Investigative Report

Systematic Silencing

Online targeting of women human rights defenders in Syria and attempts to exclude them after the fall of the regime

SEEN Association for Civil Peace
Department of Documentation and Studies

CIR CENTRE for INFORMATION RESILIENCE

An investigative report by

SEEN Association for Civil Peace

in partnership with the Centre for Information Resilience (CIR)

Team

Dommar al Soleman
Head of the Department of Documentation and Studies at SEEN Association for Civil Peace

Youssef Wannous
Head of Open-Source Intelligence (OSINT) and Archiving

Reem Fadel
Head of Media

Ola Suleiman
Head of Advocacy

Content Warning

This report quotes and describes abusive, sectarian and sexual language, as well as threats of gender-based violence. The excerpts included are necessary to the investigation.

Contents

Executive Summary 10
Key Recommendations 13
Introduction to the Investigation – The Public Sphere as a Site of Abuse 14
1.1 Relationship to Earlier Studies and the Organisation's 2025 Report 15
1.2 Contribution and Limits of Originality 16
1.3 The Continuum of Violence – From the Digital "Mental Element" to Offline Crime 16
1.4 Methodological Framework – Examining the Mechanisms and Selecting the Sample 16
Forms of Abuse and Patterns of Targeting – Mapping the Data 19
2.1 The Targeting Map – Four Semantic Categories 19
2.2 Intersectionality and the "License to Abuse" 22
2.3 Online Engagement and the Limits of Inferring Amplification 24
How Attacks Develop and Their Technical Indicators – Mechanical Mobilisation and Directed Behaviour 26
3.1 Mechanical Mobilisation – The Social Machinery of Repression 26
3.2 Directed Behaviour – Steering and Amplifying Attacks 27
3.3 The "Haha Army" Tactic and Attempts to Manipulate Algorithms 30
Effects and Consequences of the Targeting 34
4.1 The Rise of "Digital Authoritarianism" and the Shift in the Source of Threats 34
4.2 Offline Consequences – Convergence of Violence and Exclusion 36
4.3 Effects on the Defenders' Digital Activity 37
Protection Gaps and Challenges 38
5.1 Applied Protection Matrix 38
5.2 Legislation as a Tool of Repression – Selective Enforcement and Gender Blindness 39
5.3 Platforms' Failure to Exercise Due Diligence in the Syrian Context 41
Policy and Practical Recommendations 42
6.1 To Syrian Lawmakers and Authorities (Ending Selective Enforcement) 42
6.2 To Major Digital Platforms (Restoring Due Diligence) 43
6.3 To Civil Society Organisations and Feminist Movements (Shifting to Proactive Support) 43
6.4 To Donors and the International Community (Funding the Response) 44
Conclusion 45
Investigative Methodology 46
1 Core Investigative Questions 46
2 Data Scoping and Incident Definition 47
3 Monitoring, Automated Classification and Lexicon Development (SeenBERT) 47
4 Human Verification (Human-in-the-Loop) 48
5 Criteria for Selecting and Investigating Cases 48

6 Actor Analysis and the Distinction Between Mass Engagement and Directed Behaviour 49
7 Levels of Evidence and Attribution (Attribution Confidence Levels) 49
8 Measuring the Effects on Online Activity 50
9 Documentation and Preservation of Digital Evidence 50
10 Qualitative Interviews 50
11 Applied Legal Analysis 51
12 Inclusion and Exclusion Criteria 51
13 Limitations 51
14 Verification and Triangulation 52
Statistical Classification of Likely Violations 53
Methodological Note 54
References and Sources 55

Before You Read

Notes on Scope and Method

Anonymisation and Protection

Three of the women human rights defenders who took part in the qualitative interviews were assigned codes (P1, P2 and P6), while three chose to be identified by name in open defiance of the attempt to silence them: Fadwa Mahmoud, Thuraya Hijazi and Afraa Hashem. Codes were also assigned to the other defenders and entities referred to in the report (H1, O1), and to abusive accounts and influencers (M1–M3, A1, A2, C1–C4, S1–S6, X1, X2, R01–R30). Names and photographs were obscured in screenshots, and the names of a number of individuals and entities cited in the defenders' testimony were removed to prevent the defenders from being identified again and avoid exposing them or their families to retaliatory threats. The original records are kept in a secure, encrypted digital environment.

Conceptual Framework for Technology-Facilitated Gender-Based Violence (TFGBV)

The report adopts UN agencies' definition of technology-facilitated gender-based violence.

> **"Any act of violence committed or aggravated, in whole or in part, through information technology and digital media, against a person on the basis of their gender."**

The report treats practices such as stalking, exploitation through deepfakes, defamation and hate speech as precursors to physical violence, challenging the idea that the resulting harm is merely "virtual". It takes an intersectional approach, recognising that the harm intensifies when a victim's gender intersects with their human rights or civic work, making this form of violence a complex instrument of repression and silencing.

Levels of Evidence and Attribution

The report applies a four-tier evidentiary scale, detailed in Appendix A, Section 7.

01	02	03	04
Level 1 · Behavioural description	**Level 2 · Indicators of coordination**	**Level 3 · Documented links**	**Level 4 · Conclusive attribution**
Document the existence of abusive content and its semantic and gendered patterns.	Document technical and behavioural indicators of “directed behaviour”.	Establish overlaps between accounts involved in a campaign and specific influential networks or platforms.	Make causal findings or attribute criminal conduct to a specific actor only when supported by corroborating material evidence that leaves no room for doubt.

Overview

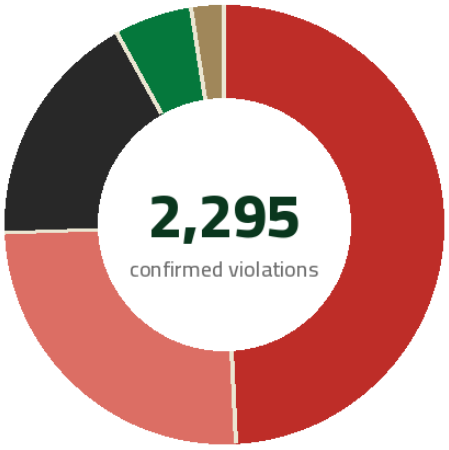
Key Figures

Findings from monitoring and verification on Facebook and Telegram from December 2024 to October 2026.

50,000+
Raw digital items in the archive (posts and comment threads)

7,793
Posts and comments screened and verified by human reviewers across 12 case files

2,295
Items in which human verification confirmed a gender-based violation



Breakdown of Confirmed Violations by Type of Targeting

Moral stigmatisation and political & sectarian framing	1,128 · 49.2%
Gendered contempt, belittling and denial of competence	577 · 25.1%
Sexual objectification and attacks on chastity	403 · 17.6%
Intimidation and explicit physical threats	133 · 5.8%
Image-based abuse and deepfakes	54 · 2.4%

70.5 %
Share classified as male among the 5,423 unique accounts that engaged with targeting campaigns

244
Accounts that recurred across multiple sites of incitement, including 59 in the sample for a single influencer

450 / 140
Average comments on a post discrediting women on the basis of gender, vs a general political post (+221%)

6
In-depth qualitative interviews with women defenders, activists and journalists (P1–P6)

29
Focal trigger posts included in a descriptive comparison with the baseline

3
Stages of the "License to Abuse": gendered punishment, withdrawal of community protection, denial of legitimacy

Executive Summary

This report examines patterns of directed online targeting[1] against **women** human rights defenders and civic activists in Syria in the period following 8 December 2024. It focuses on the functional interplay between gender-based violence and political and sectarian stigmatisation, and on how smear campaigns and threats affect public participation and human rights work. The analysis starts from the particularly precarious position of these defenders, whose two intersecting identities come into conflict with the prevailing patriarchal order. They are **women** in a society that ties honour to the body. They are also **actors in public life** who raise sensitive issues that put de facto authorities and local groups alike under scrutiny.

The report uses **intersectionality** as its organising analytical framework. Syrian women activists are targeted on the basis of their gender and their perceived sectarian or regional background, their clothing choices, and their political positions. The intersection of these factors gives rise to what the report calls the **"License to Abuse"**. A defender is first isolated on moral grounds and her chastity and conduct within the family are called into question. She is then branded with a political or identity-based label such as "regime remnant", "Nusayri", "SDF loyalist" or "agent". Once the stigma takes hold, her community's protection falls away. Sexualised attacks and defamation that would ordinarily be condemned become forms of abuse enabled and encouraged through collective complicity, under the banner of "defending the values of the revolution" or "protecting social values".

The investigation processed an archive of more than 50,000 raw digital items from Facebook and Telegram. Using **SeenBERT**, a natural language processing model, together with its dynamic gender-focused lexicon developed jointly by SEEN and the Centre for Information Resilience (CIR), the research team subjected the data to rigorous and comprehensive human-in-the-loop verification. The review removed false positives and filtered out political exchanges with no gendered component. The final

[1] Directed behaviour is a digital attack tactic that relies on vertical direction rather than horizontal coordination: an influential actor uses their reach to direct crowds of followers to intrude into the online space of a specific target in order to intimidate and silence her. Unlike "coordinated behaviour", which requires proof of prior organisation or communication among the attackers, directed behaviour relies on a public "starting signal", such as incitement or negative framing, that prompts the audience to carry out a collective attack (brigading). This produces a torrent of violations while leaving the principal instigator room to disclaim direct responsibility.

sample comprised **7,793 documented items**. Human reviewers confirmed that **2,295 posts and comments** met the report's definition of a gender-based violation. The investigation focused on 12 instances of targeting involving women human rights defenders, with the violations distributed across the main categories below.

The chart illustrates the distribution of violations across the five categories. Although a single comment or post may contain more than one category, each item was assigned to the category that best reflected its primary meaning and the main focus of the attack. This approach prevents double counting while preserving the item's principal context.

By systematically comparing SeenBERT outputs with metadata and the context of engagement, the investigation identified a recurrent pattern of behaviour that we term the **"License to Abuse"**. The findings show that the machinery of silencing does not engage with the substance of a defender's human rights work. It rests on a **composite structure of exclusion made up of three overlapping functional dimensions.** These may coincide in time or exchange roles depending on the context of a campaign, but they give the campaign its overall character: it opens with gender-based targeting and the withdrawal of community protection, paving the way for its culmination in the denial of the defender's legitimacy as a human rights and political actor.

01	02	03
Gendered punishment. Gender-based targeting, rather than debate over political positions, is the dominant signature of the wave of attacks. Monitoring recorded a flood of sexual insults, attacks on chastity, comments on appearance and the use of personal photographs to smear the defender. These attacks are intended to put her under immediate psychological pressure and undermine her personal and family reputation.	**Withdrawal of community protection.** As gender-based violence escalates and engagement peaks, mechanisms of social isolation come into play. In conservative communities, the intense focus on the defender's morals creates distance between her and those around her. Public expressions of solidarity decline, moderate voices are pushed to withdraw, and the defender is left without the support she would normally receive from her community.	**Denial of the defender's legitimacy.** Whether it coincides with the gender-based attack or represents its culmination, this dynamic is reinforced through hostile political and sectarian labels such as "regime remnant", "Nusayri", "SDF loyalist" or "agent". Political stigmatisation is not used here to open a debate. It is a decisive tool for stripping the defender of her standing as a civic actor and framing the attack on her as justified against someone cast as fair game on political and moral grounds.

Based on the monitoring findings and the links established between incidents through network analysis, the investigation concludes that this **overlapping functional interaction**, in which gender-based violence and the withdrawal of community protection prepare the ground for denying the defender's legitimacy, constitutes an effective dynamic of exclusion. Whether it emerges from an almost automatic patriarchal response within the community ("mechanical mobilisation") or is deliberately exploited by directed networks ("directed behaviour"), the process serves the same purpose: undermining women's credibility as human rights defenders in the public sphere. Rather than engaging with them on political issues, it attacks their private lives and targets them on the basis of gender, stripping them of social and legal protection in order to isolate and exclude them.

Six in-depth qualitative interviews (P1–P6) supplemented the computational monitoring. They confirmed that the digital space had become a parallel system of surveillance, compelling women to censor themselves. The effects extend offline through family pressure, the closure of organisations' premises and cross-border death threats. The report documents a climate of impunity sustained by a "double failure". Domestic legislation is inadequate, gender-blind and selectively enforced. Major digital

platforms, Meta and Telegram, fail to meet their human rights due diligence responsibilities in the Syrian conflict context.

