RAPPORT D'ENQUÊTE

Un musellement systématique

Le ciblage numérique des défenseures des droits humains en Syrie et les tentatives de les exclure après la chute du régime



Centre for Information Resilience (CIR)

CENTRE for INFORMATION RESILIENCE

×

Seen pour la paix civile

Département Documentation et études

Rapport d’enquête publié par

L’association « Seen pour la paix civile »

en partenariat avec le Centre for Information Resilience (CIR)

Équipe de travail

Dommar Al Soleman
Directeur du département Documentation et études, Association Seen pour la paix civile

Youssef Wannous
Responsable du département Sources ouvertes (OSINT) et archivage

Reem Fadel
Responsable du département Médias

Oula Suleiman
Directrice du département Plaidoyer

Avertissement sur le contenu

Ce rapport contient des citations et des descriptions de propos injurieux, confessionnels et sexuels, ainsi que des menaces de violences fondées sur le genre. Seules les citations nécessaires à l’enquête ont été conservées.

Table des matières

Résumé exécutif **8**
Principales recommandations 12
Introduction de l'enquête :l'espace public comme espace de prédation **13**
1.1 Lien avec les études antérieures et le rapport institutionnel de 2025 14
1.2 Apport scientifique et limites de l'originalité 15
1.3 Le continuum de la violence : de l'« élément moral » numérique au crime physique 15
1.4 Cadre méthodologique : démonter « la machine » et choisir l'échantillon 16
Formes d'abus et schémas de ciblage : la cartographie des données **19**
2.1 Carte du ciblage : les quatre axes sémantiques 19
2.2 Intersectionnalité et « permis d'abuser » (License to Abuse) 23
2.3 Interaction numérique et limites de l'inférence d'amplification 24
Évolution des attaques et indicateurs techniques : entre mobilisation mécanique et comportement dirigé **27**
3.1 La mobilisation mécanique : la machine de répression sociale 27
3.2 Le comportement dirigé : orienter et amplifier l'attaque 28
3.3 La tactique de « l'armée du Haha » (Haha Army) et la tentative de manipuler les algorithmes 31
Effets et conséquences du ciblage **35**
4.1 L'essor de « l'autoritarisme électronique » et le déplacement de la menace 35
4.2 Au-delà d'Internet : convergence de la violence et exclusion physique 37
4.3 Effets du ciblage sur l'activité numérique des défenseures 38
Lacunes de protection et défis **39**
5.1 Matrice des défaillances de protection (Applied Protection Matrix) 39
5.2 La législation comme outil de répression : application sélective et absence de perspective de genre 40
5.3 La dérobade des plateformes : l'absence de « diligence raisonnable » dans le contexte syrien 42
Recommandations politiques et opérationnelles **44**
6.1 Au législateur et aux autorités syriennes (démanteler l'application sélective) 44
6.2 Aux grandes plateformes numériques (rétablir la diligence raisonnable) 45
6.3 Aux organisations de la société civile et aux mouvements féministes (passer au soutien proactif) 46
6.4 Aux bailleurs de fonds et à la communauté internationale (financer la riposte) 46
Conclusion **47**
Méthodologie et cadre de l'enquête (Investigative Methodology) **48**
1 Questions fondamentales de l'enquête 48
2 Périmètre des données et définition de « l'incident » (Data Scoping & Incident Definition) 49
3 Veille, classification automatisée et construction du lexique sémantique (SeenBERT) 50
4 Vérification humaine (Human-in-the-Loop) 50

5 Critères de sélection des cas et conduite de l’enquête 51
6 Analyse des acteurs : distinguer l’interaction du public et le comportement dirigé 51
7 Niveaux de preuve et d’attribution (Attribution Confidence Levels) 52
8 Mesure des effets sur l’activité numérique 52
9 Documentation et conservation des preuves numériques 52
10 Entretiens qualitatifs (Qualitative Interviews) 53
11 Analyse juridique appliquée (Applied Legal Analysis) 53
12 Critères d’inclusion et d’exclusion 53
13 Limites de la recherche (Limitations) 54
14 Principe de vérification et de triangulation (Triangulation) 54
Classification statistique des violations probables **55**
Note de contrôle méthodologique 56
Références et sources **58**

Avant de lire

Avertissements et limites de cadrage

Anonymisation et protection des défenseures

Les identités de trois des défenseures ayant participé aux entretiens qualitatifs ont été codées (P1, P2 et P6), tandis que trois d'entre elles ont choisi de révéler leur identité, en défi ouvert à la tentative de les faire taire (Mmes Fadwa Mahmoud, Thuraya Hijazi et Afraa Hashem). Les entités mentionnées dans le rapport ont également été codées (H1, O1), de même que les comptes injurieux et les influenceurs (M1–M3, A1, A2, C1–C4, S1–S6, X1, X2, R01–R30). Les noms et les photos ont été floutés dans les captures d'écran, et les noms d'un certain nombre de personnes et d'entités figurant dans les témoignages ont été supprimés afin d'éviter toute réidentification des participantes ou toute exposition d'elles-mêmes et de leurs familles à des représailles. Les enregistrements originaux sont conservés dans un environnement numérique chiffré et sécurisé.

Cadre conceptuel de la violence numérique (TFGBV)

Le rapport s'appuie sur la définition, établie par les agences des Nations unies, de la « violence fondée sur le genre facilitée par la technologie », à savoir :

> **« Tout acte de violence commis ou aggravé, en tout ou en partie, au moyen des technologies de l'information et des médias numériques, à l'encontre d'une personne en raison de son genre »**

Le rapport ne traite pas les pratiques observées (harcèlement obsessionnel, exploitation par hypertrucage, diffamation, discours de haine) comme un « préjudice virtuel », mais les considère comme un prélude à la violence physique. Il adopte également une approche « intersectionnelle », selon laquelle la gravité de cette violence se multiplie lorsque le genre croise l'engagement militant ou civique de la victime, devenant ainsi un instrument composite de répression et de mise sous silence.

Niveaux de preuve (Attribution Levels)

Le rapport s'en tient à une échelle de preuve à quatre degrés (détaillée à l'annexe A, point 7) :

01	02	03	04
Niveau 1 (description du comportement)	**Niveau 2 (indicateurs de coordination)**	**Niveau 3 (liens documentés)**	**Niveau 4 (attribution formelle)**
Documenter l’existence du contenu abusif ainsi que sa nature sémantique et genrée.	Documenter les indicateurs techniques et comportementaux du « comportement dirigé ».	Établir le recoupement entre les comptes engagés dans la campagne et des réseaux ou des plateformes influentes identifiés.	Le rapport ne tire aucune conclusion causale et ne procède à aucune attribution pénale à une entité précise, à moins que des preuves matérielles concordantes et irréfutables ne l’étayent.

Vue d'ensemble

Chiffres clés

Bilan de la veille et de la vérification pour la période de décembre 2024 à octobre 2026, sur Facebook et Telegram.

+50 000

éléments numériques bruts dans l'archive (publications et fils de commentaires)

7 793

publications et commentaires soumis à l'examen et à la vérification humaine dans 12 dossiers d'enquête

2 295

éléments comportant une violation genrée confirmée après vérification humaine



Répartition des violations confirmées par axe de ciblage

Stigmatisation morale, politique et confessionnelle	**1 128 · 49,2 %**
Mépris genré, dénigrement et disqualification	**577 · 25,1 %**
Objectification sexuelle et atteintes à la vertu	**403 · 17,6 %**
Intimidation et menaces physiques explicites	**133 · 5,8 %**
Atteintes visuelles et hypertrucages	**54 · 2,4 %**

70,5 %

des 5 423 comptes uniques actifs dans les campagnes de ciblage sont des comptes masculins

244

comptes récurrents se déplaçant en parallèle entre plusieurs foyers d'incitation, dont 59 au sein de l'échantillon d'un même influenceur

450 / 140

commentaires en moyenne sur une publication d'attaque genrée, contre une publication politique générale (+221 %)

6

entretiens qualitatifs approfondis avec des défenseures, militantes et journalistes (P1–P6)

29

publications focales « déclencheuses » comparées à la ligne de base

3

étapes du « permis d'abuser » : féminisation de la sanction, levée de la protection sociale, retrait de la qualité de défenseure

Résumé exécutif

Ce rapport examine les formes de **ciblage numérique dirigé**[1] visant les défenseures des droits humains et les militantes de la société civile en Syrie **au cours de la période qui a suivi** le 8 8) 2024 ديسمبر décembre 2024). Il met l'accent sur l'interaction fonctionnelle entre la violence fondée sur le genre et la stigmatisation politique et confessionnelle. Il analyse également les répercussions des campagnes de diffamation et de menaces sur la participation publique et l'action en faveur des droits humains. L'analyse part d'un constat central : les défenseures se trouvent dans une position d'extrême vulnérabilité, au croisement de deux identités qui heurtent la structure patriarcale dominante. Elles sont ciblées à la fois comme « femmes », dans une société qui associe l'honneur au corps, et comme « actrices de la vie publique » ouvrant des dossiers sensibles qui embarrassent autant les autorités de facto que les groupes locaux.

Le rapport adopte la perspective de **l'« intersectionnalité » (Intersectionality)** comme cadre d'analyse principal : la militante syrienne n'est pas ciblée en tant que « femme » de manière abstraite, mais en tant que femme appartenant à un milieu ou à une origine confessionnelle et régionale supposée, et en fonction de sa tenue vestimentaire et de sa position politique. Ce croisement produit le phénomène du **« permis d'abuser » (License to Abuse)** : on s'attache d'abord à isoler moralement la défenseure et à mettre en cause sa vertu et sa conduite familiale, avant de lui accoler une étiquette politique ou identitaire (par exemple : « fouloul » [résidus du régime], « Nousayriya » [terme péjoratif désignant les Alaouites], « Qasadiya » [partisane des Forces démocratiques syriennes (FDS)], « agente »). Une fois cette étiquette fixée, la protection sociale qui l'entourait tombe, et l'agression sexuelle et la diffamation cessent d'être des actes condamnables pour devenir une curée bénéficiant de la complicité et de l'encouragement collectifs, au nom de la défense des « valeurs de la révolution » ou de la « protection des valeurs sociales ».

