posts), silencing demands (39%; 2,698 posts), rape threats (38%; 2,087 posts), and moral or religious policing (34%; 3,607 posts). Other forms concentrate on men well above the baseline for men, notably death threats/incitement and homophobic or transphobic slurs, discussed below.

The qualitative difference is as important as the quantitative one: men are attacked overwhelmingly about killing them; women are attacked about their bodies, their reputations, their families, and their right to remain in the space at all. Where a stated context is available (38.3% of the corpus, 22,494 posts), women’s abuse is more often triggered by appearance or personal life (9.4% vs 4.2% for men), relationship or marital status (3.2% vs 0.4%), or mere presence in public space (1.8% vs 0.5%), while men’s abuse is more often triggered by political expression (38.1% vs 23.4% for women).

On finer target classification, distinguishing, for instance, a woman journalist from a woman politician, activist, or influencer, the current corpus does not code target occupation or public role, so this brief cannot report a breakdown by professional category. The one indicative signal available is the sustained-targeting list (discussed in a later section), where the accompanying data note observes that the most persistently targeted women are disproportionately Lebanese media and entertainment figures, a finding about visibility, not necessarily about who is most endangered.

Aimed at a woman

share of posts aimed at a woman, by form of abuse (posts with an identifiable target)



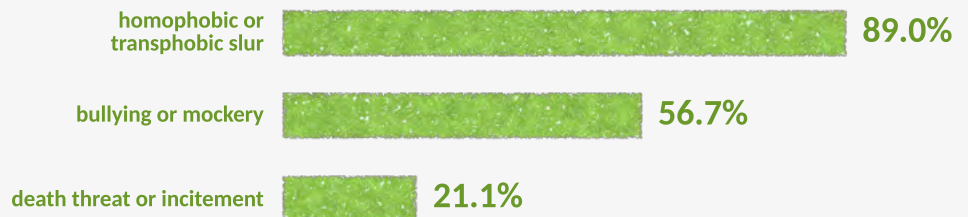
LGBTQ+ targets

LGBTQ+ individuals are a small share of directly identified targets (829 posts, 1.4% of the corpus) but show a markedly different abuse profile from either women or men. 89.0% of posts aimed at an LGBTQ+ person contain a homophobic or transphobic slur, 56.7% bullying or mockery, and 21.1% a death threat or incitement.

A separate and much larger signal sits alongside the direct-target count: 4,013 posts (6.8% of the whole corpus) use homophobic or transphobic language generally, most often deployed against men to question their masculinity (مخنث, شاذ) rather than at an identified LGBTQ+ person a pattern identical to the misogyny-against-men finding above. Read together, the data describes a goal of expulsion from the space rather than escalation to violence: severity-ranked moderation queues will rank this content low, exactly where it is most effective at achieving its purpose.

Aimed at an LGBTQ+ person

abuse profile of the 829 posts aimed at an LGBTQ+ person (1.4% of the corpus)



4,013

posts (6.8% of the whole corpus) use homophobic or transphobic language in general



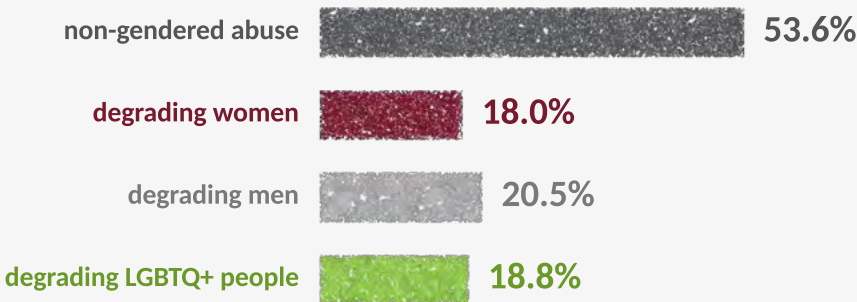
The Severity Paradox

Measurement itself seems to deprioritise gendered abuse. Non-gendered abuse in this corpus reaches the highest severity tier at 53.6%. Abuse that degrades women reaches that tier at only 18.0%, about a third of the rate, while abuse degrading men (20.5%) and abuse degrading LGBTQ+ people (18.8%) sit closer to the women’s rate than to the non-gendered rate. Notably, target gender alone barely moves this number (24.8% for men, 23.4% for women); it is the kind of abuse, not the identity of the person receiving it, that the severity scale downgrades.