The investigation addresses a central question. How is digital gender-based violence used to target women human rights defenders in Syria after the fall of the regime, and what patterns and mechanisms shape its spread and effects?

The following chapters address six subsidiary questions on the basis of the monitoring data.

1. What discursive and semantic features distinguish gender-based abuse, and how do they intersect with political stigmatisation and sectarian sorting to discredit women human rights defenders?
2. How do online attacks develop, and how do accounts and users take part?
3. What behavioural and technical indicators distinguish spontaneous mass engagement, or mechanical mobilisation, from directed behaviour and tactics for manipulating algorithms?
4. What measurable changes occur in the frequency and nature of defenders' online activity before, during and after a campaign?
5. How does the targeting affect defenders' personal and professional safety and their participation in public life?
6. What structural gaps in domestic legislation and digital platforms' due diligence policies prevent effective protection and redress for victims?

Key Recommendations

Major digital platforms. Improve understanding of Syrian dialects and contexts and provide safe, clear reporting channels. Before treating high engagement as evidence of content quality, investigate whether it reflects deliberate algorithmic amplification of hate speech.

Civil society organisations. Adopt proactive support strategies, including documenting and preserving digital evidence on behalf of those affected and with their consent, to spare survivors the psychological burden of rereading abusive material.

Donors and the international community. Support the development of secure, verified human rights databases and digital investigation tools capable of uncovering networks engaged in directed behaviour and producing conclusive evidence for attribution.

Syrian lawmakers and authorities. Introduce a standalone legal definition of technology-facilitated gender-based violence and criminalise such violence and non-consensual deepfakes. Suspend broadly worded provisions of the Cybercrime Law, move defamation and insult cases to civil jurisdiction, and require public prosecutors to act on their own initiative against incitement campaigns in accordance with the Rabat Plan of Action, without requiring a complaint from the person targeted.

01

Chapter 1

Introduction to the Investigation – The Public Sphere as a Site of Abuse

Syria's digital space has long reflected the political and social fault lines that divide life offline. The changes following the fall of the regime on 8 December 2024 ushered in a new period of openness in social and political life. Citizens' online participation grew markedly amid acute polarisation and an almost complete absence of transitional justice mechanisms and processes. In this security and legal vacuum, digital platforms, particularly Facebook and Telegram, the most widely used in Syria, became new arenas for settling political and sectarian scores. They also became instruments for policing the boundaries of the new public sphere and determining who could take part in it.

Amid these changes, women human rights defenders and civic activists faced a growing wave of technology-facilitated gender-based violence (TFGBV). The violence carried forward an established pattern of discrimination and gave it a new function as an instrument of repression and political intimidation, systematically excluding women who raised questions of accountability or challenged dominant narratives.

The current mechanisms of targeting grew out of a system of repression established under the previous regime. In January 2025, the Syrian Center for Media and Freedom of Expression (SCM) documented the systematic use of smear campaigns to make women activists outcasts in their own communities.[2] The evidence gathered for this investigation shows that the machinery of silencing has not stopped: its tools have passed to new local actors and networks with influence in those communities. As one defender (P6) explained, the shift in power moved the principal source of danger from the direct grip of the security apparatus to open social and digital spaces where abuse could go unchecked.

[2] Syrian Center for Media and Freedom of Expression (SCM), Smear Campaigns Against Women: Is the Threat Over? Targeting Women Activists as a Tool of Repression and Political Exclusion, by Tammam Saymoua and Lara Aizouqi, published as part of a joint project with Khatt 30, 27 January 2025.

> **“It happened to me several times. My profile was shared in different places, and so was my comment. I got a lot of insults, death threats and complaints. People even tagged the Ministry of Interior. Once someone wrote, ‘Where is she? Bring her in!’ And like I told you, there were threats to kill me if I went back to Syria.”**

1.1 Relationship to Earlier Studies and the Foundational 2025 Report

This investigation forms part of the work of the Department of Documentation and Studies at SEEN Association for Civil Peace and the Centre for Information Resilience (CIR) to monitor hate speech in Syria’s digital space and analyse its patterns, meanings and effects. The report Hate Speech in the Syrian Digital Space (2025)[3] found that hate speech takes the form of direct verbal abuse as well as stigmatisation, exclusion and incitement associated with social, political and sectarian identities and affiliations. In some contexts, this hostile language can intersect with social tensions and violence. The earlier report also noted that women journalists and human rights defenders faced sexual threats, smear campaigns and pressure aimed at undermining their online presence. Its main focus was on hate speech, its patterns and its spread in Syria’s digital space. It gave less attention to how women human rights defenders and activists are targeted and how gender-based violence interacts with political, social and sectarian stigmatisation.

This investigation therefore asks a more specific question. How is technology-facilitated gender-based violence used to target women human rights defenders and activists in Syria, what forms does it take, through what mechanisms does it operate, and how does it affect their public participation, sense of safety and ability to continue their work? The investigation moves from monitoring hate speech in general to examining cases and the mechanisms and contexts of targeting. It draws on digital content analysis, human verification, case studies and qualitative interviews as set out in its methodology. The investigation does not assume that the targeting of women or online gender-based violence has gone undocumented in the organisation’s previous work. Building on the earlier report, it seeks to deepen the understanding of the gender dimension, mechanisms and effects of targeting, and contributes further analysis to the

[3] SEEN Association for Civil Peace in partnership with the Centre for Information Resilience (CIR), Hate Speech in the Syrian Digital Space (2025): Patterns, Meanings and Effects in a Conflict Environment, Department of Documentation and Studies, March 2026.

organisation's body of research. Earlier findings are restated only where needed to frame the investigation or compare its results.

1.2 Contribution and Limits of Originality

Online violence against Syrian women has been documented before. This investigation combines content monitoring, human verification and case studies before examining the chronology of targeting, indicators of coordination and changes in the defenders' online activity. It uses the "License to Abuse" as a testable explanatory hypothesis.

The strength of this contribution depends on publishing clear data on the size of the monitoring sample, criteria for selecting incidents and the limits of SeenBERT, along with documented examples showing how the analysis moves from observation to findings on coordination or attribution. The report defines its contribution by what its evidence and methods establish, without claiming primacy or exclusivity.

1.3 The Continuum of Violence – From the Digital "Mental Element" to Physical Crime

Digital violence is not confined to online abuse. It can extend offline and turn into threats or physical attacks. The cases documented in this investigation indicate that incitement on digital platforms can help create an environment that encourages violence against women human rights defenders, without necessarily implying a direct causal link between online incitement and violence on the ground.

Words written on screens often help build the psychological and social infrastructure that can later legitimise an attack.

The effects of digital campaigns go beyond psychological harm. They reach into families and communities, where relatives may be pressured to silence the defenders on the pretext of averting the "scandal" generated by online stigmatisation. In several documented cases, these campaigns extended to death threats, pursuit by security services or administrative action affecting the premises of feminist organisations.

The intersection of inciting speech and physical action suggests that digital platforms can at times become spaces for mobilisation that contribute to political exclusion and the imposition of "social punishment". This underscores the urgency of addressing such speech through protection and accountability mechanisms.

1.4 Methodological Framework – Examining the Mechanisms and Selecting the Sample

The methodology combines sociological analysis and evidence-based investigation using two main approaches.

- **Data sampling and scoping.** The research team combined proactive, network-led monitoring with retrospective, target-led monitoring. It identified a set of public accounts, pages and groups active in Syrian public affairs on Facebook and Telegram. Based on monitoring of their past activity, the team had strong grounds to suspect that these accounts, pages and groups played a role in initiating or disseminating hate speech and incitement. Researchers archived the selected accounts' entire feeds while closely monitoring the public pages of several targeted defenders and activists to trace how the attacks spread. This produced a comprehensive raw archive of more than 50,000 digital items, including posts and comment threads, covering cases from December 2024 to October 2026.
- **Algorithmic classification and lexicon development (SeenBERT & Dynamic Lexicon).** The team screened the initial database using SeenBERT, a specialised language model, to identify hate speech and content used for targeting. Its purpose-built gender-focused lexicon was updated cumulatively and regularly as new terms were identified. In each round of extraction and screening, investigators recorded new terms, slurs, metaphors and coded expressions carrying gendered abuse, including metaphors about honour, adaptations of colloquial dialect and coded derogatory labels. They incorporated them into the lexicon and used them to retrain the model to detect expressions used to evade platform algorithms.
- **Human verification and incident aggregation (Human-in-the-Loop & Incident Aggregation).** Researchers reviewed and checked the automated results in full to remove purely political exchanges and false positives. They identified and examined 7,793 documented and verified posts and comments across 12 case files. Items were grouped into distinct incidents of gender-based targeting on the basis of proximity in time, a shared narrative and a common starting point. The final count included 2,295 items containing gender-based violations under the report's definition. A further 5,429 items comprising general political exchanges with no gendered component formed the control group. Reviewers classified 69 items as inconclusive.

> "To assess the reliability of the automated classification, two independent reviewers examined the initial screening results and compared their assessments for

consistency. Compared with this reference classification, the model proved highly effective at detecting explicit indicators of stigmatisation and direct insults. All ambiguous cases involving sarcasm, local dialect or double meanings were referred for full human review to ensure that no classification rested on automation alone."

- **Investigation and identification of actors.** The investigation draws a rigorous distinction between "mechanical mobilisation", the spontaneous and predictable mass engagement conditioned by the prevailing patriarchal order, and "directed behaviour", which relies on prior direction, the synchronised circulation of content and hashtags, and attempts to manipulate algorithms to amplify abuse.

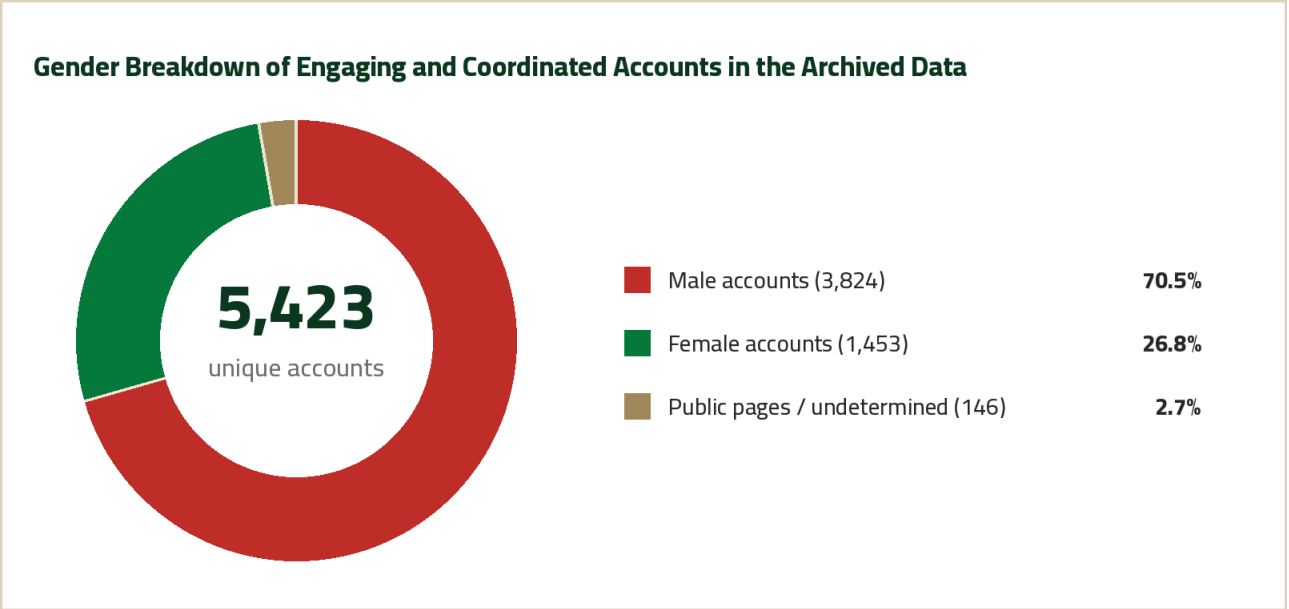


Figure 1. Male accounts predominate among accounts engaging with gender-based targeting campaigns

The chart shows the demographic breakdown of 5,423 unique accounts identified in the archived Facebook and Telegram data. Male accounts predominated, at 70.5% (3,824 accounts), and were the main drivers of the campaigns of incitement and abuse. Female accounts followed at 26.8% (1,453 accounts). Public pages and undetermined accounts made up the remaining 2.7% (146 accounts).