Sur le plan méthodologique, l'enquête repose sur le traitement d'une archive numérique de plus de 50 000 éléments bruts issus de Facebook et de Telegram. À l'aide du modèle de traitement automatique du langage naturel **(SeenBERT)** et de son lexique sémantique genré et dynamique — développé conjointement par l'association Seen et le Centre for Information Resilience

[1] Tactique d'attaque numérique fondée sur une orientation verticale plutôt que sur une coordination horizontale : un acteur influent mobilise son influence pour orienter des foules d'abonnés afin d'envahir l'espace numérique d'une victime déterminée, dans le but de l'intimider et de la réduire au silence. Contrairement au « comportement coordonné », qui exige d'établir l'existence d'une organisation ou d'une communication préalable entre les assaillants, le comportement dirigé repose sur un « signal de départ » public (tel que l'incitation ou le cadrage négatif) qui pousse le public à mener une attaque collective (*Brigading*), générant ainsi un flot de violations tout en offrant à l'instigateur principal une marge pour se dédouaner de toute responsabilité directe.

(CIR) —, les données ont été soumises à une vérification humaine rigoureuse et exhaustive (Human-in-the-Loop) afin d'écarter les faux positifs et d'exclure les polémiques politiques non genrées. L'échantillon examiné s'est finalement établi à **7 793 éléments documentés**, parmi lesquels la vérification humaine a établi la présence d'une violation genrée dans **2 295 éléments (publications et commentaires)**, selon la définition retenue dans ce rapport. L'analyse s'est concentrée sur 12 cas de ciblage visant des défenseures des droits humains, répartis selon les axes principaux suivants :



Graphique illustrant la répartition des violations entre les cinq catégories. Bien qu'un même commentaire ou une même publication puisse relever de plusieurs des catégories mentionnées, nous l'avons classé statistiquement selon l'axe le plus saillant ou le poids sémantique dominant du texte, afin d'éviter les doubles comptages et de rattacher chaque élément à son contexte offensif principal.

En comparant de manière systématique les résultats du modèle (SeenBERT) avec les métadonnées (Metadata) et les contextes d'interaction, le rapport identifie un schéma comportemental récurrent que nous avons nommé « permis d'abuser » (License to Abuse).Les résultats montrent que ce dispositif de réduction au silence ne débat pas du contenu militant de la défenseure ; il repose sur une structure d'exclusion composite articulée autour de trois dimensions fonctionnelles. Bien qu'elles puissent coïncider ou intervertir leurs rôles selon le contexte, elles suivent généralement une même logique : le ciblage genré et le retrait de la protection sociale préparent la disqualification de sa légitimité civique et de sa qualité de défenseure des droits humains :

01	02	03
La féminisation de la sanction : le ciblage genré constitue l'empreinte dominante de la vague d'attaques, plutôt que la discussion des positions politiques. La veille enregistre un flot dense d'insultes sexuelles, de mises en cause de la vertu et de commentaires sur l'apparence physique, ainsi que l'utilisation de photos personnelles à des fins diffamatoires, afin d'exercer une pression psychologique rapide et d'enfermer la défenseure dans le registre de l'atteinte à sa réputation personnelle et familiale.	**Le retrait de la protection et du soutien social :** parallèlement à l'escalade de la violence genrée et à son « pic d'interaction », les mécanismes d'isolement social se déclenchent. Dans les milieux conservateurs, l'insistance sur la dimension morale de la défenseure crée une distance entre elle et son entourage, ce qui entraîne le recul de la solidarité publique à son égard et pousse les voix modérées à se retirer : la défenseure se retrouve privée du soutien social habituel.	**Le retrait de la qualité de défenseure :** qu'elle coïncide avec l'attaque genrée ou qu'elle la couronne, cette dynamique s'achève par le recours à des formules politiques et confessionnelles hostiles (par exemple : « fouloul », « Nousayriya », « Qasadiya », « agente »). La stigmatisation politique n'est pas avancée ici comme une accusation ouvrant un débat, mais comme un outil décisif pour dépouiller la défenseure de sa qualité civique et présenter l'agression comme un acte justifié contre une personne politiquement et moralement « livrée à la curée ».

Sur la base des résultats de la veille et de la mise en réseau des incidents, l'enquête conclut que cette interaction fonctionnelle imbriquée (où la violence genrée et le retrait de la protection et du soutien social préparent le retrait de la qualité militante) constitue une dynamique d'exclusion efficace. Qu'il résulte d'une réaction sociale masculiniste quasi automatique (mobilisation mécanique) ou qu'il soit exploité par des réseaux orientés (comportement dirigé), ce parcours poursuit un même objectif : affaiblir la crédibilité de la femme engagée dans l'espace public. Plutôt que de lui opposer des arguments politiques, il cible sa sphère personnelle et genrée et la prive de sa protection sociale et juridique, afin de l'isoler et de l'exclure.

La veille computationnelle a été complétée par six entretiens **qualitatifs approfondis (P1 à P6)**, qui confirment la transformation de l'espace numérique en une autorité de contrôle parallèle imposant une autocensure contrainte, dont les effets se prolongent dans le monde physique : pressions de l'entourage familial, fermeture de locaux d'organisations, menaces de mort transnationales. Le rapport documente un environnement d'impunité alimenté par un « double échec » : une législation nationale défaillante, marquée par l'aveuglement au genre et une application sélective, et une dérobade persistante des grandes plateformes numériques

(Meta et Telegram) face à leurs obligations de diligence raisonnable dans le contexte du conflit syrien.

> **L'enquête cherche principalement à répondre à la question centrale suivante : comment la violence numérique fondée sur le genre est-elle utilisée pour cibler les défenseures des droits humains en Syrie depuis la chute du régime, et quels schémas et mécanismes régissent sa diffusion et ses effets ?**

Pour décomposer cette question, le rapport s'attache, dans ses chapitres suivants, à répondre aux sous-questions ci-après à partir des données recueillies :

1. Quels sont les marqueurs discursifs et sémantiques des abus fondés sur le genre, et comment se combinent-ils avec la stigmatisation politique et l'assignation confessionnelle pour disqualifier les défenseures ?
2. Comment les attaques numériques évoluent-elles, et quelle est la nature de la participation des comptes et des utilisateurs ?
3. Quels indicateurs comportementaux et techniques distinguent l'interaction mécanique du public (mobilisation mécanique) du comportement dirigé et des tactiques de manipulation algorithmique ?
4. Quels changements mesurables affectent le rythme et la nature de l'activité numérique des défenseures au cours des trois périodes d'attaque (avant, pendant et après la campagne) ?
5. Quels sont les effets du ciblage sur la sécurité personnelle et professionnelle des défenseures et sur leur participation publique ?
6. Quelles sont les lacunes structurelles du système législatif national et des politiques de diligence raisonnable des plateformes numériques qui empêchent une protection et une réparation effectives pour les victimes ?

Principales recommandations

Aux grandes plateformes numériques : mieux comprendre les dialectes et les contextes syriens, offrir des voies de signalement sûres et claires, et ne pas tenir pour acquis qu'un niveau d'interaction élevé est la preuve exclusive de la qualité d'un contenu sans examiner la possibilité d'une amplification algorithmique délibérée du discours de haine.	**Aux organisations de la société civile :** adopter des stratégies de soutien proactif, notamment en documentant et en conservant numériquement les preuves au nom des personnes touchées (avec leur consentement), afin d'épargner aux survivantes le fardeau psychologique de relire les abus.	**Aux bailleurs de fonds et à la communauté internationale :** orienter le soutien vers la création de bases de données documentées et sécurisées sur les droits humains, et vers le développement d'outils d'enquête numérique capables de mettre au jour les réseaux de comportement dirigé et de fournir des preuves d'attribution solides.	**Au législateur et aux autorités syriennes :** inscrire une définition juridique autonome incriminant la « violence numérique fondée sur le genre et facilitée par la technologie » (TFGBV) et les hypertrucages non consentis ; suspendre les dispositions vagues de la loi sur la cybercriminalité et transférer les affaires de diffamation et d'injure à la compétence civile ; activer la saisine d'office du ministère public pour poursuivre les campagnes d'incitation conformément au « Plan d'action de Rabat », sans exiger de plainte personnelle.

01

Chapitre premier

Introduction de l’enquête : l’espace public comme espace de prédation

L’espace numérique syrien n’a jamais été isolé des lignes de fracture politiques et sociales qui déchirent la réalité physique ; toutefois, les changements qui ont suivi la chute du régime le 8 décembre 2024 ont inauguré une nouvelle ère d’ouverture sociale et politique. Cette ouverture, qui s’est traduite par une hausse notable de l’engagement numérique des citoyens, s’est accompagnée d’une forte polarisation et d’une absence quasi totale d’outils et de processus de justice transitionnelle. Dans ce vide sécuritaire et juridique, les plateformes numériques — au premier rang desquelles Facebook et Telegram, les plus répandues en Syrie — sont devenues des arènes de substitution pour régler des comptes politiques et confessionnels. Elles fonctionnent aussi comme des instruments coercitifs qui redessinent les frontières du nouvel espace public et déterminent qui est autorisé à y participer.

Au cœur de ces transformations, les défenseures des droits humains et les militantes de la société civile se sont retrouvées face à une vague croissante de violences numériques fondées sur le genre et facilitées par la technologie (TFGBV). Cette vague n’est pas la simple continuation d’une forme traditionnelle de discrimination : elle s’est manifestée comme un outil de répression et de dissuasion politique « actualisé », visant l’exclusion systématique des femmes qui ouvrent des dossiers de redevabilité ou contestent les récits dominants.