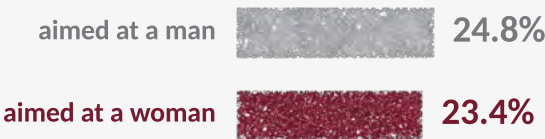
The cause is structural rather than a hidden bias against any one group. Severity scales are built around proximity to physical violence, and political or sectarian threat language dominates the top tier, while gendered abuse concentrates in sexual degradation and reputational destruction, content which is defined in a proximity-to-violence scale scores as mid-tier. A moderation queue designed strictly by severity score will therefore surface conflict speech first and gendered abuse last, systematically, every time, regardless of intent. The vocabulary evidence in this corpus reinforces the same conclusion from a second angle: the highest-severity, non-gendered word cloud is built almost entirely from political and sectarian terms, with almost no gendered terms present.

The Severity Paradox share of posts reaching the highest severity tier

By kind of abuse



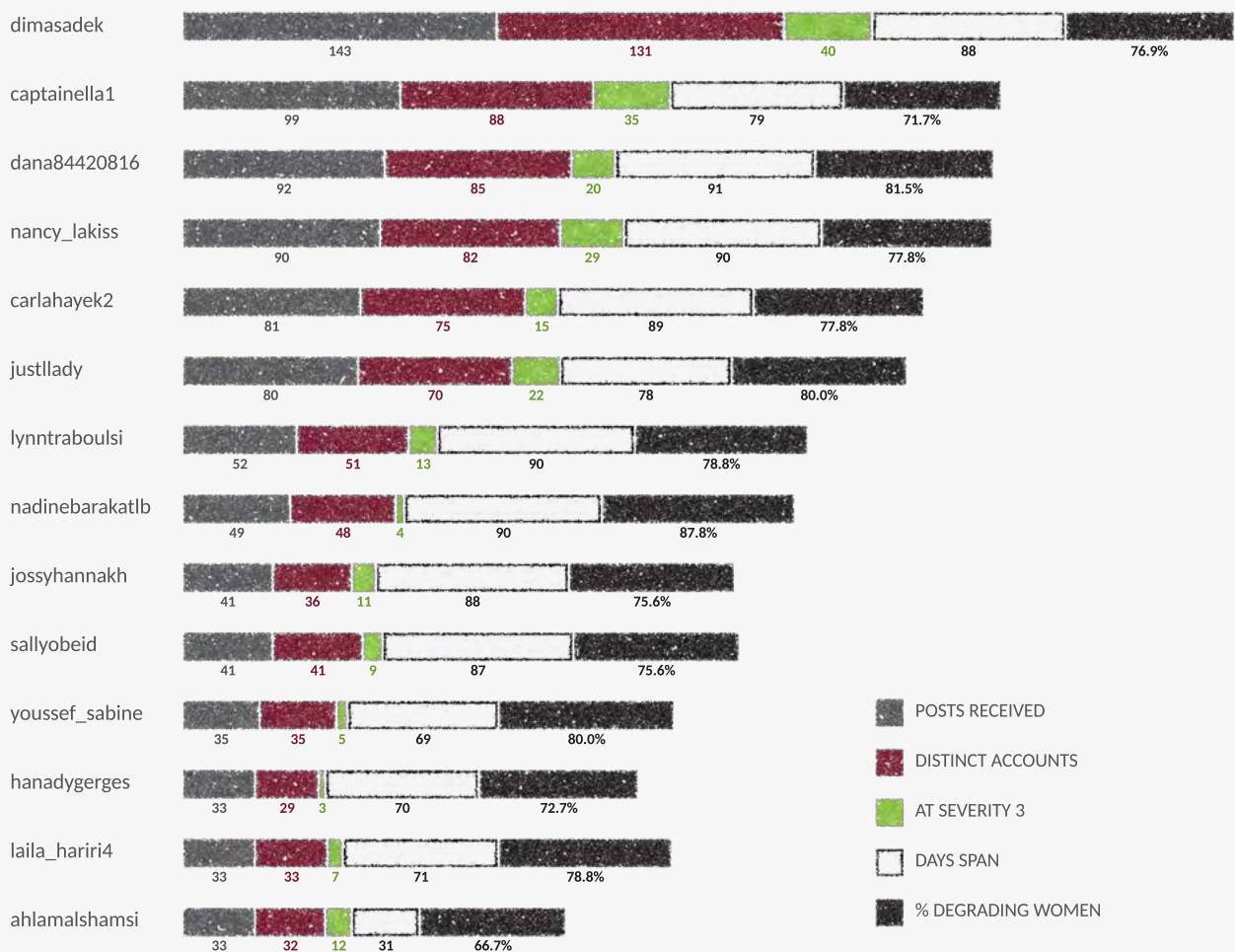
By target gender



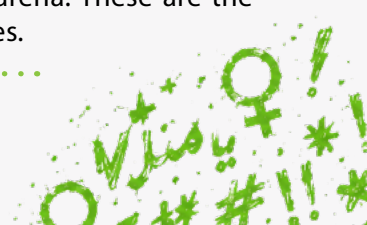
Dispersed abusers and relentless targets