Researchers inferred the gender associated with each account from the account name, declared gender, profile photograph and the masculine or feminine grammatical forms used in posts. Ambiguous accounts and public pages were classified as "undetermined". The classification may differ from the actual identities of those who operate the accounts.

The team linked these data to case studies and in-depth qualitative interviews with Syrian women human rights defenders. To ensure its findings could withstand scrutiny, the investigation drew on a four-level evidentiary framework, from behavioural

description and indicators of coordination to documented links and conclusive attribution.

The following chapters address a central question. How does the machinery of online silencing operate in Syria, and who is responsible for stopping it?

02

Chapter 2

Forms of Abuse and Patterns of Targeting – Mapping the Data

Personal testimonies alone cannot explain the dynamics of digital exclusion affecting Syrian women human rights defenders. Moving from documenting individual harm to establishing a systematic pattern of digital repression requires close analysis of the wider body of data. This chapter combines outputs from the SeenBERT language-monitoring model with the defenders' qualitative testimony. It maps the main forms of attack and examines how political polarisation intersects with gender-based violence to produce what the report calls the "License to Abuse".

2.1 The Targeting Map – Four Semantic Categories

Participants' accounts indicate a shift in the nature of attacks after the fall of the regime. Direct political debate gave way to attacks that focused on their bodies, honour and families. Fadwa Mahmoud (P3), a veteran activist with decades of public work behind her, said she had never faced so many insults about her "honour" or her "appearance" throughout her career.

The patterns found in the archived data reflect this shift. They allow the violations to be grouped into four main semantic categories.

- **Moral and family-based stigmatisation.** In Syria, this form of abuse is distinctly family-based. Attackers do not stop at branding activists as "immoral" or attacking their choice of clothing; they go on to target the men in their families. Expressions such as "There are no men in the house" are a clear attempt to turn the family into an instrument of repression against the defender. The attackers routinely focus on private life. The final count for this category reached 1,128 violations (49.2% of the total), concentrated on attacks on honour and accusations of treachery based on sectarian or political affiliation. Attackers also set up fake accounts impersonating defenders to spread false information, subject them to sexual blackmail and smear them daily.

text	threadingDepth	gender_targeting_categc
يعني مين هنن حتى هلاء يجوا ويحكوا وقت كان الشعب عم ينذبح ويتنكل به وين كانت هي [redacted] وين كانوا يروحوا طموا راسهم بالطين	0	الترهيب والتهديد الجسدي
[redacted] يا قحبة يا شرموطة يا مرابعة يا بنت المرابعي :دعسناكم وطلعناكم بملابس النسوان مثل الكلاب وعواكم وصل قبرص وماضل حدا ما ساعدكم يا ولاد القحبة ورغم هيك انقلعتو وانذليتو ودعسناكم بالصرامي يا عاهرة ولسا رح تندعسو يا مرابعة 👌	2	الترهيب والتهديد الجسدي
[redacted] معلش نحنا منحب لارها..بي بكفي د.عس لا.قليا.ت وحكم لا.قليا.ت ولعلما.نية تبعك يا احية.١٤ سنة نحنا عم نتق.صف وانتو متل لكلا.ب بدكن حكم متل مابدكن. وبدكن س.نة يعملو د.ين متل مابدكن ك.س اختكن واخت شر.فكن الله يديم صر.مايتنا عليكن احبة نسو.ية متحر.رة	2	الترهيب والتهديد الجسدي
[redacted] يا عجوز يا شمطاءيا شمبوقةيقول الشاعر ابو عمشة : العلوي الجبد هو العلوي الذي لايتنفس🤣🤣🤣🤣🤣😅🤣	1	الترهيب والتهديد الجسدي
[redacted] بدي انيجها واشق ثمها هل عاهرة	0	الترهيب والتهديد الجسدي

Figure 2. A sample from the archive of SEEN Association for Civil Peace and direct messages received by some activists

English translation of the text in the screenshots:

"So who are these people coming out to speak now? When people were being slaughtered and brutalised, where was she, [redacted]? [Redacted] Where were they? They buried their heads in the mud." *(thread depth 0; category: Intimidation and physical threats)*

"[Redacted] You whore, you slut, you muraabia [derogatory term], you daughter of a muraabi. We trampled you and drove you out in women's clothes like dogs; your howling reached Cyprus, and no one came to your aid, you sons of a whore. Even so, you were driven out and humiliated. We trampled you under our shoes,

you whore, and we told you that you would be trampled, you muraabia." (thread depth 2; category: Intimidation and physical threats)

"[Redacted] Fine, we love the terrorist. Trampling minorities, minority rule, and that secularism of yours are enough, you whore. For 14 years we were being shelled while you barked like dogs, wanting a regime that suits you. Fuck your sister and your honour. May God preserve our boots over you, you liberated feminist whore." *(thread depth 2; category: Intimidation and physical threats)*

"[Redacted] You old woman, you grey-haired hag, you *shambouqa*. As the poet Abu Amsha says, 'The good Alawite is the one who isn't breathing.'" *(thread depth 1; category: Intimidation and physical threats)*

"[Redacted] I want to get her and rip her mouth open, that whore." *(thread depth 0; category: Intimidation and physical threats)*

"Anyway, let's see how long you stay in Erbil. After those comments, people are bound to file complaints against you. Let Erbil bring your family over so you can live near them, if you can."

"Glory be to whoever humiliated you lot." "I shit and piss on you." "On your religion." "You filthy, cursed [person]." "Impure filth." "I swear to God, every day I'm more certain that killing Alawites is an act of worship. You can't possibly be human, you pimps." "You adulteress." "You whose honour has been violated." "And your honour and religion …"

14 May, 03:35: "You slut, there's no point blaming you." "Damn the womb that carried you."

8 June, 20:28: "The least I can say about you is that you're a first-rate slut." "Just wait till we get hold of you and piss on you, you whore." "I swear to God, you look disgusting. Fuck your looks."

12 June, 02:13: "Shit on you, [redacted]. [Redacted] … You're just a girl who's come to pontificate about affairs of state. You're disgusting. We'll crush people like you and everything you do in this country."

- **Gendered contempt, belittling and denial of competence (Gendered Contempt & Devaluation).** The investigation recorded 577 likely violations in this category (25.1%). These attacks sought to undermine defenders' intellectual and political competence through belittling labels such as "shambouqa" and "female truck driver", insults about age or appearance such as "Go cover yourself up, old hag" and "unveiled woman", and attempts to restrict women to a position of dependence on men.
- **Sexual objectification and attacks on chastity.** This reduces the defender to her body and subjects her to crude sexual remarks. The investigation recorded 403 likely violations in this category (17.6% of all recorded violations), including direct attacks on chastity, sexually defamatory insults and vulgar sexual language. P2 described how attackers disregard the substance of her work and go after "the way I look, my body, who I am … They attack a woman because she's a woman, without regard to her work."
- **Intimidation, physical threats and image-based abuse (Physical Threats & Image-Based Abuse).** The investigation recorded 187 likely violations in this category (8.1% of the total), divided between two functionally related patterns:
  - **Intimidation and physical threats (133 cases, 5.8%):** explicit threats to kill and cause physical harm ("We'll bring you in", "rip her open"), tracking down where defenders live, and tagging security agencies to call for the activists' arrest.

- **Image-based abuse and deepfakes (54 cases, 2.4%):** any use of the target's personal photographs without her consent in defamatory contexts, humiliating manipulation or cropping of images, and the use of artificial intelligence to generate misleading, fabricated images or videos intended to destroy her reputation.

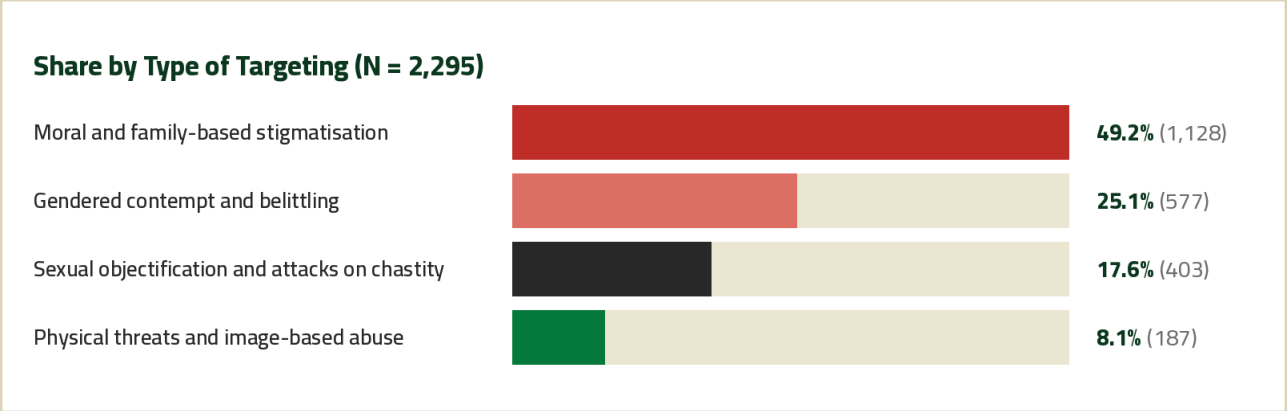


Figure 3. Distribution of gender-based targeting in the archived data

2.2 Intersectionality and the "License to Abuse"

Pattern analysis reveals a central mechanism that helps explain the intensity of gender-based violence in Syria's complex political context. Women who do not belong to the groups that hold power, or who work on sensitive transitional justice issues such as missing persons and abducted women, are more readily treated as fair game.

The monitoring shows how the machinery of silencing operates in practice: rather than engaging with an activist's political positions, it sets in motion a dynamic of exclusion. Gender-based targeting often delivers the "opening blow", intended to cause rapid psychological shock and isolate the defender socially. As the campaign of moral defamation escalates and cuts the target off from her supporters in conservative communities, political and sectarian stigmatisation intersects with the gender-based attack, whether it occurs alongside it or represents its culmination. It is not an accusation to be proven but a decisive tool for stripping the defender of her standing as a human rights defender and framing the attack on her as a "justified act". P1's testimony captures this final framing precisely: after the gender-based attack against her escalated, a sectarian background was invented for her to complete her exclusion from civic life.

> "Why did he call me a Nusayri? To get people to turn on me … He's making me fair game for anyone who wants to come after me."

The “License to Abuse” is most pronounced when gender explicitly intersects with sectarian polarisation. In-depth interviews and monitoring records of direct messages and threats show that women journalists and activists from particular sectarian backgrounds living inside Syria, including in Latakia, Hama and Damascus, can be targeted without engaging in prominent political activity. They receive persistent messages calling them “Assad’s orphans”, “female captives” or “Nusayris”, accompanied by explicit threats. This intersection creates a state of terror online. A woman’s presumed sectarian identity accelerates the violence, and impunity makes it easier to target her.

Figure 4. Direct messages targeting defenders on the basis of a presumed sectarian identity, reinforced through the attacks but not necessarily reflecting their actual background

English translation of the text in the screenshots:

22 January, 11:28 p.m. “[Redacted] … may you be orphaned, you girl from al-Ataa. [Redacted] … sister of your shitty sect, daughter of Abu Taqa.”

“Iranian Alawite. What else could we expect from you? The Umayyads trampled your honour because …”

9 June, 00:00. “You told me you were Sunni, you loose woman, you bastard. Did you think we would let you keep lying like this without keeping tabs on you, you worthless piece of shit? Damn you, and damn those filthy Alawite criminals. How many Alawites, you whore, you adulteress, you daughter of a dog, have come out and apologised in public, from the start of the revolution until today, for what the regime did and what the rabid dogs of your whore of a sect did? Damn your religion, you bastard, you criminal, you pig.”

5 January, 12:20 p.m. “[Redacted] you, your demands and your whole sect.”

29 March, 11:08. “Alawite woman, I want to ask you something.” “Why have you turned off comments, Alawite woman? Is that how it is? I didn’t expect an educated Alawite woman like you to do that. God forgive you. God knows Alawites don’t do that. I really wanted to support your position, Alawite woman.”

2.3 Online Engagement and the Limits of Inferring Amplification

Measuring engagement with abusive posts can help describe how users respond to the content and how it spreads. Engagement alone cannot establish that the platform’s algorithms increased the content’s reach or favoured it over other material. Engagement rates are affected by factors such as the size of the account’s audience, the timing and format of the post, the number of shares and reposts, and the nature of the event discussed.