La machine de ciblage actuelle ne surgit pas de nulle part : elle hérite d’une structure répressive mise en place par l’ancien régime. Le Centre syrien pour les médias et la liberté d’expression (janvier 2025)[2] a ainsi documenté l’usage de la diffamation comme « élément systématique » pour transformer les militantes en parias au sein de leurs propres communautés. Aujourd’hui, les données de terrain montrent que le système de mise sous silence ne s’est pas arrêté : ses outils sont passés à de nouveaux acteurs locaux et à des réseaux d’influence sociale. C’est ce qu’a exprimé l’une des défenseures (participante P6), en soulignant que le basculement des

[2] Centre syrien pour les médias et la liberté d’expression (SCM) : enquête « Campagnes de dénigrement des femmes : la menace est-elle passée ? Le ciblage des militantes comme outil de répression et d’exclusion politique », par Tamam Saymoua et Lara Aizouqi, publiée dans le cadre d’un projet commun (Khatt 30), 27 janvier 2025.

rapports de force a déplacé le centre du danger de l'emprise sécuritaire directe vers un espace ouvert, livré à la curée sociale et numérique.

> **« Par exemple, moi, ça m'est arrivé plusieurs fois : mon profil a circulé à plusieurs endroits, mon commentaire a circulé à plusieurs endroits. J'ai reçu énormément d'insultes et de menaces de mort, des plaintes, et même des mentions adressées au ministère de l'Intérieur ; une fois, par exemple : "Elle est où, celle-là ? Ramenez-la !", et des menaces de mort si je rentrais au pays. »**

1.1 Lien avec les études antérieures et le rapport institutionnel de 2025

Cette enquête s'inscrit dans le programme de recherche mené par le département Documentation et études de l'association « Seen » pour la paix civile et le Centre for Information Resilience (CIR) afin de surveiller le discours de haine dans l'espace numérique syrien et d'en analyser les schémas, les significations et les effets. Le rapport « Le discours de haine dans l'espace numérique syrien (2025) »[3] a montré que ce discours ne se limite pas aux insultes verbales directes, mais se manifeste aussi par des formes de stigmatisation, d'exclusion et d'incitation liées aux identités et aux appartenances sociales, politiques et confessionnelles, dans des contextes où le langage hostile peut rejoindre les tensions sociales et la violence. Ce rapport signalait également que des journalistes et des défenseures des droits humains avaient été exposées à des menaces à caractère sexuel ainsi qu'à des campagnes de diffamation et de pression visant leur présence en ligne. Pour autant, l'analyse approfondie des mécanismes propres au ciblage des femmes défenseures des droits humains et des militantes, tout comme l'interaction entre violence fondée sur le genre et stigmatisation politique, sociale et confessionnelle, ne constituait pas un axe central de la conception de ce rapport, essentiellement consacré à l'analyse du discours de haine, de ses formes et de sa diffusion dans l'espace numérique syrien.

Dans cette perspective, la présente enquête pose une question plus précise : comment s'exerce le ciblage numérique fondé sur le genre et facilité par la technologie (TFGBV) contre les femmes défenseures des droits humains et les militantes en Syrie, quels schémas et mécanismes prend-il, et comment affecte-t-il leur participation publique, leur sentiment de sécurité et leur capacité à poursuivre leur engagement ? Pour y répondre, l'enquête entend passer de la surveillance générale du discours de haine à l'étude des cas, des mécanismes de ciblage et de leurs contextes, en s'appuyant sur l'analyse du contenu numérique, la vérification humaine des

[3] Association « Seen pour la paix civile », en partenariat avec le « Centre for Information Resilience (CIR) » : rapport « Le discours de haine dans l'espace numérique syrien (2025) : schémas, significations et effets dans un environnement de conflit », département Documentation et études, mars 2026.

résultats, ainsi que les études de cas et les entretiens qualitatifs prévus par sa méthodologie. Ainsi, l'enquête ne part pas du principe que le ciblage des femmes ou la violence genrée en ligne n'auraient jamais été documentés dans les travaux de l'association : elle s'appuie au contraire sur ce que le rapport précédent a établi, cherche à approfondir la compréhension de la dimension de genre, des mécanismes de ciblage et de leurs effets, et à fournir un matériau analytique complémentaire à la base de connaissances institutionnelle, en évitant de reprendre les résultats antérieurs au-delà de ce qui est nécessaire pour cadrer le sujet et comparer les conclusions.

1.2 Apport scientifique et limites de l'originalité

L'enquête ne présume pas que la violence numérique à l'encontre des femmes syriennes soit un phénomène jusqu'ici non documenté. Son apport réside dans la combinaison de la veille des contenus, de la vérification humaine et de l'étude de cas, puis dans l'examen de la chronologie du ciblage, des indicateurs de coordination et des évolutions de l'activité numérique des personnes ciblées. Le terme « permis d'abuser » y est employé comme une hypothèse explicative vérifiable.

La solidité de cet apport dépend de la publication de données claires sur la taille de l'échantillon de veille, les critères de sélection des incidents, les limites de performance de SeenBERT, ainsi que d'exemples documentés montrant comment l'analyse passe de l'observation à la conclusion de coordination ou d'attribution. C'est pourquoi le rapport ne se présente pas comme « le premier » ou « le seul » dans ce domaine ; il définit sa contribution à partir de ce que ses matériaux et ses outils établissent effectivement.

1.3 Le continuum de la violence : de l'« élément moral » numérique au crime physique

La violence numérique ne se limite pas aux abus en ligne : elle peut s'étendre au monde réel et se transformer en menaces ou en agressions physiques. Les cas documentés dans cette enquête indiquent que le discours d'incitation sur les plateformes numériques peut contribuer à créer un environnement favorable à la violence contre les défenseures des droits humains, sans que cela implique nécessairement un lien de causalité direct entre l'incitation numérique et la violence sur le terrain.

Les mots écrits sur les écrans contribuent souvent à bâtir l'infrastructure psychologique et sociale qui pourra ensuite légitimer l'agression.

En ce sens, les effets des campagnes numériques ne se limitent pas au préjudice psychologique : ils débordent de l'écran pour franchir les murs de la famille et de la société (par des pressions

exercées sur les familles des défenseures pour les faire taire, sous prétexte d'« éviter le scandale » lié à la stigmatisation en ligne), et se prolongent dans plusieurs cas en menaces de mort, en poursuites sécuritaires ou en mesures administratives frappant les locaux d'organisations féministes (comme le montrent les cas documentés).

Cette imbrication manifeste entre le discours d'incitation et l'acte physique indique que les plateformes numériques peuvent parfois devenir des espaces de mobilisation contribuant à produire l'exclusion politique et à distribuer la « sanction sociale », d'où la nécessité urgente de traiter ce discours dans des cadres de protection et de redevabilité.

1.4 Cadre méthodologique : démonter « la machine » et choisir l'échantillon

La méthodologie a été conçue pour associer description sociologique et enquête fondée sur les preuves, selon deux axes principaux :

- **Délimitation de l'échantillon et des foyers de veille (Data Sampling & Scoping) :** la construction de l'échantillon repose sur une stratégie de veille double, combinant **l'axe proactif (en réseau)** et **l'axe inverse (centré sur les foyers)** : l'équipe de recherche a identifié un ensemble ciblé de comptes, de pages publiques et de groupes actifs dans la vie publique syrienne sur Facebook et Telegram, à l'égard desquels existaient des indices de suspicion et un degré élevé de doute méthodique, fondé sur leurs antécédents dans la conduite et la diffusion de discours de haine et d'incitation. L'ensemble des flux de ces comptes a été archivé, parallèlement à une veille approfondie des pages publiques de plusieurs défenseures et militantes ciblées, afin de suivre le reflux des attaques. Ce balayage a produit une archive brute exhaustive de plus de **50 000 éléments numériques** (publications et fils de commentaires) couvrant des cas survenus entre décembre 2024 et octobre 2026.
- **Classification algorithmique et construction du lexique sémantique (SeenBERT & Dynamic Lexicon) :** la base de données initiale a été soumise au modèle linguistique spécialisé **(SeenBERT)** afin de trier le discours de haine et d'identifier les textes porteurs de ciblage. Le modèle s'appuie sur un lexique sémantique genré spécifique, qui n'a pas été construit de manière figée et préalable, mais selon une méthode de **« mise à jour cumulative et périodique »** : à chaque étape d'extraction et de tri des contenus des comptes archivés, l'équipe d'enquête relevait tout terme, qualificatif, métaphore ou formule d'euphémisation nouvelle porteuse d'une charge genrée injurieuse (métaphores de l'honneur, détournement des dialectes parlés, sobriquets dévalorisants codés, etc.), l'intégrait au lexique et en nourrissait le modèle pour le réentraîner à repérer les formes de contournement des algorithmes des plateformes.

- **Vérification humaine et unité de mesure des incidents (Human-in-the-Loop & Incident Aggregation) :** pour garantir l'intégrité scientifique, les résultats automatisés ont fait l'objet d'une révision et d'un contrôle humains exhaustifs, afin d'écarter les polémiques purement politiques et les faux positifs : 7 793 publications et commentaires documentés et vérifiés ont ainsi été recensés et examinés à travers 12 dossiers d'enquête validés. Un critère de regroupement des éléments en « incidents de ciblage genré distincts » a été appliqué (proximité temporelle, unité du récit, point de départ), aboutissant à un bilan de 2 295 éléments comportant une violation genrée selon la définition retenue dans ce rapport, face à 5 429 éléments constituant le groupe témoin de la polémique politique générale non genrée, et 69 éléments jugés non concluants.