Abuse in this corpus is generally dispersed rather than coordinated. The top 1% of accounts produce only 9.2% of posts, and 28,814 accounts (76.5% of all 37,673 accounts in the corpus) post exactly once. Yet those single-post accounts together account for 49.0% of all abuse. There is no single brigade to identify and ban; there is a large, low-effort base of one-time contributors.

This dispersion is precisely what allows sustained targeting to go undetected by account-based enforcement. 7amleh identifies 998 targets who received abuse from five or more distinct accounts across thirty or more days; 137 of them are women. For the woman receiving it, this is a months-long campaign; for a platform's enforcement system measuring violations per account, it is a hundred-plus separate first-time offenders, not one of which trips a repeat-offender threshold. The table below illustrates the pattern with the most heavily targeted individual women in the corpus.



Note: The sustained-target list is dominated by Lebanese media and entertainment figures, consistent with Lebanon having the highest woman-target share by arena. These are the corpus's most visible women, not necessarily its most endangered ones.





A review of the women on this list points to a shared profile rather than a random sample: most are vocal, publicly opinionated figures who engage regularly with contested regional political questions, such as Hezbollah, Iran, Israel, and Syria chief among them, largely through work in media or commentary. Several occupy positions that run counter to majority sentiment in their own context, such as commentators known for pro-Israel or anti-Hezbollah views all appear on the list. One entry, captainella1, is a female Arabic spokesperson for the Israeli Army. In one or two cases on the list, 7amleh was not able to authenticate whether the account is in fact operated by the woman it purports to represent.

This suggests that sustained volume is not randomly distributed even within the set of women who receive abuse: it concentrates on women whose public political positioning is itself contested, and whose visibility and engagement levels likely also drive the volume of replies they draw. We note this as an observation about who accumulates volume in this corpus, not as a claim about the relative merits of any of the political positions involved.

Platform patterns

Facebook is a small share of the corpus (8.9%) but a qualitatively distinct one: 36.3% of Facebook abuse reaches the highest severity tier, against 28.2% on Twitter/X, and 59.2% of Facebook abuse degrades women against 54.8% on Twitter/X. Facebook also carries disproportionately more honour/immorality accusations (34.4% vs 21.8%), doxxing (4.1% vs 0.8%), and silencing demands (7.7% vs 4.3%). Twitter/X abuse skews political; Facebook abuse is embedded in kinship and community networks and more often exposes private information, and on a platform where real names and shared social circles are the norm, the reputational and offline risk to a targeted woman is materially higher.

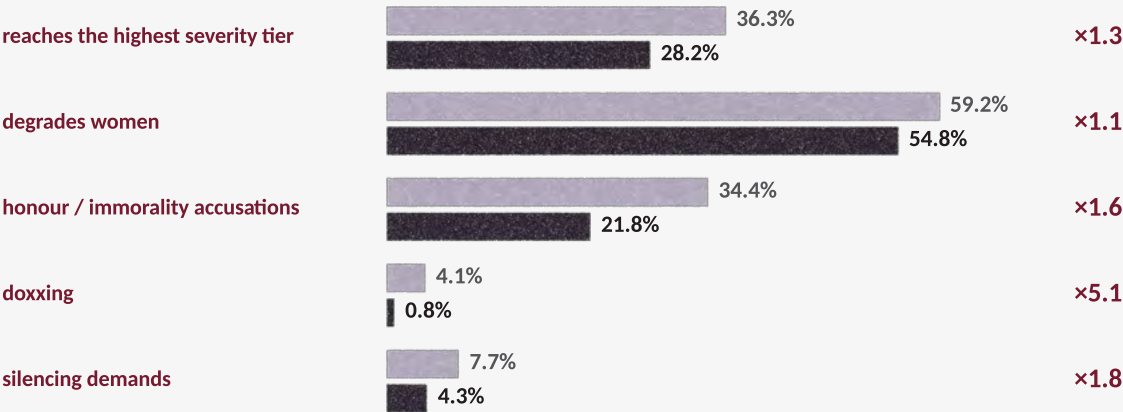
8.9%

of the corpus comes from Facebook,
but its abuse is more severe

Facebook
Twitter/X

share of each platform's abuse (%)

× Facebook vs Twitter/X



Over the four-month monitoring window, the corpus shows a late shift: from early April 2026, the share of posts reaching the highest severity tier falls to 16.4% while the share of posts degrading women rises to 65.6%. Reading alongside the severity paradox, this shift would, on a severity-first system, look like abuse is easing, when in fact gendered abuse is simply increasing its share of a metric that already under-scores it.