Where reliable data on reach, impressions or platform recommendations are unavailable, the report uses “relative engagement with abusive content” in place of “relative engagement indicator”.

2.3.1 Results of the Relative Engagement Analysis

To gauge how much engagement gender-based hate speech generated, the report conducted a descriptive comparison using a purposive sample of 29 focal posts on Facebook and Telegram. These were selected as the trigger posts that set off targeting campaigns through the influential accounts under review. Given the unusually high engagement typical of trigger posts, the research team established a baseline of 29 general political posts with no gender-based targeting. To make the comparison, it matched each focal post to a general post from the same account or page and the same period. Quantitative engagement data were gathered using data-scraping tools.

Descriptive results. Posts that attacked women on the basis of gender or questioned their chastity drew 3.2 times the rate of engagement and reposting recorded for the political baseline posts. Together, the 29 focal posts drew about 25,000 reactions and 9,000 comments. The highest concentrations of engagement occurred on posts targeting defenders by name and using their personal photographs. Focal posts averaged 450 comments, compared with a baseline average of no more than 140 comments per general political post without a gendered component, an increase of 221%.

Behavioural significance. The observed difference indicates that sensational content and incitement framed around “honour” are highly likely to polarise audiences and generate engagement. The descriptive finding leaves open whether a pre-existing algorithmic bias promoted such content, as platforms do not provide open access to

application programming interface (API) data for independent scrutiny of recommendation algorithms.

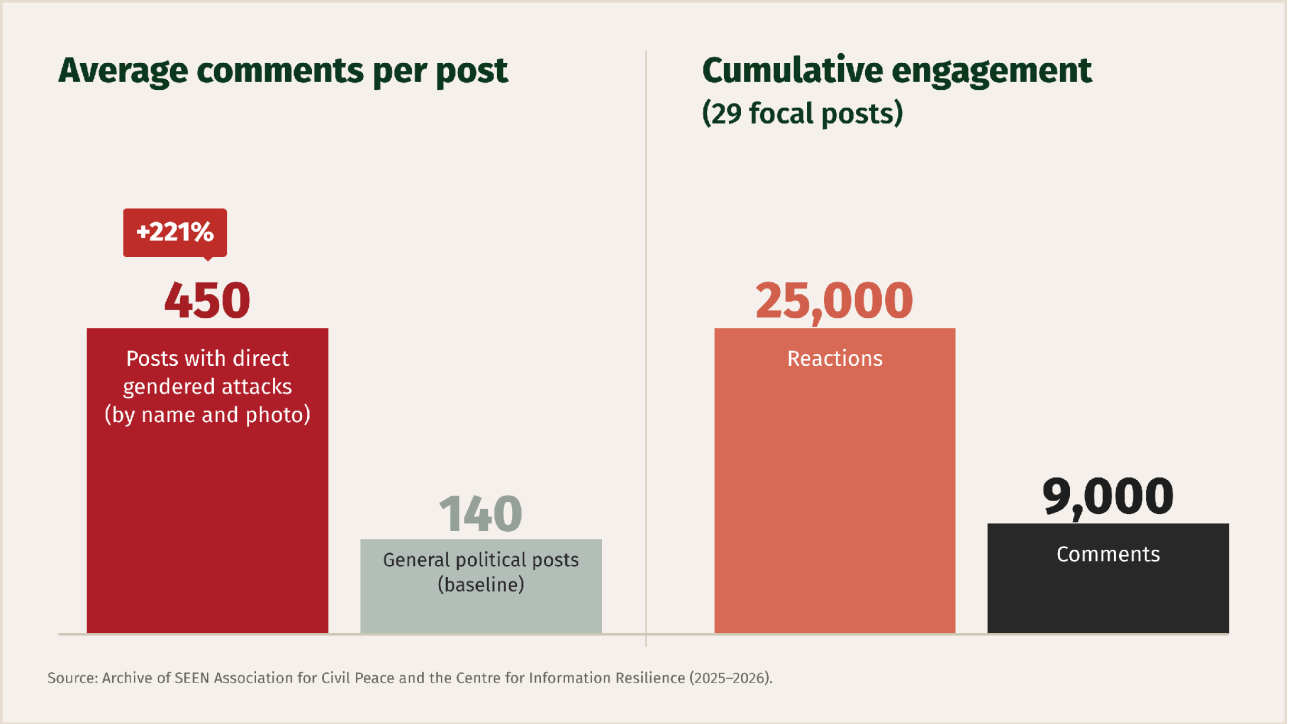


Figure 5. Descriptive comparison of engagement: focal posts versus the baseline

The chart shows two levels of measurement. On the left, it compares average comments per post for posts containing gender-based attacks (450 comments) with political baseline posts (140 comments), an increase of 221%. On the right, it shows the total cumulative engagement across the 29 focal posts combined: 25,000 reactions and 9,000 comments.

Source. Archive of SEEN Association for Civil Peace and the Centre for Information Resilience (2025–2026).

03

Chapter 3

How Attacks Develop and Their Technical Indicators – Mechanical Mobilisation and Directed Behaviour

The investigation of specific incidents shows that digital violence in Syria operates through two parallel mechanisms that can take over from one another. The first is a “social mechanism” that responds mechanically and predictably to patriarchal cues. The second is a “technical mechanism” that uses directed networks to amplify attacks and evade platform policies. Unorganised mass engagement and directed behaviour converge in their immediate effect. They exclude women human rights defenders from public life and undermine their personal and professional safety.

Of the 29 focal incidents examined through network and time-series analysis, 20 were classified as “mechanical mobilisation”. Seven more complex incidents met the second evidentiary level (directed behaviour), and two remained inconclusive because their metadata were incomplete.

3.1 Mechanical Mobilisation – The Social Machinery of Repression

The cases examined suggest that much of the violence directed at women human rights defenders is driven by what the report calls “mechanical mobilisation” and can happen without an “operations room” or direct instructions from a central authority.

Mechanical mobilisation refers to a spontaneous dynamic in which ordinary users join a collective attack without prior coordination, responding to patriarchal values and conservative social structures. Once the initial stigma is introduced, online crowds join in a conditioned response aimed at policing and punishing any woman whose public presence challenges dominant narratives or raises sensitive questions of accountability. Digital violence becomes contagious and self-reinforcing. Commenters

identify with the hate speech under the illusion that they are fulfilling a “duty to protect the public sphere”. The effects can be as destructive as those of an organised attack.

The case of Fadwa Mahmoud (P3), an activist working on justice and enforced disappearances, illustrates both mechanical mobilisation and the stages of the “License to Abuse”. After she took a public, critical position on accountability processes, attackers sidestepped her human rights arguments. The first wave targeted her on the basis of gender and focused on her appearance and age, using belittling language and stigmatising labels. The emphasis on her morals and appearance quickly began to isolate her. Her supporters expressed less solidarity in public to avoid social embarrassment in a conservative community. One interviewee described this as a situation in which the target becomes fair game for everyone, as users join in simply to imitate others on the platform.

To complete the attack and justify participation in collective bullying, the third dimension came into play: stripping her of her standing as a human rights defender and accusing her of political betrayal by branding her a “regime remnant”. Fadwa Mahmoud’s case shows how political polarisation can break down solidarity among those working on human rights and prompt people to echo hate speech spontaneously, even in the absence of networks of fake accounts. P2 compared the bullying to a “national wedding celebration”. People joined the mobbing simply because “everyone’s bullying so-and-so, so I have to join in”.

3.2 Directed Behaviour – Steering and Amplifying Attacks

Spontaneous mass mobilisation alone cannot explain every pattern of digital violence. The archived data reveal converging technical and behavioural indicators that meet the second evidentiary level for coordination. These indicators do not, on their own, establish direct attribution to a security agency or political actor.

Account tracing confirms a pattern of organised brigading, or directed attack. By matching accounts across posts by influencers M1, M2 and M3 and posts by media outlets that hosted the defenders, the investigation found 244 recurring accounts operating in parallel across several focal points of incitement. They included 59 accounts that moved systematically across multiple posts in the sample for a single influencer. Their activity went beyond comments on individual posts. They acted as amplifier accounts for the campaign, moving as soon as new material appeared to pursue the defender and flood her space with the same templates for moral stigmatisation, sexual objectification and contrived hashtags, such as “Sabaya al-

Ataa”[4], as well as accusations of betrayal. They followed her onto media outlets’ pages and commented on clips of her interviews, engaging in directed brigading that went beyond spontaneous activity.

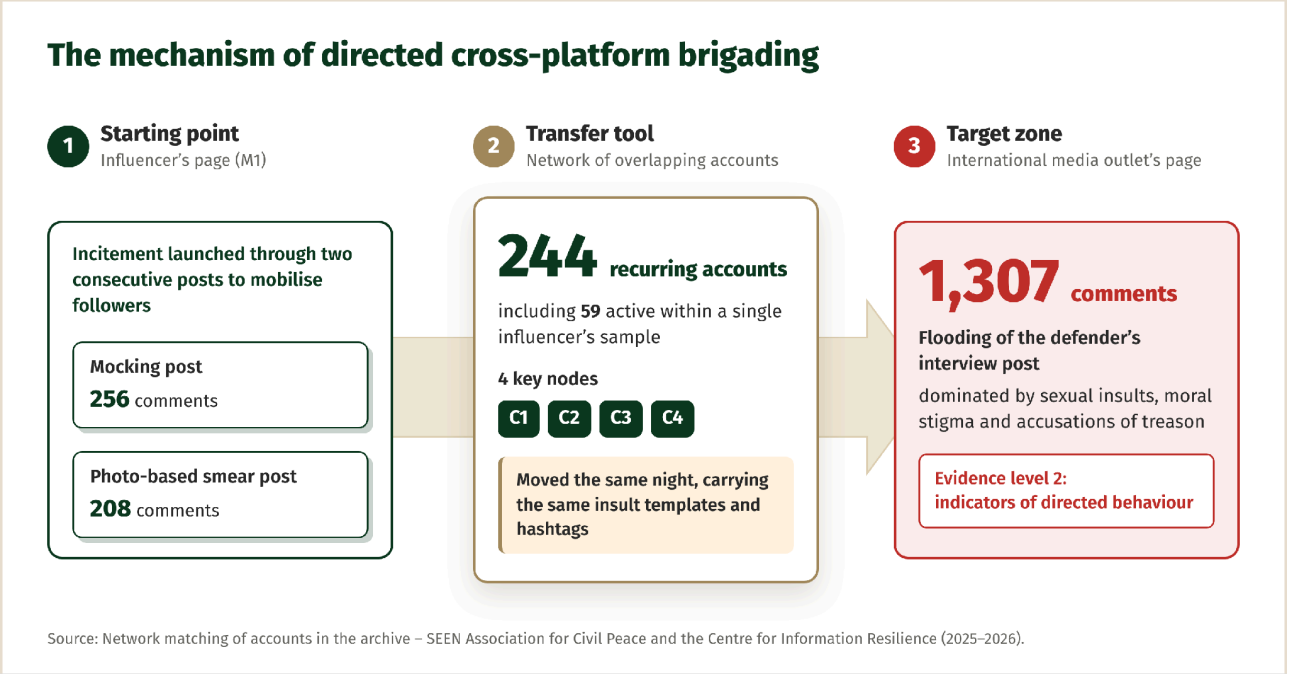


Figure 6. The path of cross-platform brigading

The diagram shows how behaviour was coordinated between the central inciting post on the influencer’s page and a post by an international media outlet. The investigation documented dozens of matching accounts moving in sync that night to pursue the defender, flooding the space where she appeared in the media with the same accusations of betrayal and gender-based slurs in an effort to undermine her credibility as a human rights defender.

The network pattern is consistent with the targeting of one activist after she posted critical replies to an inciting post. The investigation traced a systematic campaign led by accounts with overlapping technical indicators, which moved in sync from the influencer’s page to the defender’s personal page. The timeline shows two accounts, A1 and A2, moving to her personal page after 1.34 hours and 12.19 hours, respectively. They flooded six separate posts with accusations of being an agent and attacks on her honour.

[4] The term “Sabaya al-Ataa” (meaning “young women who give”) originally referred to women’s support groups affiliated with the National Defence Forces (NDF) and other formations loyal to the former regime. These women were mobilised for propaganda purposes and to provide logistical and media support to government forces; the name thus remained associated with the regime’s apparatus of power. Following the fall of the regime, the term was repurposed as a smear hashtag. It exploits the close phonetic resemblance between ṣabāyā (صبايا, “young women”) and sabāyā (سبايا, “female captives” or “women taken captive in war”) to portray the targeted activists as having served the former regime in a sexual capacity. The hashtag thus combines political and sexual stigmatisation in a single label, embodying the mechanism of “granting permission to attack” described in this report.