> « Pour évaluer la fiabilité de la classification automatique, deux relecteurs indépendants ont soumis les résultats du tri initial à un double contrôle humain, afin d'assurer la cohérence de l'arbitrage ; comparé à cette classification de référence, le modèle s'est montré très performant pour détecter les marqueurs explicites de stigmatisation et les insultes directes, tandis que tous les cas ambigus relevant de l'ironie, des dialectes locaux et du double sens ont été renvoyés à une révision humaine complète, afin qu'aucune classification automatique ne soit retenue isolément. »

- **Enquête et identification des acteurs :** l'enquête distingue rigoureusement la « mobilisation mécanique » (interaction mécanique et stéréotypée du public, régie par une réaction conditionnée par la structure patriarcale) du « comportement dirigé », qui repose sur une orientation préalable, un transfert simultané de contenus et de hashtags, et une tentative de manipulation des algorithmes pour amplifier les abus.

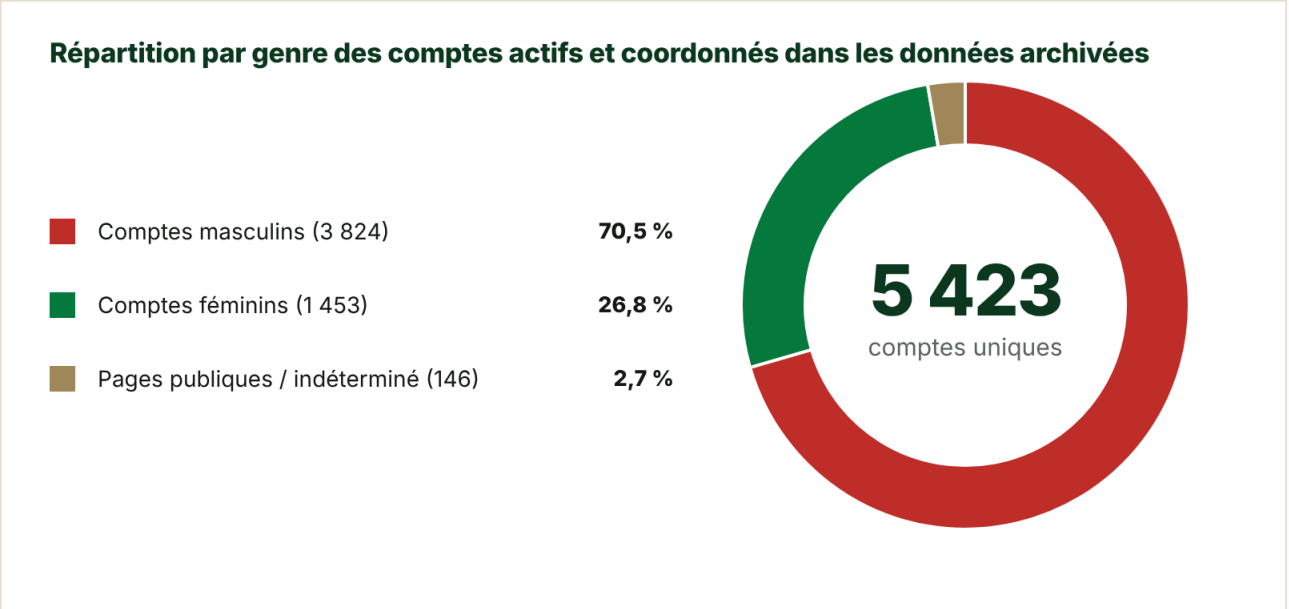


Figure (1) : la domination masculine parmi les comptes actifs dans les campagnes de ciblage genré

Le graphique présente l'analyse démographique de 5 423 comptes uniques relevés dans les données archivées sur Facebook et Telegram : les comptes d'utilisateurs masculins dominent avec 70,5 % (3 824 comptes) et constituent la principale force motrice des campagnes d'incitation et d'abus. Viennent ensuite les comptes féminins, avec 26,8 % (1 453 comptes), tandis que les pages publiques et les comptes indéterminés représentent les 2,7 % restants (146 comptes).

Le genre des comptes a été déterminé à partir du nom et du genre déclaré, de la photo de profil et des marques grammaticales du masculin et du féminin dans les publications ; les cas ambigus et les pages publiques ont été classés dans la catégorie « indéterminé ». Cette classification ne reflète pas nécessairement l'identité réelle des personnes qui gèrent les comptes.

Ces mécanismes ont été complétés en reliant ces données aux études de cas et aux entretiens qualitatifs approfondis menés avec les défenseures syriennes, et en s'appuyant sur les « niveaux de preuve » (description du comportement, indicateurs de coordination, liens documentés, attribution formelle) afin de rendre afin de prémunir les conclusions du rapport contre la contestation

Les chapitres suivants cherchent à répondre à une question centrale : comment fonctionne la machine de mise sous silence numérique en Syrie, et à qui incombe la responsabilité de l'arrêter ?

02

Chapitre deux

Formes d’abus et schémas de ciblage : la cartographie des données

Les dynamiques d’exclusion numérique des défenseures syriennes ne peuvent être comprises à partir des seuls récits individuels : passer de la preuve du « préjudice personnel » à celle de la « méthode répressive » exige une dissection minutieuse de données massives. Ce chapitre combine les résultats du modèle de veille linguistique (SeenBERT) et les témoignages qualitatifs des défenseures afin de dresser une carte précise des axes d’attaque et de mettre au jour la manière dont la polarisation politique et la violence genrée se conjuguent pour produire ce que le rapport appelle le « permis d’abuser ».

2.1 Carte du ciblage : les quatre axes sémantiques

Les témoignages qualitatifs des participantes indiquent une transformation de la nature des attaques depuis la chute du régime : le débat politique direct recule au profit d’un ciblage concentré sur le corps, l’honneur et la famille. Mme Fadwa Mahmoud (P3), militante chevronnée ayant connu des décennies d’engagement public, affirme n’avoir jamais subi, de toute sa carrière, autant d’insultes touchant à « l’honneur » ou à « l’apparence ».

Cette transformation qualitative concorde avec les schémas dégagés des **données archivées**, qui ont permis de répartir les violations selon des axes sémantiques principaux :

- **Stigmatisation morale (Moral Stigmatization) :** dans le contexte syrien, cette stigmatisation revêt un caractère éminemment familial. Les assaillants ne se contentent pas de taxer la militante de « dépravation » ou de s’en prendre à ses choix vestimentaires : ils s’en prennent aussi aux hommes de sa famille. Des formules telles que « il n’y a pas d’homme à la maison » traduisent une tentative manifeste de transformer la famille en instrument de répression contre la défenseure. Les assaillants se concentrent de manière stéréotypée sur la vie privée : la répartition finale de cet axe atteint **1 128 violations (soit 49,2 % de l’ensemble des violations)**, centrées sur l’atteinte à l’honneur et l’accusation de trahison fondée sur l’appartenance confessionnelle et politique. Les assaillants créent

également de faux comptes au nom des défenseures pour diffuser de fausses informations, les faire chanter sexuellement et les diffamer au quotidien.

text	threadingDepth	gender_targeting_category
يعتي مين منن حلي هلاه يحوا ويحكوا وقت ثان الشعب عم ينتلج ويتتل يا وين كانت هي [redacted] وين كتلوا بروهوا طعما راسيم بالطين	0	الترهيب والتهديد الجسدي
[redacted] يا قحبة يا شرمرطة يا مرابعة يا بلت المرابعي الحستتم وطتطتك وملابس النسوان مثل الكلاب وهواكم وصل فيرص وماضل هدا ما ساعتكم يا ولاد القعبة ورعم هيك الطلعو والتلينو واعستتكم بالصرامي يا عاقرة ولسا رح تتدعسيو يا مراهة	2	الترهيب والتهديد الجسدي
[redacted] معلش تعنا منحب لارهارمي بثكي درعس لابقهات وحكم لابقهات ولعلما لية نبعك يا اهبة. ! ١ سنة نحنا هم تكن صف والفو مثل لكلاب بتكن حكم مثل مابدكن. ويدكن مريكة بصلو دبين مثل مابدكن لد من اهتكن واهت لعرافقن الله بديم بدم. مابتة علبكن اسية تسورية متحوى؟	2	الترهيب والتهديد الجسدي
[redacted] يا عجوز يا شمطاعيا شميوقةبلول الشاعر ابو عمانة : العلوي الجيد هو العلوي الذي لايأتنس	1	الترهيب والتهديد الجسدي
[redacted] بدي اليجها واشق نمها هل هاهرة	0	الترهيب والتهديد الجسدي



Figure (2) : échantillon tiré des archives de l'association Seen pour la paix civile, ainsi que de messages directs reçus par certaines militantes

- **Mépris genré, dénigrement et disqualification (Gendered Contempt & Devaluation) : 577 violations probables (25,1 %) ; cet axe vise** à dépouiller les défenseures de leur compétence intellectuelle et politique au moyen de diminutifs méprisants (par exemple : « chamboûqa » [terme à connotation sexuelle assimilant la femme à une actrice de films pornographiques], « chauffeuse de camion »), de l'âgisme et des remarques sur le physique (par exemple : « vieille, va te couvrir », « sâfira »), et à cantonner de force le rôle des femmes à la **subordination aux hommes**.
- **Objectification sexuelle et atteintes à la vertu (Sexual Objectification) :** réduction de la défenseure à son corps et commentaires sexuels grossiers. L'enquête a enregistré **403 violations probables (soit 17,6 % du total des violations relevées)**, consistant en des attaques directes contre la vertu et en l'usage de termes diffamatoires et d'obscénités sexuelles. La participante (P2) confirme ce schéma, expliquant que les assaillants délaissent le fond de son travail pour s'en prendre à « mon apparence, mon corps, qui je suis… Ils attaquent la femme en raison de son sexe et non de son travail ».
- **Intimidation, menaces physiques et atteintes visuelles (Physical Threats & Image-Based Abuse) :** 187 violations probables (8,1 % du total), réparties entre deux schémas fonctionnellement liés :
  - **Intimidation et menaces physiques (133 cas, soit 5,8 %) :** menaces explicites de meurtre et d'atteinte physique (« on va te ramener », « fends-la en deux »), traque des lieux de résidence et « mentions » publiques adressées aux services de sécurité pour faire arrêter les militantes.
  - **Atteintes visuelles et hypertrucages (54 cas, soit 2,4 %) :** toute utilisation des photos personnelles de la personne ciblée sans son consentement dans des contextes diffamatoires, manipulation ou recadrage dégradant des images, ainsi que le recours à l'intelligence artificielle pour générer des images ou des vidéos trompeuses et falsifiées à des fins d'assassinat moral.