Regional implications

6,050 posts (10.3% of the corpus) accuse the target of collaboration or treason. 48.2% of these also degrade women in the same post, and the combination reaches the highest severity tier at 34.0%, above the 29.0% corpus baseline. In the Palestinian and wider Levantine context, an accusation of collaboration carries an implication of offline physical danger that no standard platform policy category currently captures. It is typically scored, if at all, as ordinary political insult, missing the compounded sexual and honour-based framing that accompanies it in practice.

A common objection to intervention is that Arabic-language abuse is too dialectal, coded, or evasive to moderate reliably. The evidence here points the other way: 95.1% of the content is lexically explicit, with no euphemism or coding, and the explicit rate is even higher among the most severe posts (93.1%). Only 2.5% of posts use any orthographic evasion technique. 10,771 severity-3 posts meet a stated-intent or incitement bar on their face, of which 1,906 are aimed at women. The persistence of this content is a resourcing and enforcement decision, not a technical limitation of Arabic-language moderation.

Accusations of collaboration or treason

10.3% of the corpus (6,050 posts) accuse the target of collaboration or treason

of these, also degrade women in the same post

48.2%

the combination reaches the highest severity tier

34.0%

29.0% corpus baseline

Stated intent or incitement, on its face

10,771

severity-3 posts meet a stated-intent or incitement bar

Arabic-language abuse is lexically explicit

95.1% of the content is lexically explicit, with no euphemism or coding

lexically explicit among the most severe posts

93.1%

of posts use any orthographic evasion technique

2.5%

1,906

of them are aimed at women





Conclusion

The headline figure in this corpus is not that women are attacked online, that is well established, but that most of the misogynistic language circulating in Levantine Arabic online spaces is not, in fact, aimed at women at all. It is aimed at men, through women, in a shared honour economy that platform policy currently has no category to describe. Addressing TFGBV in the Levant therefore cannot rest on identifying and removing attacks on women alone, it requires dismantling the gendered vocabulary itself, recalibrating what «severe» means, and measuring harm at the level of the person who receives it rather than only the account that sends it.

None of this requires new detection technology as over 95% of this content is explicit and unambiguous in Levantine Arabic. What it requires is dialect-accurate monitoring sustained over time, which is precisely the capacity 7amleh has built through this tool, and which this brief offers as a resource to platforms, regulators, and civil society partners working on digital safety across the region.

Recommendations

For platforms

Create a policy category for gendered abuse used as a vehicle against a third party, distinct from both harassment (requires a harassed target) and hate speech (requires a targeted protected group). Nearly half of misogynistic content in this corpus falls into this gap today.

Move from severity-only classification to a combined severity-and-volume-received model for individual targets, so that months-long, multi-account campaigns against one person are flagged even when no single account reaches a repeat-offender threshold.

Recalibrate severity scoring so that sexual degradation, reputational destruction, and honour-based threats are not systematically under-scored relative to threats of physical violence, including in automated moderation systems.

Apply Levant-specific classifiers for honour-system vocabulary and treason/collaboration-plus-gendered-insult compounds, which generic Arabic-language or Modern-Standard-Arabic classifiers are likely to miss or under-score.

For civil society

Require or encourage platform transparency reporting to disaggregate gendered abuse by whether the woman referenced is the actual addressee or an instrument used against a third party.

Fund and institutionalise dialect-specific monitoring capacity of the kind demonstrated here as national or pan-Arab aggregation is not an adequate substitute, and the regional consistency finding shows that Levant-specific tooling is transferable across Syria, Lebanon, Palestine, and the wider region.

Extend target coding in future monitoring rounds to include public role (journalist, politician, activist/WHRD, influencer, organisation) so that protection resources can be prioritised by risk profile.

For donors

Support continuation and extension of this monitoring pipeline beyond the current window, and its extension to additional platforms and Levantine sub-contexts, given the demonstrated capacity to sustain dialect classification and multi-level severity coding at scale.



Contact us:

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

Find us on social media: 7amleh