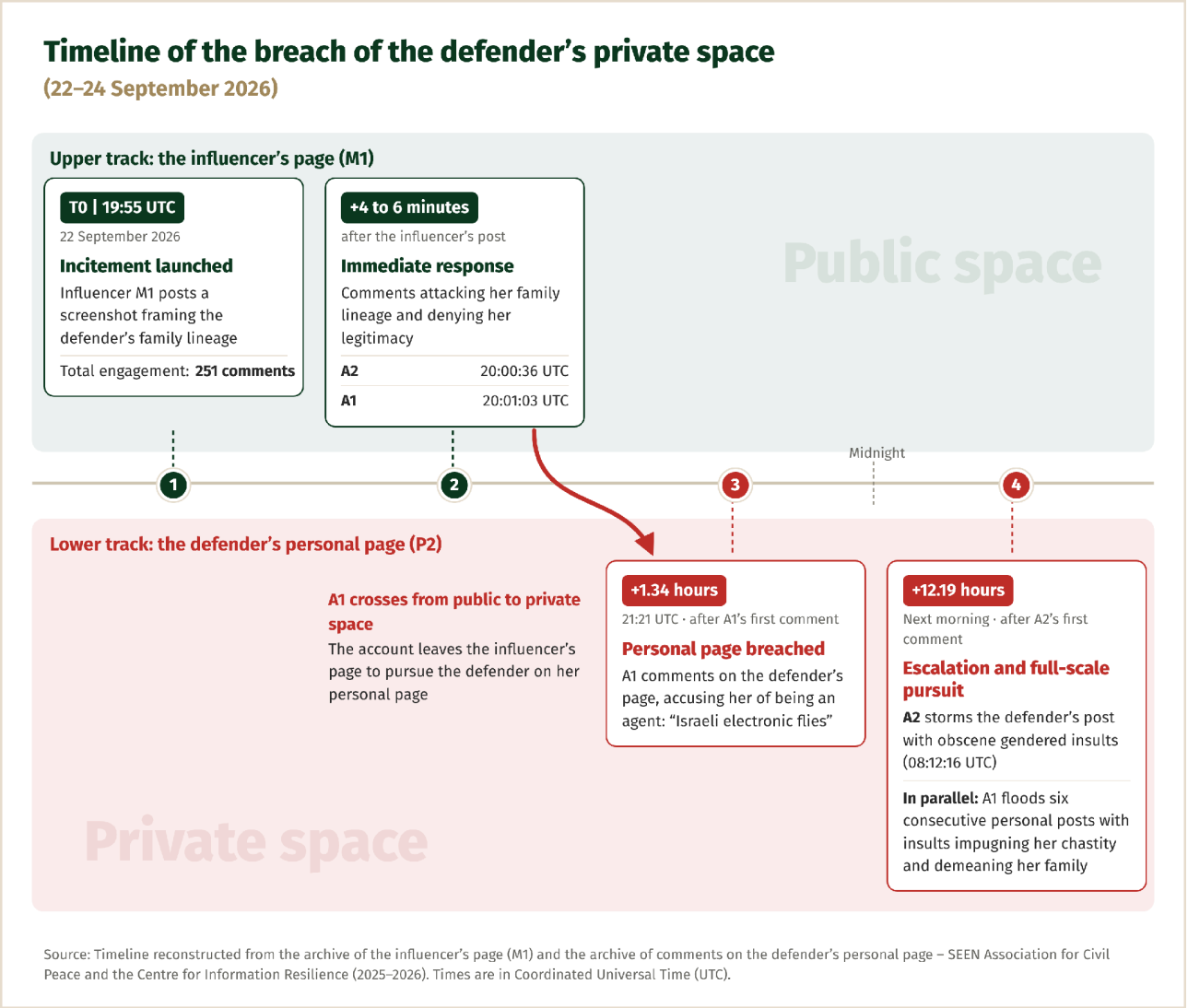


Figure 7. Sequential and temporal tracking of account movements and intrusion into the defender's personal online space

The diagram traces the organised movement of inciting accounts from M1's post at 19:55 UTC. The time-series analysis documents an initial flow of comments just four minutes later, followed by A1's directed move after 1.34 hours to intrude into the defender's personal online space and attack her with insults impugning her chastity and accusing her of being an agent. The directed behaviour peaked at 12.19 hours, when A2 intervened to pursue the activist simultaneously across six separate posts on her personal page, flooding them with accusations of betrayal and the "Sabaya al-Ataa" hashtag in an instance of directed brigading.

3.2.1 Criteria for Documenting Directed Behaviour

- **Account classification.** Distinguish carefully between public pages, accounts belonging to influencers and content creators, and accounts that available indicators suggest are fabricated or operated under assumed identities (sockpuppets).
- **Spikes and timing.** Measure sudden influxes of comments containing gender-based abuse within narrow time windows as evidence of scheduled or directed posting.

- **Network mapping.** Identify the seed accounts that initiate the incitement, the amplifiers that share and boost it, and the accounts that flood the targeted spaces.
- **Shared linguistic patterns.** Track identical wording and verbatim copying of phrases and ready-made hashtags used to frame the person targeted.
- **Ethical and legal safeguards.** Document influential public figures and entities accurately while anonymising individuals behind ordinary accounts to protect their privacy.

3.2.2 The Case of P1

The case of P1, a woman human rights defender from north-eastern Syria, is a practical example of "compound directed abuse". After showing solidarity with abducted women and taking part in civic activities, she faced random abuse as well as attacks using sophisticated, systematic technical methods. The investigation documented the following indicators of coordination in her case.

- **Deepfakes.** A fabricated video depicting the defender in an intimate situation with a stranger was published to destroy her reputation quickly in a conservative community.
- **False identities and amplification.** Media pages run under false identities, including one run by a man using a woman's name, had large audiences in the defender's city. They published her name and photograph at the same time, a tactic intended to single her out for their audiences to attack.
- **Targeted hashtags.** Accounts systematically flooded her spaces with ready-made hashtags that assigned her invented identities. They claimed she was a "Nusayri" to mobilise sectarian hostility in her community and repeatedly called her an "SDF loyalist" and a "protector of minorities".

3.3 The "Haha Army" Tactic and Attempts to Manipulate Algorithms

Directed engagement from fake accounts is one of the more recent and consequential forms of directed behaviour used to target Syrian women human rights defenders. The archived material shows a recurring pattern of serious human rights posts, on detainees, abducted women, transitional justice and the documentation of violations, being flooded with hundreds or thousands of organised mocking reactions using Facebook's "Haha" button.

A review of the accounts' profile information and geographic patterns reveals a marked disconnect. The visual analysis documented concentrations of accounts listing locations far removed from Syria, including Lusaka in Zambia, Antananarivo in Madagascar and Jhapa District in Nepal, as well as locations in Indonesia, Thailand, Myanmar and Sri Lanka. These geographic discrepancies strongly suggest that the synchronised influx of reactions lacked an organic basis and was likely manufactured.

Campaign infrastructure and operations. The indicators are consistent with internationally documented infrastructure for generating inauthentic engagement.

- **Click farms and automated voting.** These operations use thousands of phones with inexpensive SIM cards, or use residential proxies that simulate residential internet addresses in different countries to evade Meta's detection systems.
- **Social media marketing panels (SMM panels).** These non-public marketplaces act as intermediaries, allowing instigators to buy packages of specific reactions, such as 1,000 "Haha" reactions for US$2 to US$5, and precisely schedule their delivery to a target post.[5] Although the documentation team was unable to obtain conclusive financial evidence identifying entities that funded this activity, the clear behavioural similarities between these attacks and the operating methods of such platforms lend strong support to the hypothesis that paid services were used deliberately.
- **Hijacked and abandoned accounts (zombie accounts).** Originally created by real users in the Global South, these accounts are taken over through phishing or abandoned for years before being sold in illicit data bundles and forcibly reactivated by automated networks without their owners' consent.

The tactic is intended to provoke the defender psychologically and to manipulate how algorithms distribute her posts and how local audiences understand them. It does so in two complementary ways.

- **Algorithmic redirection.** Facebook's algorithms interpret the flood of "Haha" reactions as a signal of exceptionally high engagement with controversial content. Rather than curbing the attack, the recommendation system amplifies the post and redirects it away from its intended organic audience, such as families of missing persons or people working on human rights, and pushes it towards unintended audiences. This separates the defender from her supporters and exposes her posts to wider audiences who lack context and are more susceptible to polarisation, increasing the volume of spontaneous hostile reactions.

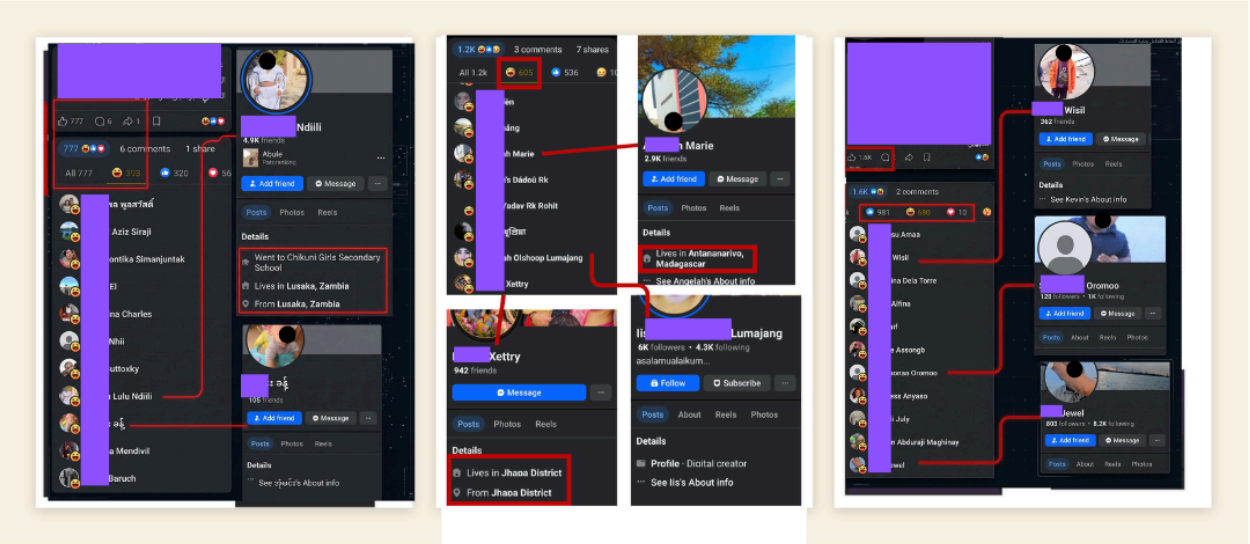
[5] Reports by the Atlantic Council's Digital Forensic Research Lab (DFRLab) and the Institute for Strategic Dialogue (ISD) on the economics of social media marketing (SMM) panels and click farms.

- **Shaping perceptions through social proof.** This pattern undercuts the moral and human significance of material documenting human rights violations. When users see a post about serious abuses surrounded by hundreds of inauthentic laughing reactions, social proof can lead them unconsciously to assume that mocking the victim and dismissing her account is the natural collective response. This primes online audiences to join the attack through “mechanical mobilisation”, unaware that the apparent consensus was manufactured and sustained using paid services.

The practice in Syria resembles documented strategies in other conflicts that use cross-border gender-based digital violence as an instrument of repression. Examples include the targeting of Filipina investigative journalist Maria Ressa, whose reporting was flooded with mocking reactions in an attempt to mislead recommendation algorithms, and tactics documented in Brazil, where Asian click farms were hired to systematically ridicule and smear independent women human rights defenders and politicians and undermine public solidarity with them. It also resembles tactics used by online armies in the Middle East that engage in rapid brigading to circumvent platform policies.[67]

Taken together, the evidence strongly supports the hypothesis that actors coordinated in advance and used inauthentic means of generating engagement to amplify targeting, silence defenders and force them to censor themselves. The available data cannot identify the purchaser, the means of purchase or the specific infrastructure used in the cases examined.

[6] UNESCO study on the case of Maria Ressa.
[7] Reports by the Inter-Parliamentary Union and Amnesty International on click farms in Brazil.