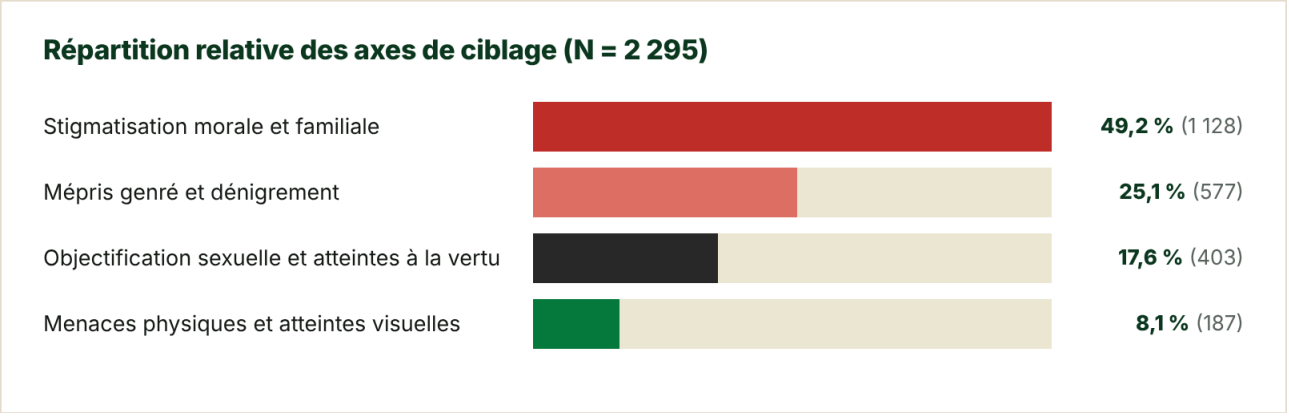


Figure (3) : répartition relative des axes de ciblage genré dans les données archivées

2.2 Intersectionnalité et « permis d'abuser » (License to Abuse)

L'analyse des schémas révèle un mécanisme central qui explique l'intensité de la violence genrée dans le contexte politique syrien complexe : les femmes qui n'appartiennent pas aux groupes dominant le pouvoir, ou qui travaillent sur des dossiers sensibles de justice transitionnelle (personnes disparues, femmes enlevées), deviennent « plus faciles à livrer à la curée ».

Plutôt que de débattre des positions politiques de la militante, les résultats de la veille montrent comment la machine à faire taire applique concrètement sa dynamique d'exclusion : le ciblage genré occupe souvent le devant de la scène comme « coup d'envoi », afin de provoquer un choc psychologique rapide et d'encercler socialement la défenseure. Parallèlement à l'escalade de la campagne de diffamation morale et à l'isolement de la personne ciblée de son entourage solidaire dans les milieux conservateurs, la stigmatisation politique et confessionnelle croise l'attaque genrée — qu'elle coïncide avec elle ou qu'elle la couronne —, non comme une accusation à prouver, mais comme un outil décisif pour dépouiller la défenseure de sa qualité de militante et présenter l'agression comme un « acte justifié ». Les déclarations de la participante (P1) reflètent précisément ce cadrage final : après l'escalade de l'attaque genrée contre elle, on lui a inventé des origines confessionnelles pour couronner sa mise à l'index civique.

> « Pourquoi a-t-il utilisé le mot Nousayriya ? Pour qu'on se mobilise contre moi… Ça fait de moi une cible permise pour n'importe qui. »

Le « permis d'abuser » atteint son apogée lorsque le genre croise explicitement la polarisation confessionnelle. Le suivi des témoignages de journalistes et de militantes issues de milieux confessionnels déterminés et vivant en Syrie (Lattaquié, Hama, Damas), tel que documenté par les entretiens qualitatifs approfondis et les relevés de veille des messages et des menaces directes, montre que l'attaque à leur encontre n'exige pas un engagement politique marquant : elles reçoivent sans cesse des messages menaçants les désignant comme *« orphelines d'Assad »* ou *« sabaya » [femmes réduites en captivité et en esclavage] ou « Nousayriya »*, assortis de menaces explicites. Ce croisement engendre un véritable terrorisme numérique. L'identité confessionnelle supposée de la femme devient ainsi un « accélérateur » de la violence numérique, l'impunité renforçant la facilité avec laquelle ces femmes sont prises pour cibles.



Figure (4) : messages directs reçus par les défenseures, fondés sur une origine confessionnelle supposée, ancrée dans les attaques qu'elles ont subies sans que cette origine soit nécessairement réelle

2.3 Interaction numérique et limites de l'inférence d'amplification

Mesurer l'interaction avec les publications injurieuses peut aider à décrire la façon dont le contenu est reçu et diffusé, mais cela ne suffit pas, à soi seul, à prouver que les algorithmes de la plateforme ont multiplié sa portée ou l'ont favorisé par rapport à d'autres contenus. Les taux d'interaction dépendent notamment de la taille de l'audience du compte, du moment de la publication, du type de contenu, du nombre de partages et de republications, ainsi que de la nature de l'événement traité.

C'est pourquoi, en l'absence de données fiables sur la portée, les impressions ou les recommandations de la plateforme, l'expression « interaction relative avec le contenu injurieux » est employée au lieu d'« indice d'interaction relative ».

2.3.1 Résultats effectifs de la mesure de l'interaction relative

Pour apprécier l'ampleur de l'effet d'interaction du discours de haine genré, le rapport a procédé à une « comparaison descriptive » (Descriptive Comparison) portant sur un échantillon raisonné de 29 publications focales diffusées sur Facebook et Telegram. Ces publications ont été retenues précisément en tant qu'« étincelles déclenchantes » (Trigger Posts) ayant lancé les campagnes de ciblage via les comptes influents sous surveillance. Afin de contrôler la comparaison et de mesurer l'écart réel (les publications déclenchantes bénéficiant par nature d'une interaction exceptionnelle), une « ligne de base » (Baseline) de référence a été définie : 29 publications politiques générales (exemptes de ciblage genré), publiées par les mêmes comptes et pages, sur la même période, et sélectionnées selon un critère d'appariement interne, à raison d'une publication générale pour chaque publication focale. Les données quantitatives d'interaction ont été collectées à l'aide d'outils d'extraction de données.

Résultats descriptifs de la comparaison : la comparaison descriptive montre que les contenus fondés sur les attaques genrées et les atteintes à la vertu ont généré des taux d'interaction et de republication 3,2 fois supérieurs à ceux enregistrés sur les publications politiques de la « ligne de base ». Les 29 publications focales ont totalisé ensemble quelque 25 000 réactions (Reactions) et 9 000 commentaires. La plus forte concentration d'interactions a été relevée sur les publications visant les défenseures par leur nom et leurs photos personnelles : le nombre moyen de commentaires a bondi à 450 par publication focale, contre une moyenne habituelle de la ligne de base ne dépassant pas 140 commentaires par publication politique non genrée (soit une hausse de +221 %).

Cet écart descriptif constitue un indicateur comportemental rendant hautement probable (Highly Likely) la capacité des contenus scandaleux et de l'incitation fondée sur « l'honneur » à accroître la polarisation du public et à faire monter l'interaction ; toutefois, cette conclusion descriptive ne permet pas d'affirmer de manière catégorique l'existence d'un biais algorithmique préalable (Algorithmic Bias) favorisant ce contenu, tant que les plateformes n'offrent pas un accès ouvert aux données de leurs interfaces de programmation (API) permettant d'auditer de façon indépendante les indicateurs des algorithmes de recommandation.

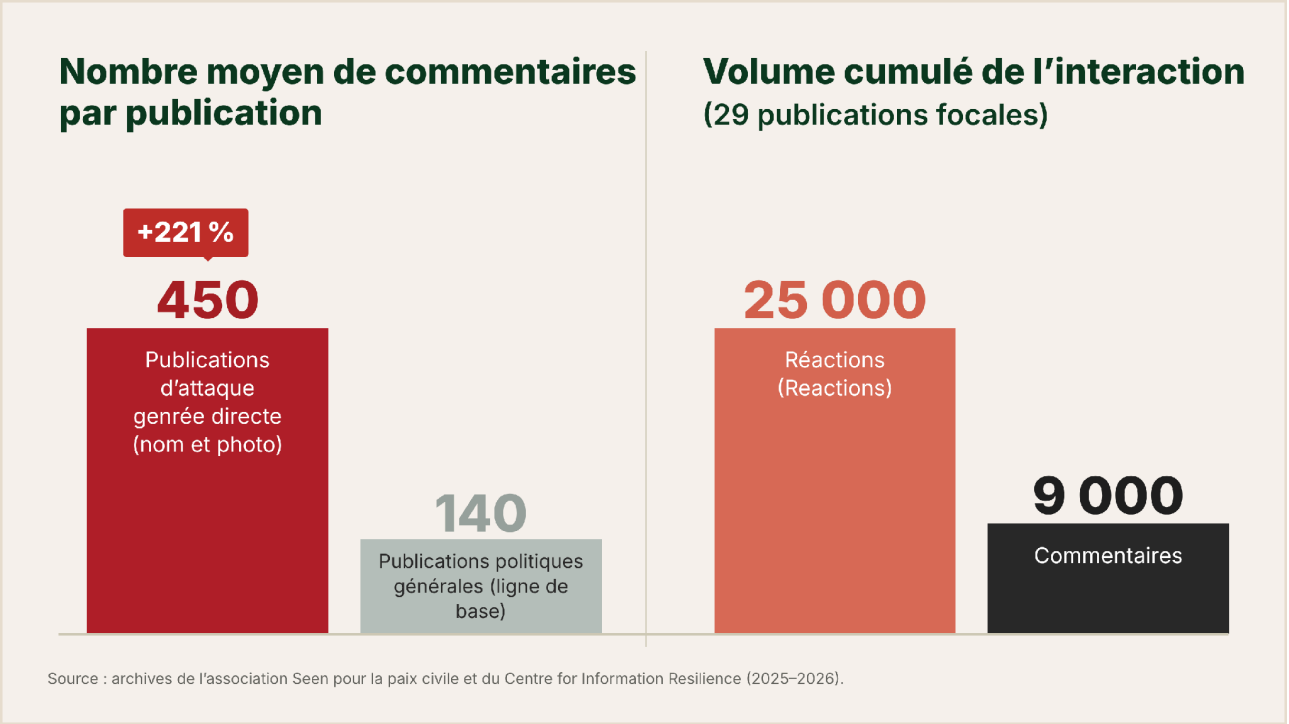


Figure (5) : comparaison descriptive de l'interaction : publications focales et ligne de base