Figures 8, 9 and 10. Screenshots of accounts using the “Haha” reaction and their self-reported locations

The first names of the accounts and the text of the original posts were obscured, and the accounts were assigned codes R01–R30. Reaction counts and self-reported geographic locations were retained as evidence.

English translation of the text in the screenshots:

R01–R11: 1.6K reactions, 2 comments. Visible totals include 981 Likes, 680 Haha reactions and 10 Loves. [Post text redacted.] R02: 362 friends. R07: 128 followers, 1K following. R11: 803 followers, 8.2K following.

R12–R19: 1.2K reactions, 3 comments, 7 shares. Visible totals include 605 Haha reactions and 536 Likes. R14: lives in Antananarivo, Madagascar. R18: digital creator. R19: lives in and is from Jhapa District (942 friends).

R20–R30: 777 reactions, 6 comments, 1 share. Visible totals include 393 Haha reactions, 320 Likes and 56 Loves. [Post text redacted.] R27: lives in and is from Lusaka, Zambia. R28: [redacted].

04

Chapter 4

Effects and Consequences of the Targeting

4.1 The Rise of "Digital Authoritarianism" and the Shift in the Source of Threats

Qualitative interviews with women human rights defenders (P1–P6) reveal a marked shift in the risks that shape Syria's public sphere. As the centralised security apparatus that previously enforced self-censorship lost its grip, the sources of threat became more diffuse, giving online crowds and content creators greater influence. The change coincided with the rise of influencers with large followings. Their reach allows them to shape public perceptions, stir up collective hostility and direct waves of defamation, even where there is no evidence that their campaigns were authorised in advance.

The defenders describe overlapping fears. The greatest was being unable to protect family members inside Syria from the consequences of cross-border online incitement. Feeling that they had nowhere to turn for protection, some activists were forced to close their accounts for long periods and withdraw entirely from public debate.

> **"I went quiet. I was scared my family might be killed. The direct threats were to kill me if I went back to Syria, or to find out where my family was. It got to the point where they tagged the Ministry of Interior: 'Where is she? Bring her in!'"**
>
> **"There's absolutely nowhere you can turn. … There's no law to protect you at all. All you can do is run, stay away, or keep your distance from the law of the jungle that rules social media now. … The law itself gets used when it suits them. There's a law that protects a lot of the people doing the inciting, and the authorities give them full legitimacy when directing them to attack women activists."**

> ***— Interview with P6, an independent human rights and political activist.***

> **"It turns into a 'national wedding celebration' of bullying. The attackers drop the substance of the discussion and the human rights work and go after my appearance, my body and the details of my private life. People join in just because everyone else is doing it, so 'I have to join in'."**

> ***— Interview with P2, an investigative journalist and co-founder of a media outlet.***

The defenders' accounts point to a fundamental shift in what drives self-censorship. Previously, they censored themselves to avoid reprisals from the security apparatus. Today, they seek to avoid open incitement led by influencers with large audiences. Some influencers attended meetings and public events with official bodies and local authorities after December 2024. Any finding of direct institutional backing or direction for these campaigns remains subject to the report's strict evidentiary standards. The investigation establishes that the parties make use of one another's political and social influence. Establishing that these influencers are organisationally subordinate to the authorities would require conclusive technical and judicial evidence.

Field monitoring also shows that digital violence can overlap with offline harassment and stalking. In one documented incident in Damascus following a civic gathering, a defender was harassed in the street and the incident was filmed. When the video was shared on public pages, the incident became a fierce smear campaign. At the height of the initial wave, her page received about 100 abusive comments and sexual insults per minute whenever she posted a clarification, according to monitoring by SEEN's Department of Documentation and Studies. She eventually disabled her account to protect her psychological and physical safety.

These dynamics produce a collective "chilling effect".[8] The intimidation silences the defender directly targeted and reaches other women beginning to speak out, prompting them to censor themselves and withdraw from public life.

[8] UNESCO, The Chilling: Global Trends in Online Violence Against Women Journalists, research discussion paper.

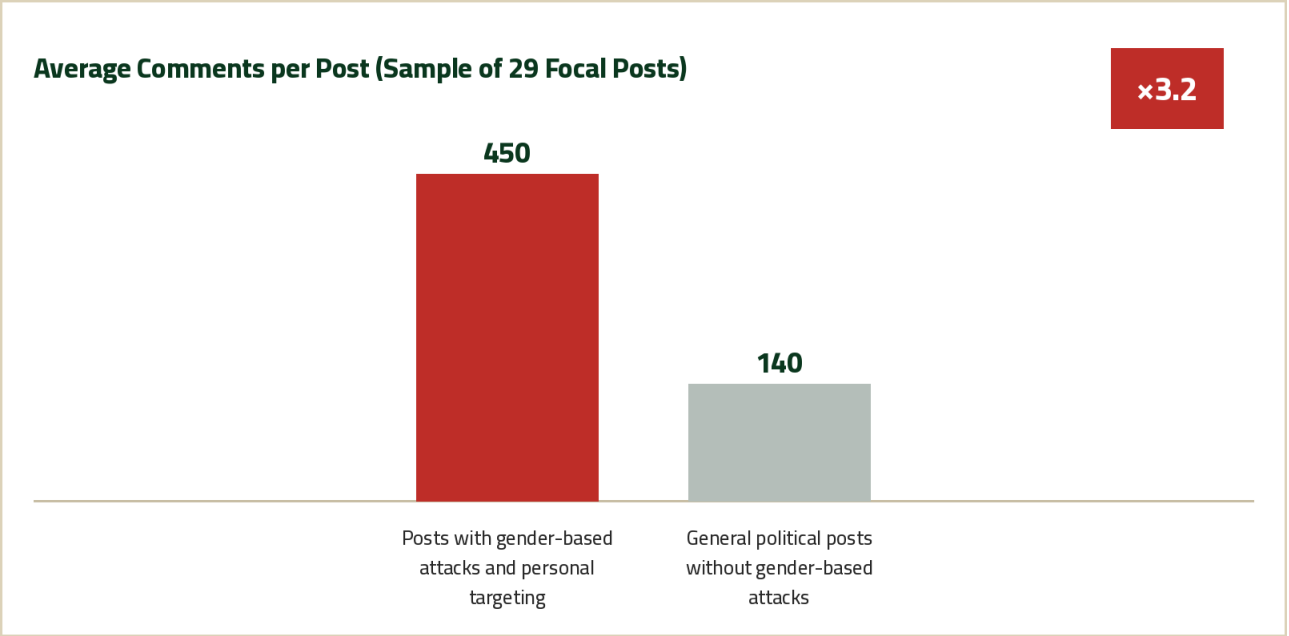


Figure 11. Relative engagement multiplier for posts containing gender-based attacks compared with general posts

The chart compares engagement in the sample of 29 focal posts examined. Average comments rise from 140 for general political posts to 450 for posts that target women on the basis of gender and attack them personally. That is 3.2 times as many comments, an increase of 221%. Those behind the attacks exploit this to inflict immediate psychological shock and force the targets to withdraw.

4.2 Offline Consequences – Convergence of Violence and Exclusion

Digital violence against Syrian women human rights defenders is not merely "virtual harm". Documented cases show that it can precede or coincide with physical violence. The report documents this "convergence", in which online incitement overlaps with offline punitive measures and social pressure.

- **Judicial and administrative measures.** The case of H1, a prominent defender who heads the feminist organisation O1, is the most serious example. In April 2025, a fierce online smear campaign built around an excerpt from a leaked recording of unknown origin coincided with threats to burn down the organisation's office. Within days, the online incitement was followed by official action. Local authorities filed a complaint against the defender, the police closed the organisation's only office in north-western Syria, and her family was pursued. The timing of the online campaign and official measures suggests that the campaign provided cover for the authorities' actions.
- **Coercion through family pressure (proxy censorship).** Attackers target defenders' families to pressure them into silence. The attack on P1 extended to her

family through inciting remarks such as “There are no men in the house”. P6 reported that relatives living inside Syria had been threatened with death and arrest in an effort to silence her. In conservative settings, this stigmatisation compels families, driven by fear or a desire to avert “scandal”, to exert coercive pressure on activists to resign or withdraw from public life.

- **Cross-border threats.** Documented cases show that living outside Syria, in a country of refuge, does not shield defenders from physical threats. Fadwa Mahmoud (P3) received direct threats through her private accounts giving the precise street and location of her home in a European city, with messages such as “You’re in (Berlin), and we can get to you”. This illustrates how online intimidation can escalate from verbal harassment to transnational repression aimed at instilling fear of being killed and restricting the defender’s freedom of movement and human rights work.

4.3 Effects on the Defenders’ Digital Activity

A comparison of the public online activity of 12 targeted defenders, measuring average weekly posts across three periods (a baseline of two weeks before a campaign began, the peak of the targeting, and a post-campaign period of two weeks after the wave of attacks subsided), showed that average posting fell from 14 posts per week before a campaign to 3 per week during it, recovering partially to 5 per week afterwards. Four defenders temporarily closed their accounts, and eight tightened their privacy settings and restricted comments. Human rights content concerning public affairs fell from 78% to 24% of their output. Where qualitative testimony corroborates the observed changes, it supports a direct causal link with the campaigns. Other changes are presented as temporal associations.

05

Chapter 5

Protection Gaps and Challenges

Online targeting relies on social engagement and technical mechanisms to spread. The legal and regulatory environment allows it to persist and magnifies its effects. Applied legal analysis of the documented incidents reveals conditions conducive to impunity, sustained by a twofold, systemic gap. The cases in this chapter show how selective enforcement of domestic legislation turns the law, which should offer protection, into an instrument of intimidation and deterrence. At the same time, major digital platforms fail to meet their due diligence responsibilities (duty of care) and disregard the Syrian context, leaving women human rights defenders without technical safeguards against campaigns of mobilisation.

5.1 Applied Protection Matrix

Rather than providing a theoretical review of legal provisions, the investigation asks a practical question: what protection should have been available to each defender in the documented cases, and where did it fail?

Documented case	Observed online conduct	Legal or regulatory protection that should apply	Where protection failed
Participant P1	Image manipulation, an AI-generated fabricated video, operation of a page under a false identity, and systematic sectarian incitement.	Domestic: Article 13(2) of the Constitutional Declaration and the cybercrime law's provisions on impersonation and online defamation. Platforms: policies prohibiting and providing for removal of non-consensual deepfakes.	Domestic: no safe complaint channels for residents outside areas under direct control, and a lack of trust in the courts. Platforms: slow responses and a page takedown that failed to prevent instigators from immediately setting up replacement pages.

Documented case	Observed online conduct	Legal or regulatory protection that should apply	Where protection failed
Defender H1	A multi-pronged smear campaign, public calls to burn the organisation's office, and death threats.	International: Article 20(2) of the International Covenant on Civil and Political Rights (prohibition of incitement to violence). Domestic: laws protecting civil society organisations and the right to engage in civic work.	Convergence of violence: local authorities used the online campaign as cover and a security pretext to close the organisation's only office in north-western Syria and pursue the defender and her family, directly advancing the perpetrators' objectives.
Participant Fadwa Mahmoud (P3)	A direct threat of physical harm that included her home address in a European city, political stigmatisation and accusations of being an agent.	International: criminal law in the country of residence and the European Union's Digital Services Act (DSA), which is binding on platforms in the EU.	The entire burden of documenting evidence for criminal proceedings fell on the victim, who was already psychologically exhausted. No accredited institutional intermediary (trusted flagger) was available to follow up on cross-border violations.
Participant P2 / institutional targeting	A feminist media outlet was targeted to stop it from exposing sexual violence and violence on the ground.	Domestic: Article 13(1) of the Constitutional Declaration, which guarantees freedom of the press, publishing and media work.	The courts and the public treated reporting on sexual crimes as an offence against public morals, leaving investigative journalism without professional protection.
Participant P6	Death threats, pursuit of relatives in Syria to put pressure on her, and public tags calling on law enforcement to arrest her.	Domestic: criminal provisions on threats, psychological blackmail and misuse of the internet. Platforms: protocols to protect defenders from cross-border coordination and the pursuit of family members (doxxing).	Filing a complaint became a risk in itself, increasing the danger of arrest and reprisals against family members in Syria. Platform algorithms failed to connect the targeting of the defender with threats to her relatives, forcing her to censor herself and suspend her activity for long periods.

5.2 Legislation as a Tool of Repression – Selective Enforcement and Gender Blindness

Syria's legal framework for online activity, in particular Cybercrime Law No. 20 of 2022, presents a structural problem that has deepened during the transition following the December 2024 shift in control. The law remains in force in practice. Its broadly worded

provisions are at odds with the principles of the Constitutional Declaration and international standards on freedom of expression and gender justice.

In June 2026, the Ministry of Justice announced its intention to review the law and redraft its problematic provisions.[9] It subsequently issued Circular No. 26[10] to regulate wanted notices and restrict detention to cases in which the public prosecution has authorised it on the basis of material evidence. These limited procedural measures left the substantive offences intact and were not accompanied by a suspension of the law, as independent human rights organisations had demanded. Two features therefore continue to characterise the law's application in practice.

- **Structural gender blindness.** The law does not define or recognise technology-facilitated gender-based violence (TFGBV). Courts and cybercrime units process cases involving women human rights defenders under inadequate conventional criminal provisions, such as Articles 23 and 24 on online defamation and insult and Articles 21 and 22 on violations of privacy. This legal framing overlooks the following.
  - The collective and networked nature of campaigns, including mechanical mobilisation, directed behaviour and engagement-flooding tactics.
  - Threats involving non-consensual uses of AI and deepfakes specifically designed to destroy women's reputations.
  - The complete absence of safeguards and urgent interim measures, such as emergency judicial protection orders, procedures compelling service providers to remove content immediately, and confidential complaint channels that take account of the sensitivity of honour and family stigmatisation in local communities.
- **Continued selective enforcement as a means of deterrence.** Despite restrictions on pretrial detention, the documented cases show that defamation provisions and broadly framed offences are applied with exceptional flexibility and speed when criticism targets influential figures in the new administration or local actors with an armed presence. Influential pages and networks, meanwhile, run documented campaigns to demonise and defame defenders and accuse them of betrayal. The campaigns meet the threshold for incitement to physical violence under the UN Rabat Plan of Action. Law enforcement agencies refrain from acting

[9] Statement by the Syrian Ministry of Justice, June 2026: review of the Cybercrime Law (Law No. 20 of 2022).
[10] Judicial Circular No. 26, issued by the Ministry of Justice of the Syrian Arab Republic on 9 June 2026, regulating procedures in cybercrime cases and safeguarding citizens' rights and freedoms.

on their own initiative, citing the absence of a direct complaint by the person affected or the difficulty of tracing accounts operating under false identities.

This disparity transforms the law and its enforcement agencies from sources of redress into instruments that perpetuate intimidation and impunity. It helps explain why the defenders and survivors interviewed, including P1 and P6, chose to stay away from the courts. Filing a complaint in these conditions can expose their private lives and endanger their families in a judicial system lacking safeguards for judicial independence.

5.3 Platforms' Failure to Exercise Due Diligence in the Syrian Context

Social media companies have an independent and direct human rights responsibility under the UN Guiding Principles on Business and Human Rights. A comparison of platform policies with the events recorded in the digital archive shows clear failures of human rights due diligence in Syria.

- **Meta's policy rollback and the erosion of proactive protection.** In early 2025, Meta relaxed its hate speech moderation policies, allowing more degrading language in political debate and weakening women's online safety. The end of "proactive enforcement" shifted the full burden of reporting and substantiation to the victim. Algorithms and automated screening systems have limited capacity to pick up meanings conveyed through Syrian dialects or coded sectarian and gendered language, including indirect metaphors about honour and derogatory epithets. As a result, these systems fail to intervene and help amplify abusive content on the basis of high engagement.[11]
- **Telegram as a safe haven and centre for coordination.** Despite the platform's repeated claims that it is developing its moderation tools, Telegram channels and feeds remain readily available "safe havens" for sharing manipulated material, coordinating smear campaigns and identifying targets. Content is then transferred to platforms with wider reach, such as Facebook. These campaigns take advantage of the one-way structure of channels and closed groups, which leaves victims unable to reply or challenge the claims.

Inadequate legislation and selective enforcement in Syria, combined with international platforms' failure to take technical action, create conditions for structural impunity. In this environment, online violence operates unchecked, supplying the "mental element"

[11] Meta Hateful Conduct policy updates, January 2025.

and preparing the ground for silencing and isolating defenders offline. Syria's public sphere is becoming increasingly hostile to women's active presence and depriving them of their fundamental right to civic and political participation.

06

Chapter 6

Policy and Practical Recommendations

The findings of this investigation show that the online targeting of Syrian women human rights defenders is not merely “virtual harm” or a transient technical failure, but a complex mechanism of repression in which patriarchal structures, political polarisation and platform algorithms reinforce one another. Given the systemic and pervasive gaps in protection and access to remedy at both the governmental and technical levels, dismantling this machinery of silencing requires intervention on multiple levels. This means moving beyond victims’ individual responses to establish robust institutional, technical and legal safeguards.

Based on the documented evidence, SEEN Association for Civil Peace and the Centre for Information Resilience (CIR) make the following urgent recommendations.

6.1 To Syrian Lawmakers and Authorities (Ending Selective Enforcement)

- **Recognise the gendered nature of digital violence in law.** Introduce a precise, standalone legal definition of technology-facilitated gender-based violence (TFGBV) as part of comprehensive reforms to the Cybercrime Law and the general Penal Code. Explicitly criminalise the production and dissemination of non-consensual deepfakes, smear campaigns and directed brigading, and threats directed against defenders and their families. Establish urgent interim measures, including emergency judicial protection orders, to require the immediate blocking of abusive content and protect victims’ online spaces.
- **Suspend broadly worded provisions.** Immediately suspend broadly worded political provisions of the Cybercrime Law, particularly those concerning “weakening national sentiment” or “undermining the prestige of the state”. These are used in practice to deter women human rights defenders and critics. Move defamation and insult cases from criminal to civil jurisdiction, and end the use of pretrial detention and wanted notices as pre-emptive forms of punishment.

- **Address the "mental element" of attacks.** Put a decisive end to the practice of overlooking online incitement campaigns under the pretext of freedom of expression when they reach the threshold for incitement to hatred and violence under the six-part test of the UN Rabat Plan of Action. Enforce the public prosecution's duty to act on its own initiative against accounts, public figures and content creators that lead such incitement, preventing online incitement from paving the way for or justifying offline harm, such as the closure of civil society organisations' premises or threats to kill.

6.2 To Major Digital Platforms (Restoring Due Diligence)

- **Assess risks specific to Syria.** Require Meta and Telegram to regularly conduct in-depth risk assessments of the Syrian digital space as a "highly fragile and divided conflict environment", in accordance with the UN Guiding Principles on Business and Human Rights. Reverse the more permissive moderation policies adopted in early 2025 by Meta, which left women without protection from waves of mechanical mobilisation and directed behaviour.
- **Adapt algorithms to local usage and update lexicons.** Invest immediately in improving automated and human moderation of Syrian dialects and coded gendered language. Establish technical cooperation with local monitoring teams, which can regularly supply updated lexicons and lists of new expressions used to evade moderation.
- **Create rapid escalation channels.** Set up trusted emergency reporting channels for women human rights defenders and accredited civil society organisations. Ensure a response within 24 hours to requests to remove life-threatening content and fabricated videos. Establish a mechanism that prevents perpetrators from setting up replacement pages.

6.3 To Civil Society Organisations and Feminist Movements (Shifting to Proactive Support)

- **Adopt a proactive support strategy.** Move beyond a model that waits for victims to request help, which they rarely do because of self-censorship and family pressure. Set up early-intervention teams to detect campaigns as they begin and reach out confidentially to those targeted, offering digital safety protocols and psychological support before the harm escalates.

- **Put "proxy archiving" into practice.** Establish human rights networks to document and archive abusive content, comments and threats on behalf of the women targeted. This preserves evidence for future criminal proceedings and spares survivors the psychological burden of reading the insults and threats directed at them and their families.
- **Pursue strategic and cross-border litigation.** Build a coalition of volunteer lawyers to follow cases of transnational repression and invoke universal jurisdiction to document and prosecute those who send threats to defenders living in countries of refuge. Document the handling of domestic complaints to expose selective enforcement with legal evidence.

6.4 To Donors and the International Community (Funding the Response)

- **Support digital investigation infrastructure.** Direct funding towards projects that develop and use tools to analyse and assess online activity in Syria. This will support the transition from traditional descriptive documentation to digital investigations capable of uncovering networks engaged in "directed behaviour" and producing conclusive evidence for attribution.
- **Link transitional justice with the digital sphere.** Make the level of online protection for women human rights defenders and an end to systematic silencing a binding criterion when assessing local authorities' compliance with human rights standards and participation in peacebuilding processes. Include these indicators in reports of the Independent International Commission of Inquiry on the Syrian Arab Republic (COI) and the International, Impartial and Independent Mechanism (IIIM).

Conclusion

The investigation shows that the online targeting of Syrian women human rights defenders cannot be reduced to a by-product of unchecked freedom of expression. It is a systematic strategy of political exclusion that is redrawing the boundaries of public life in post-Assad Syria. Gendered punishment is at the forefront of this targeting dynamic, through smears focusing on women's bodies, attacks on their chastity and pressure on their families. This is accompanied by the withdrawal of community protection from the defender. It culminates in political stigmatisation and sectarian sorting that grant a "License to Abuse", leading to forced withdrawal, the closure of offices or direct death threats. Its destructive impact extends beyond the immediate target, producing a collective "chilling effect" that reaches any woman considering taking part in public life or raising questions of accountability and transitional justice.

P1's words, "We're being silenced … One way or another, we're being silenced", are a warning about what is at stake during the transition. The words of Fadwa Mahmoud (P3), "I'll keep writing, and I'll keep speaking out", show that this machinery of silencing is not inevitable. Law enforcement agencies, local authorities, major digital platforms, civil society organisations and international actors involved in peacebuilding share a direct responsibility to ensure that Syrian women can speak and participate in public life without paying for it with their safety and dignity.

Appendix A

Investigative Methodology

The report combines large-scale digital monitoring, automated text processing using the specialised SeenBERT language model, rigorous human-in-the-loop verification, investigations using open-source intelligence (OSINT), in-depth qualitative interviews and applied legal analysis. It uses these methods to examine patterns of technology-facilitated gender-based violence (TFGBV) and hate speech directed at Syrian women human rights defenders and activists after 8 December 2024, and to measure the direct chilling effect on their civic and political participation.

The data constitute an indicative, statistically verified sample of the public online spaces that could be monitored. They do not provide a complete count of all violations. The investigation uses the four levels of evidence set out in Section 7 to distinguish predictable social responses, or mechanical mobilisation, from directed behaviour. It attributes activity to a specific political or security actor only when corroborating evidence of a standard suitable for judicial proceedings is available.

The methodological framework comprises the following 14 stages and criteria.

1 Core Investigative Questions

The investigation seeks to answer six central questions.

1. What discursive and semantic features distinguish gender-based abuse, and how do they intersect with political stigmatisation and sectarian sorting to discredit women human rights defenders?
2. How do online attacks develop, and how do accounts and users take part?
3. What behavioural and technical indicators distinguish spontaneous mass engagement, or mechanical mobilisation, from directed behaviour and tactics for manipulating algorithms?
4. What measurable changes occur in the frequency and nature of defenders' online activity before, during and after a campaign?
5. How does the targeting affect defenders' personal and professional safety and their participation in public life?
6. What structural gaps in domestic legislation and digital platforms' due diligence policies prevent effective protection and redress for victims?

2 Data Scoping and Incident Definition

The investigation used two parallel approaches to collect and analyse publicly available content on Facebook and Telegram.

- **Network-led monitoring.** The team tracked public Syrian news pages, channels and groups where narratives or discussions concerning women human rights defenders and activists appeared. It focused on content showing signs of gender-based violence, defamation, incitement or threats.
- **Retrospective, target-led monitoring.** The team closely monitored the public accounts and pages of a number of targeted women human rights defenders and activists as focal points for tracing incoming attacks and their development, and for tracking changes in activity before, during and after targeting.

Definition of an Incident

For the purposes of this report, not every abusive comment or post is treated as a separate incident. A distinction is drawn between an individual item and an incident.

An incident comprises a series of posts, comments or other digital items targeting a particular defender or activist over a short period, with a common objective, narrative, starting point or pattern of dissemination. This makes it possible to analyse them as a single wave of attacks. Similar content alone does not establish a single incident. The links between items were assessed using verifiable indicators appropriate to each case, including timing, similar messaging, reposting, common starting points and the sequence in which the content spread.