Le graphique présente deux niveaux de mesure : (à gauche) la comparaison du nombre moyen de commentaires par publication entre les publications d'attaque genrée (450 commentaires) et les publications politiques de la ligne de base (140 commentaires), soit une hausse de +221 % ; (à droite) le volume cumulé total de l'interaction sur les 29 publications focales réunies, soit 25 000 réactions (Reactions) et 9 000 commentaires.

Source : archives de l'association Seen pour la paix civile et du Centre for Information Resilience (2025–2026).

03

Chapitre trois

Évolution des attaques et indicateurs techniques : entre mobilisation mécanique et comportement dirigé

L'examen approfondi des incidents de ciblage montre que la violence numérique en Syrie oscille entre deux mécanismes parallèles qui se relaient : d'une part, la « machine sociale », qui réagit de façon mécanique et stéréotypée aux stimuli masculinistes ; d'autre part, la « machine technique », qui mobilise des réseaux orientés pour amplifier l'attaque et contourner les politiques des plateformes. Les finalités de l'interaction désorganisée du public et du comportement dirigé convergent vers un même effet d'exclusion directe, qui étouffe la présence des défenseures et compromet leur sécurité personnelle et professionnelle.

Sur les **29** incidents de ciblage focaux soumis à une analyse en réseau et chronologique, **20** ont été classés dans la « mobilisation mécanique », **7** incidents composites ont atteint le niveau 2 de preuve (comportement dirigé), tandis que **deux incidents (2)** sont restés non concluants faute de métadonnées complètes.

3.1 La mobilisation mécanique : la machine de répression sociale

Les cas examinés par l'enquête indiquent qu'une grande partie de la violence dirigée contre les défenseures ne suppose pas nécessairement l'existence de « salles d'opérations » ou d'une direction centrale, mais repose sur ce que le rapport qualifie de **« mobilisation mécanique » (Mechanical Mobilization)**.

On entend par là la dynamique mécanique par laquelle des utilisateurs ordinaires se précipitent pour participer à une attaque collective sans coordination préalable, en réponse à un système de valeurs masculiniste et à des structures sociales conservatrices. Dès que la première « étincelle de stigmatisation » est lancée, les foules numériques se joignent à l'attaque par une réaction conditionnée visant à contrôler et à punir toute présence féminine qui conteste les récits dominants ou soulève des questions sensibles de redevabilité. La violence numérique devient alors un comportement collectif contagieux qui s'auto-alimente : les commentateurs adhèrent au

discours de haine dans l'illusion d'accomplir un « devoir de protection de l'espace public », ce qui confère à cette violence une force destructrice comparable à celle des attaques organisées.

Le cas de la participante Mme Fadwa Mahmoud (P3), militante engagée sur les questions de justice et de disparitions forcées, offre une illustration concrète du mécanisme de mobilisation mécanique et du parcours du « permis d'abuser » : après qu'elle eut exprimé publiquement des positions critiques sur les processus de redevabilité, les assaillants ont évité de discuter ses arguments, et le « coup d'envoi » a pris la forme d'un ciblage genré intensif centré sur son apparence et son âge, au moyen de termes dévalorisants et de qualificatifs stigmatisants. Très vite, cette focalisation sur la morale et l'apparence a conduit à son isolement progressif et au recul du soutien public de ses soutiens, soucieux d'éviter l'embarras social dans un milieu conservateur. L'une des participantes aux entretiens qualitatifs décrit ce phénomène comme une sorte d'« état de curée généralisée », où les utilisateurs se ruent dans l'attaque par simple imitation du comportement dominant sur la plateforme.

Pour couronner cette curée et justifier la participation collective au harcèlement, la troisième dimension s'est activée : la priver de sa qualité de défenseure et la désigner comme traîtresse en la qualifiant de « résidu du régime » (fouloul). Cette séquence, dans le cas de Mme Fadwa Mahmoud, montre comment la polarisation politique peut fragmenter la solidarité militante et pousser les individus à adhérer spontanément au discours de haine, sans qu'il soit besoin de faux réseaux. La participante (P2) décrit cette dynamique comme une sorte de « fête nationale » du harcèlement, où chacun se joint à l'attaque grégaire simplement parce que « tout le monde harcèle Une telle, alors il faut que j'en sois ».

3.2 Le comportement dirigé : orienter et amplifier l'attaque

À l'inverse, toutes les formes de violence numérique ne peuvent s'expliquer par la seule mobilisation sociale spontanée : les données archivées révèlent des indicateurs techniques et comportementaux concordants qui atteignent le niveau 2 de preuve (**preuve de la coordination et du comportement**), sans préjuger de l'existence d'une attribution sécuritaire ou politique directe.

Le suivi des mouvements des comptes confirme l'existence d'un schéma organisé de raids / **d'attaques dirigées (Brigading)** : l'appariement numérique entre les publications des influenceurs (M1, M2 et M3) et celles des médias ayant reçu les défenseures a permis de repérer **244 comptes récurrents** se déplaçant en parallèle entre plusieurs foyers, dont **59 comptes actifs** au sein de l'échantillon d'un même influenceur, passés méthodiquement d'une publication à l'autre. L'activité de ces comptes ne se limitait pas à des commentaires ponctuels : ils agissaient comme comptes d'amplification de la campagne, se déplaçant dès la publication de tout contenu pour traquer la défenseure et inonder son espace des mêmes modèles de

stigmatisation morale, d'objectification sexuelle et d'association forcée à des hashtags fabriqués (comme le hashtag « Sabaya al-Ataa »[4] et les accusations de trahison), jusqu'à la poursuivre sur les médias et à commenter les extraits de ses entretiens, dans un comportement de raid dirigé (Brigading) qui dépasse les limites de l'interaction mécanique.

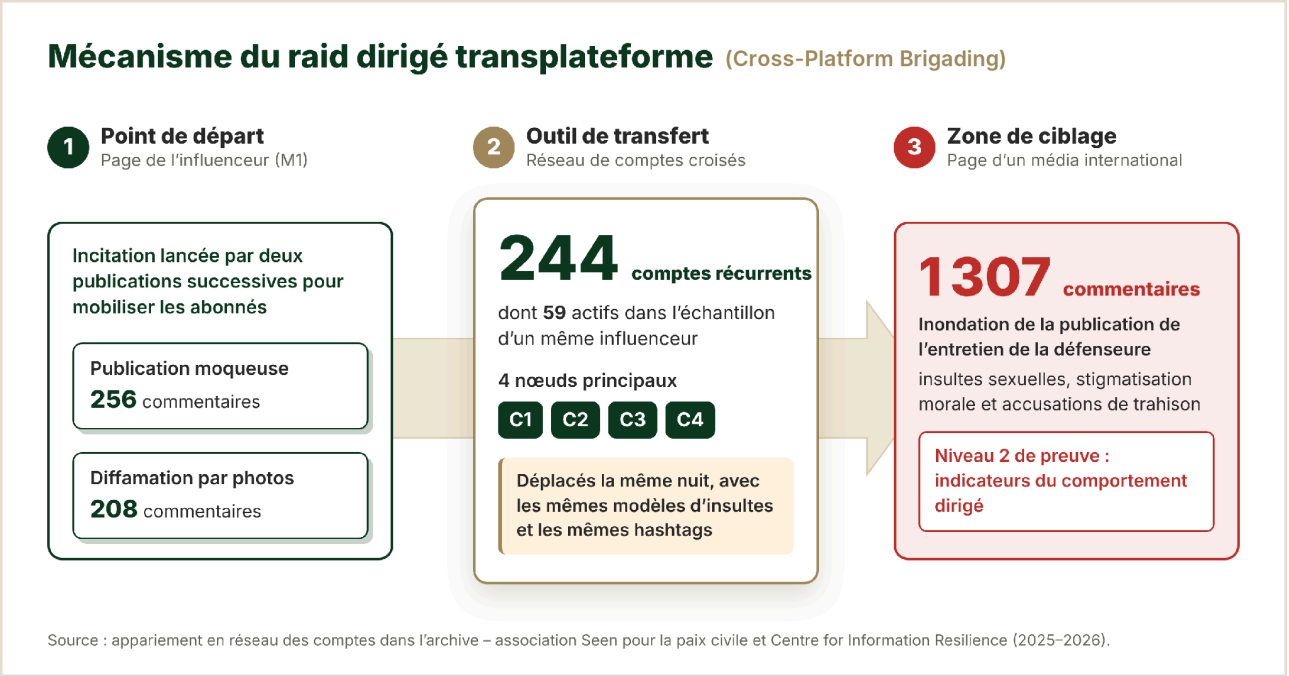


Figure (6) : parcours du raid transplateforme (Cross-Platform Brigading)

Le schéma illustre le mécanisme de coordination comportementale entre la publication d'incitation centrale diffusée sur la page de l'influenceur et la publication d'un média international : l'enquête a documenté le déplacement simultané, la même nuit, de dizaines de comptes identiques pour traquer la défenseure et inonder son espace médiatique des mêmes accusations de trahison et attaques générées, afin de briser sa crédibilité de militante.

Ce schéma en réseau correspond à ce que l'enquête a observé lors du suivi d'un cas concret : le ciblage d'une militante après ses réponses critiques à une publication incitative. Un réseau de comptes techniquement imbriqués y a mené une campagne systématique fondée sur le passage simultané de la page de l'influenceur à la page personnelle de la défenseure. L'analyse des flux chronologiques montre le déplacement de deux comptes identifiés (A1 et A2) dans des intervalles de temps resserrés (1h34 pour le premier et 12h19 pour le second) pour envahir sa page personnelle et inonder 6 de ses publications d'accusations de collaboration et d'atteintes à son honneur.

[4] Sabaya al-Ataa » (littéralement « les jeunes femmes du don ») désigne, à l'origine, des groupes féminins de soutien rattachés aux Forces de défense nationale et à d'autres formations loyales à l'ancien régime, mobilisés à des fins de propagande et d'appui logistique et médiatique aux forces gouvernementales. Le nom est ainsi resté associé à l'appareil du pouvoir. Après la chute du régime, il a été détourné en hashtag diffamatoire : jouant sur la quasi-homophonie entre sabaya (« jeunes femmes ») et sabaya (« captives »), il sert à prêter aux militantes visées un rôle sexuel au service de l'ancien régime. Le hashtag cumule donc en un seul mot la stigmatisation politique et la stigmatisation sexuelle, illustrant le mécanisme du « permis d'abuser » décrit dans ce rapport.

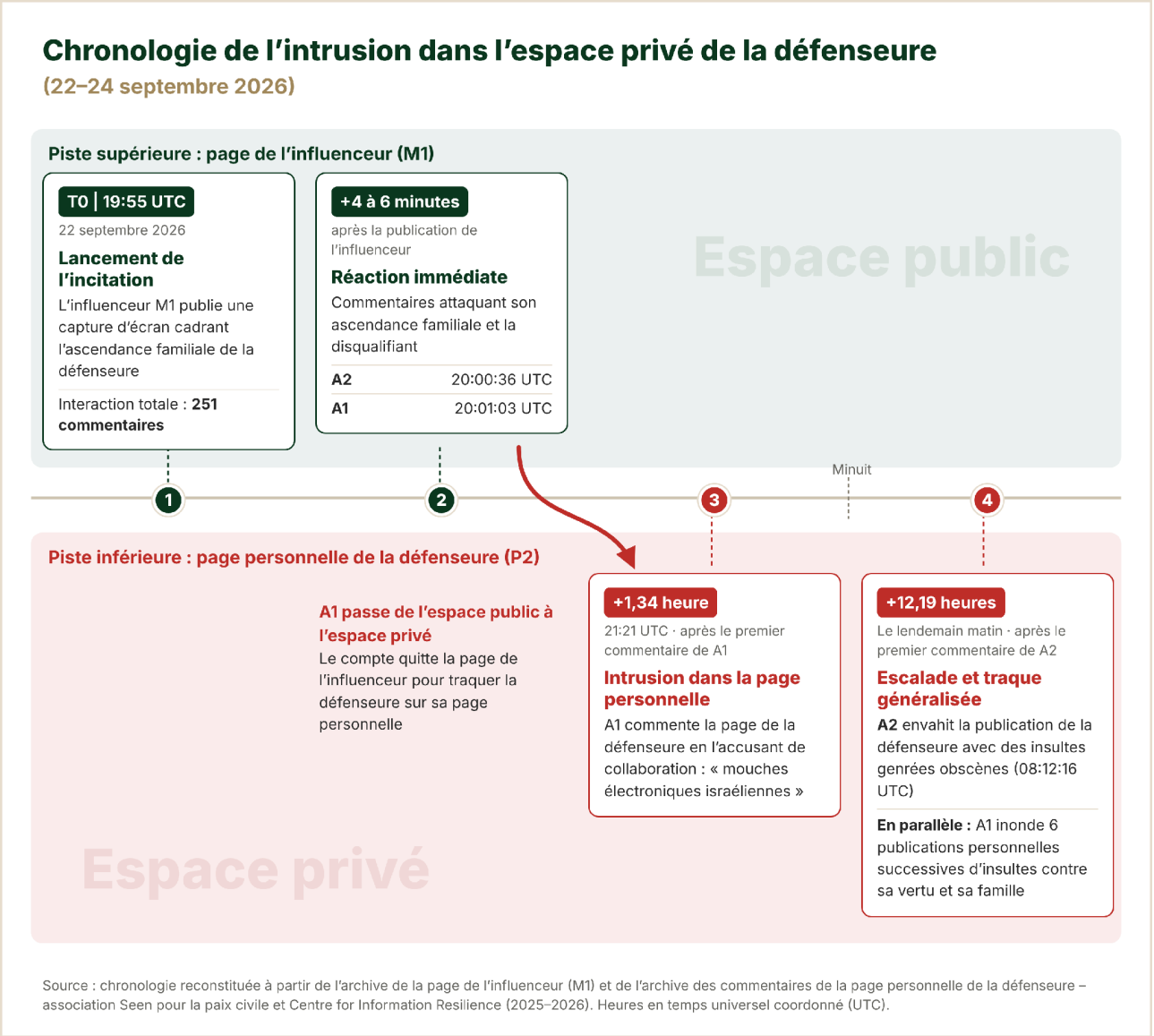


Figure (7) : suivi séquentiel et chronologique du mouvement des comptes et de l'intrusion dans l'espace privé de la défenseure

Le schéma retrace le parcours organisé des comptes d'incitation à partir de la publication de l'influenceur (M1), mise en ligne à 19:55 UTC. L'analyse chronologique documente un premier flux de commentaires 4 minutes seulement après, suivi du déplacement dirigé du compte (A1) 1,34 heure plus tard pour envahir la page personnelle de la défenseure et l'accabler d'insultes portant atteinte à sa vertu et l'accusant de collaboration. Le comportement dirigé a atteint son paroxysme 12,19 heures plus tard avec l'intervention du compte (A2), qui a traqué simultanément la militante sur 6 publications privées différentes en les inondant d'accusations de trahison et du hashtag « Sabaya al-Ataa », dans un comportement de raid dirigé.

3.2.1 Critères de documentation du comportement dirigé

- **Classification de la nature des comptes :** distinction stricte entre pages publiques, comptes d'influenceurs et de créateurs de contenu, et comptes dont certains indices laissent penser qu'ils sont fabriqués ou d'emprunt (Sockpuppets).

- **Repérage des indicateurs temporels et rythmiques (Spiking & Timing) :** mesure des afflux soudains de commentaires à contenu genré injurieux dans des fenêtres de temps resserrées, attestant d'une injection programmée et dirigée de contenu.
- **Identification des rôles dans le réseau (Network Mapping) :** distinction entre comptes d'incitation initiale (Seed Accounts), comptes de relais et d'amplification (Amplifiers) et comptes de raid sur le terrain.
- **Schémas linguistiques uniformes :** suivi de la concordance textuelle et de la copie mot pour mot de formules et de hashtags préfabriqués pour cadrer la personne ciblée.
- **Garde-fous éthiques et juridiques :** documentation précise des entités et des personnalités publiques influentes, avec anonymisation des comptes ordinaires afin de protéger la vie privée.

3.2.2 Le cas de la participante (P1)

Cette forme de ciblage est confirmée par le cas de la participante (P1), défenseure des droits humains originaire du nord-est de la Syrie, exemple de « curée composite dirigée ». Après s'être montrée solidaire des femmes enlevées et avoir participé à des activités civiques, elle n'a pas seulement subi des insultes aléatoires : elle a été visée par des outils techniques avancés et systématiques. L'enquête a documenté dans son cas les indicateurs de coordination suivants :

- **Hypertrucage (Deepfakes) :** diffusion d'une vidéo falsifiée montrant la défenseure dans une situation intime avec un inconnu, dans le but de détruire rapidement sa réputation morale dans un milieu conservateur.
- **Usurpation d'identité et amplification :** recours à des pages d'information gérées sous des identités fictives (par exemple une page tenue par un homme sous un nom féminin) et disposant d'une large audience dans la ville de la personne ciblée, pour diffuser simultanément son nom et sa photo, selon une tactique visant clairement à « désigner la victime » (Targeting) au public en vue de la curée.
- **Hashtags dirigés :** inondation systématique de ses espaces par des hashtags préfabriqués construisant des identités inventées (par exemple en prétendant qu'elle est « Nousayriya » pour attiser contre elle les instincts confessionnels dans son milieu, ou en la qualifiant de manière répétée de « Qasadiya » et de « protectrice des minorités »).

3.3 La tactique de « l'armée du Haha » (Haha Army) et la tentative de manipuler les algorithmes

L'interaction dirigée au moyen de faux comptes constitue l'une des formes de comportement dirigé les plus récentes et les plus efficaces dans le ciblage des défenseures syriennes : le

matériau d'enquête relève un schéma récurrent consistant à inonder des publications militantes sérieuses — portant sur les détenus, les femmes enlevées, la justice transitionnelle et la documentation des violations — de centaines, voire de milliers de réactions moqueuses organisées (réaction « Haha » / Haha Reaction).