3 Monitoring, Automated Classification and Lexicon Development (SeenBERT)

The locally developed language model SeenBERT was used to screen the large volume of data, detect patterns and identify material requiring in-depth human review.

The monitoring used a gender-focused lexicon developed for Syria and updated to capture idioms, local dialects, coded words and expressions used to evade automated content moderation.

The system initially grouped the material it flagged into five main analytical categories. Figure 3 combines the last two.

1. Sexual objectification and sexual insults
2. Moral and gender-based stigmatisation and defamation

3. Gendered contempt, belittling and denial of competence
4. Intimidation and physical threats
5. Image-based abuse and manipulated media

The classification serves as an analytical tool for monitoring. Items identified by the system undergo human review before entering the final database, so the nature of the content and whether a violation occurred are never determined by automation alone. To reduce classification errors, human reviewers assessed ambiguous cases involving sarcasm, double meanings, local expressions and religious or social language that could be used in more than one sense.

4 Human Verification (Human-in-the-Loop)

Researchers reviewed the initial automated results to do the following.

- Verify that the relevant forensic criteria were met and establish the context in which the material circulated.
- Remove false positives and distinguish purely political exchanges from gender-based targeting.
- Interpret double meanings, sarcasm and ambiguous local expressions.
- Keep likely instances of gender-based violence distinct from the control group of general political exchanges.
- Determine whether different items belong to the same incident or to separate incidents.

Where the evidence did not support classification with sufficient confidence, researchers treated the cases as inconclusive and did not use them to draw definitive findings.

5 Criteria for Selecting and Investigating Cases

The investigation moved from large-scale monitoring to the study of individual cases. Criteria included the severity and reach of the targeting, the clarity of the available digital evidence, indicators of synchronised or coordinated activity, and the ability to trace how the incident unfolded and measure changes in the targeted defender's activity. The cases were selected for their explanatory and investigative value in examining how targeting works, spreads and affects defenders. The sample was not designed to be statistically representative of all women human rights defenders or all forms of digital violence in Syria.

The 12 defenders were selected on the basis of their public activity after December 2024, exposure to a documented campaign and the existence of a public account that could be monitored. The selection also took into account geographic and professional diversity.

6 Actor Analysis and the Distinction Between Mass Engagement and Directed Behaviour

The investigation distinguished two types of digital behaviour, recognising that similarity in language alone does not establish coordination.

- **Unorganised mass engagement (mechanical mobilisation).** Users join an attack or repeat abusive language in similar ways and almost automatically in response to entrenched patriarchal cues. There is insufficient evidence of central direction or prior coordination.
- **Directed behaviour.** A combination of indicators may point to deliberate coordination among accounts or entities seeking to amplify content, a narrative or an activity artificially. These include unusual synchronisation, repeated similarities in content, sequential reposting or anomalous links between accounts. Several independent indicators are needed to establish a coordinated attack. Establishing behavioural coordination and attributing it to a particular actor require separate findings.

7 Levels of Evidence and Attribution (Attribution Confidence Levels)

The investigation used four levels to keep findings within the limits of the evidence.

1. **Behavioural description.** Establish the existence of abusive content and identify its semantic and gendered characteristics.
2. **Indicators of coordination.** Identify behavioural, temporal and network evidence of coordinated behaviour.
3. **Documented links.** Establish overlaps between accounts involved in a campaign and specific influential networks or platforms.
4. **Conclusive attribution.** Link a campaign to a specific security or political actor only where corroborating material evidence leaves no room for doubt.

8 Measuring the Effects on Online Activity

To assess the effects of targeting campaigns, the investigation analysed the public accounts of the women affected across three periods. Section 4.3 presents the findings.

1. **Baseline period.** Activity recorded during the two weeks before the campaign began.
2. **Targeting period.** The period spanning the escalation and peak of the wave of attacks.
3. **Post-campaign period.** Activity recorded during the two weeks after the wave of attacks subsided or ended.

The analysis covered posting frequency, types of content, the ratio of personal to public-interest content, temporary pauses, deletion of content, moves to more restrictive privacy settings, and changes in subject matter. Changes in activity are treated as direct effects only where supported by independent testimony and qualitative evidence. Other changes are presented as temporal associations.

9 Documentation and Preservation of Digital Evidence

The investigation preserved digital evidence in line with forensic protocols for investigations using open-source intelligence (OSINT). It recorded permalinks and UTC timestamps, captured full screenshots, extracted metadata and retained archived copies to safeguard material against later alteration or deletion by accounts involved in the incitement.

10 Qualitative Interviews

Six in-depth qualitative interviews were conducted with Syrian women human rights defenders, activists, journalists, and leaders of civil society and women's organisations.

- **P1.** Human rights defender and peacebuilder.
- **P2.** Investigative journalist and co-founder of an independent media outlet.
- **P3, Fadwa Mahmoud.** Political and human rights activist working on enforced disappearances, based abroad.
- **P4, Thuraya Hijazi.** Head of a women's rights and human rights organisation.
- **P5, Afraa Hashem.** Human rights defender and civic activist.
- **P6.** Independent human rights and political activist.

Interviews focused on violations that could not be documented through open sources, including private messages, attempted blackmail, physical threats and proxy censorship imposed through relatives. They were guided by the Murad Code and the principle of “do no harm”. The investigation required informed consent and encrypted the records, keeping them on secure, isolated storage media. Participants’ accounts were triangulated with independent digital evidence.

11 Applied Legal Analysis

The legal analysis examined the Syrian legislative framework, including Cybercrime Law No. 20 of 2022 and judicial circulars issued in 2026, along with the International Covenant on Civil and Political Rights and relevant international conventions. It examined gaps in implementation, selective enforcement of criminal provisions and the absence of gender-specific legal definitions. It also assessed whether major digital platforms met their due diligence responsibilities under UN standards on business and human rights.

12 Inclusion and Exclusion Criteria

Items and incidents were included in the final database if the activity fell within the study period, it was linked to a verifiable account or online space, the person targeted and the subject of the attack were clear, and there were sufficient indicators of one of the forms of violence covered. Material whose source, context or timing could not be verified was excluded.

13 Limitations

- **Closed spaces and deleted content.** The report focuses on public content. Private correspondence and closed groups are covered only where participants documented the material. Account deletion can also leave gaps in the timeline.
- **Visibility bias.** Available data centre on activists and defenders who are more prominent in public life. This may give an incomplete picture of the targeting of local activists whose presence on platforms is less visible.
- **Limits of automated classification.** SeenBERT can misinterpret dialect, double meanings or sarcasm. Human review therefore supplemented automated classification throughout.
- **Limits of attribution.** Similar language alone does not establish directed coordination. Attribution to a specific actor requires independent evidence.

- **Limits of impact measurement.** Changes in online activity can be associated with several factors beyond the campaign itself.
- **Limits of the qualitative sample.** The six interviews were selected purposively for explanatory depth. The sample does not lend itself to statistical generalisation.

14 Verification and Triangulation

The investigation triangulated the evidence by cross-checking digital sources, time-series analysis, network data, interviews and open sources. The report distinguishes among three kinds of findings.

1. What the digital evidence can establish directly.
2. What converging indicators point to with a high degree of confidence.
3. What remains an unresolved analytical hypothesis given the limits of the available data.

Appendix B

Statistical Classification of Likely Violations

Following rigorous human-in-the-loop review and verification of the archived data, the team updated the statistical classification matrix to reflect the consolidated forensic breakdown of items containing likely gender-based violations. It identified 2,295 items containing a likely violation, grouped into 29 distinct incidents. The investigation screened 7,793 items in the archive. A further 5,429 items formed the control group of general political exchanges with no gendered component, and 69 were classified as inconclusive.

The table gives the categories, their operational definitions, counts, proportions and units of measurement.

Violation category	Operational definition	Items with likely violations	Share of total (N = 2,295)	Counting method and unit
1. Moral and family-based stigmatisation	Attacks on honour and sexual reputation, incitement against the family, pressure through relatives ("There are no men in the house"), and accusations of betrayal tied to sectarian or political sorting to strip the defender of community protection.	**1,128**	**49.2%**	One post or comment, verified in context, containing a likely gender-based violation.
2. Gendered contempt, belittling and denial of competence	Denial of the defender's intellectual and political competence, belittling diminutives, discrimination based on age or appearance, and attempts to confine women to stereotyped roles of domestic dependence.	**577**	**25.1%**	One post or comment, verified in context, containing a likely gender-based violation.
3. Sexual objectification and attacks on chastity	Reducing the defender to her body, crude comments on appearance and intimate body parts, pornographic insinuations, sexually defamatory language and explicit vulgarity.	**403**	**17.6%**	One post or comment, verified in context, containing a likely gender-based violation.

Violation category	Operational definition	Items with likely violations	Share of total (N = 2,295)	Counting method and unit
4. Intimidation and physical threats	Direct threats to kill the defender or drag her through the streets, disclosure of where she lives or where she goes (doxxing), and public tags calling on security agencies to pursue her and her family.	**133**	**5.8%**	One post or comment, verified in context, containing a likely gender-based violation.
5. Image-based abuse and deepfakes	Degrading cropping of personal photographs, defamatory designs and drawings, and doctored videos or non-consensual AI-generated deepfakes.	**54**	**2.4%**	One verified visual item or instance of manipulated composite media.
Total	**Items meeting all criteria for a gender-based violation.**	**2,295**	**100%**	**Based on human review of 7,793 archived and verified items.**

Methodological Note

- Some summary charts, including Figure 3, combine the last two categories under “intimidation, physical threats and image-based abuse”. They total 187 cases, or 8.1% of the sample. Combining them highlights the functional relationship between image manipulation and threats of physical assault.
- All 5,429 items consisting of purely political exchanges and general criticism with no gendered component were excluded from the violations tally. They were retained as an ongoing control group to establish the baseline and compare engagement statistically.

References and Sources

1. Previous Institutional Reports and Studies

- **Syrian Center for Media and Freedom of Expression (SCM).** Tammam Saymoua and Lara Aizouqi, Smear Campaigns Against Women: Is the Threat Over? An investigation into the targeting of women activists as a tool of repression and political exclusion. Published as part of a joint project with Khatt 30, 27 January 2025.
- **SEEN Association for Civil Peace, in partnership with the Centre for Information Resilience (CIR).** *Hate Speech in the Syrian Digital Space (2025): Patterns, Meanings and Effects in a Conflict Environment. Department of Documentation and Studies, March 2026.*
- **SEEN Association for Civil Peace.** *Updated database and digital archive of technology-facilitated gender-based violence (TFGBV) in Syria. Department of Documentation and Archiving, 2025–2026.*
- **Independent International Commission of Inquiry on the Syrian Arab Republic (COI).** *Periodic reports to the United Nations Human Rights Council on violations of freedom of expression, restrictions on civic space and attacks on human rights defenders, 2024–2026.*

2. International Human Rights Frameworks and Covenants

- **United Nations.** Universal Declaration of Human Rights, Article 19 (freedom of opinion and expression) and Article 12 (protection from arbitrary interference with privacy).
- **United Nations General Assembly.** International Covenant on Civil and Political Rights (ICCPR), particularly Articles 19 and 20 (prohibition of incitement to hatred, discrimination or violence).
- **Murad Code.** Global standards for collecting and documenting evidence of sexual and gender-based violence in conflicts and political transitions, in line with the principle of “do no harm”.

3. Digital Platform Policies and Regulatory Standards

- **Meta.** Community Standards and updated policies on hate speech, bullying and harassment of public figures and women activists, 2025 and 2026 versions.
- **Telegram.** Terms of Service and content moderation guidelines on public channels and reporting abusive material, threats and defamation.
- **European Union.** Digital Services Act (DSA), used as a benchmark for the due diligence and risk assessment obligations of major platforms in highly polarised environments.

4. Technical Methods and Network Analysis

- **Natural language processing (NLP) models.** Research on developing and testing specialised language models, their application to local dialects and semantic analysis (Devlin et al., 2019, BERT, and Antoun et al., 2020, AraBERT), and the criteria used to refine SeenBERT's semantic classifications.
- **Social network analysis (SNA).** Methods for tracking information flows, identifying indicators of directed behaviour and analysing metadata on comment activity and synchronised engagement (François, 2019, the ABC Framework, and DFRLab protocols for analysing directed behaviour).

Systematic Silencing

Online targeting of women human rights defenders in Syria and attempts to exclude them after the fall of the regime

CIR CENTRE for INFORMATION RESILIENCE ×