L'examen des métadonnées et des schémas géographiques de ces comptes révèle un paradoxe contextuel flagrant : les analyses visuelles ont établi que les comptes ayant réagi aux publications se concentraient dans des zones géographiques totalement éloignées du contexte syrien, notamment la Zambie (Lusaka), Madagascar (Antananarivo) et le Népal (district de Jhapa), ainsi que l'Indonésie, la Thaïlande, le Myanmar et le Sri Lanka. Ces disparités géographiques indiquent **avec une forte probabilité (High Probability)** l'absence de toute motivation authentique derrière cet afflux simultané, ce qui en suggère le caractère artificiel.

Infrastructure technique et opérationnelle des campagnes : ces indicateurs concordent avec des schémas documentés à l'échelle internationale d'utilisation d'infrastructures d'interaction inauthentique, parmi lesquelles :

- **Fermes à clics et votes automatisés (Click Farms) :** centres techniques regroupant des milliers de téléphones équipés de cartes SIM bon marché, ou pilotés via des serveurs mandataires (Residential Proxies) simulant des adresses résidentielles dans différents pays afin de contourner les systèmes de détection de sécurité de Meta.
- **Plateformes de services de marketing numérique (SMM Panels) :** marchés intermédiaires non déclarés permettant aux instigateurs d'acheter des lots d'interactions précis (par exemple : un lot de 1 000 réactions « Haha » pour un prix compris entre 2 et 5 dollars seulement),[5] avec la possibilité de programmer précisément le moment de l'injection et de désigner le lien ciblé. Bien que l'équipe de documentation n'ait pu accéder à aucune preuve financière décisive établissant que des bailleurs se tiennent derrière cette activité, la concordance comportementale manifeste de ces attaques avec le fonctionnement de ces plateformes rend hautement probable l'hypothèse d'un recours délibéré à ces services rémunérés.
- **Comptes d'emprunt et comptes abandonnés (Zombie Accounts) :** comptes créés à l'origine par de véritables utilisateurs dans des pays du Sud global, puis piratés au moyen de logiciels d'hameçonnage, ou abandonnés pendant des années avant d'être vendus dans des lots de données au marché noir à des réseaux d'automatisation et réactivés de force.

L'objectif de cette tactique systématique ne se limite pas à la provocation psychologique de la défenseure : elle remplit un double rôle d'ingénierie visant à détourner les algorithmes et à façonner la perception du public local, à travers deux mécanismes complémentaires :

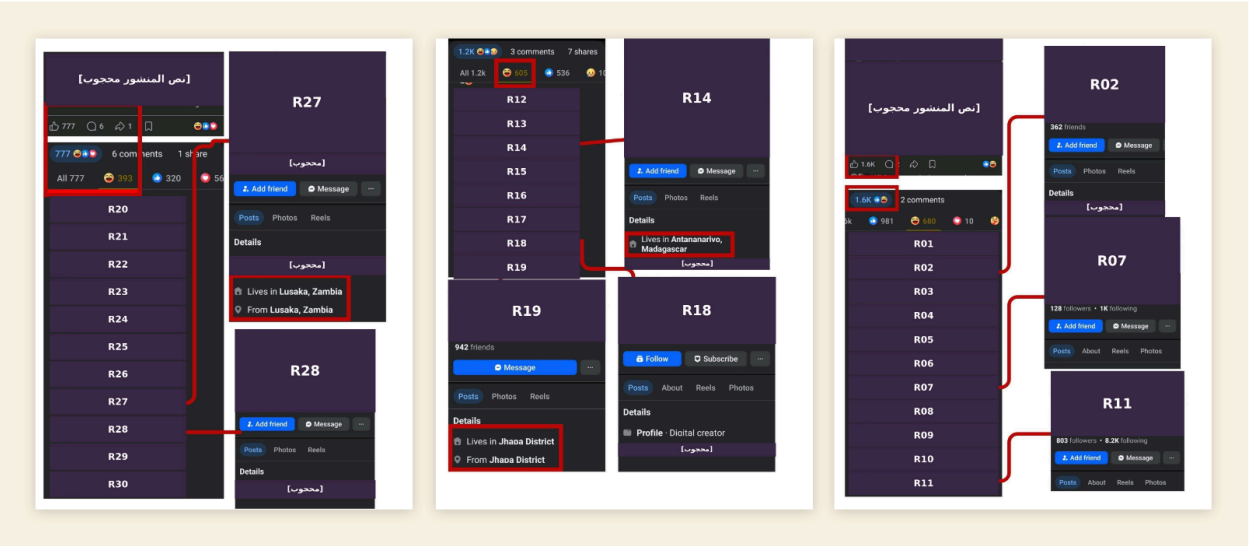
[5] Rapports du DFRLab (Digital Forensic Research Lab) et de l'Institute for Strategic Dialogue (ISD) sur l'économie des plateformes de marketing numérique (SMM Panels) et des fermes à clics

- **Redirection algorithmique (Algorithmic Redirection) :** les algorithmes de Facebook interprètent l'afflux massif de réactions « Haha » comme la preuve que le contenu suscite une « interaction polémique exceptionnelle ». Au lieu de freiner l'attaque, l'algorithme amplifie la diffusion de la publication et l'arrache à son « public organique ciblé » (familles de personnes disparues, acteurs du champ des droits humains) pour la pousser vers les fils d'un « public non ciblé » (Unintended Audience). Ce déplacement coupe la défenseure de son milieu de soutien et expose son discours à des publics généraux dépourvus de contexte et plus susceptibles d'être polarisés, ce qui multiplie le flux hostile aléatoire dirigé contre elle.
- **Cadrage préalable :** ce schéma prive le contenu militant de son poids moral et humain : lorsqu'un utilisateur ordinaire tombe sur une publication documentant des violations graves entourée de centaines de rires artificiels, il subit inconsciemment l'effet de la « preuve sociale et du conformisme » (Social Proof), et la dérision envers la victime et le déni de sa souffrance apparaissent comme une position collective normale. Ce cadrage psychologique préalable ouvre la voie à l'entrée de la rue numérique dans l'attaque selon la dynamique de la « mobilisation mécanique », sans que les participants comprennent qu'ils suivent un faux consensus fabriqué et alimenté par des outils rémunérés.

Dans le contexte syrien, cette pratique n'est pas dissociable de stratégies mondiales documentées dans les conflits, qui utilisent la violence numérique genrée transfrontalière comme outil central de répression. Elle rappelle des schémas documentés par des rapports internationaux, notamment les campagnes visant la journaliste d'investigation Maria Ressa aux Philippines[6], où des réseaux d'interactions moqueuses ont servi à submerger ses enquêtes et à tromper les algorithmes de recommandation ; elle rappelle aussi les tactiques observées au Brésil[7], où des fermes à clics asiatiques ont été louées pour mener des campagnes systématiques de dérision et d'accusations morales contre des défenseures et des femmes politiques indépendantes afin de briser la solidarité populaire à leur égard, ou encore les méthodes des armées électroniques au Moyen-Orient, qui recourent aux raids rapides (Brigading) pour contourner les politiques des plateformes.

Ces données rendent hautement probable l'hypothèse d'une coordination préalable et du recours à des outils d'influence inauthentique pour amplifier le ciblage et imposer le silence et l'autocensure aux défenseures, sans que les données dont dispose cette enquête permettent d'identifier l'acheteur, le mécanisme d'achat ou l'infrastructure utilisée dans les cas étudiés.

[6] Étude de l'UNESCO sur Maria Ressa
[7] Rapports de l'Union interparlementaire et d'Amnesty International sur les fermes à clics au Brésil



Figures (8), (9) et (10) : captures d'écran des comptes ayant réagi par « Haha » et de leurs localisations déclarées

Les prénoms des comptes ayant réagi aux publications et les textes des publications originales ont été masqués et remplacés par des codes (R01–R30) ; les compteurs d'interaction et les localisations déclarées ont été conservés à titre de preuve.

04

Chapitre quatre

Effets et conséquences du ciblage

4.1 L'essor de « l'autoritarisme électronique » et le déplacement de la menace

Les entretiens et les témoignages qualitatifs des défenseures des droits humains (P1 à P6) révèlent une transformation notable de la structure des risques qui pèsent sur l'espace public syrien : avec l'effritement de l'emprise sécuritaire centrale qui imposait autrefois l'autocensure, les sources de menace se sont largement fragmentées au profit de l'emprise de la rue numérique et de l'influence des créateurs de contenu. Cette évolution a coïncidé avec l'essor d'influenceurs disposant de vastes communautés d'abonnés, dont les discours acquièrent un poids de cadrage qui attise l'hostilité collective et oriente les vagues de diffamation, même en l'absence de preuves d'un mandat préalable pour leurs campagnes.

Les témoignages des défenseures révèlent la nature de cette inquiétude composite : la plus grande crainte est désormais l'impuissance à protéger sa famille restée en Syrie contre les retombées de l'incitation numérique transfrontalière, ce qui a engendré un sentiment d'absence totale de protection et poussé des militantes à fermer leurs comptes, contraintes, pendant de longues périodes, et à se retirer complètement du débat public :

> **« Je me suis tue, j'ai eu peur que ma famille soit tuée. La menace directe, c'était la mort si je rentrais au pays, ou qu'on découvre où vit ma famille, et c'est allé jusqu'aux mentions au ministère de l'Intérieur : "Elle est où, celle-là ? Ramenez-la !" »**
>
> **« Il n'y a absolument aucune instance vers qui se tourner… Aucune loi ne protège, il ne reste que la fuite, s'éloigner, se tenir à l'écart de la loi de la jungle qui règne en ce moment sur les réseaux sociaux… La loi elle-même est utilisée selon les intérêts, et il y a une loi qui protège beaucoup d'instigateurs, que les autorités orientent en toute légitimité vers l'attaque contre les militantes. »**

> ***— Entretien avec la participante (P6), militante indépendante des droits humains et militante politique.***

> **« Ça devient comme une “fête nationale” du harcèlement ; les assaillants laissent de côté le fond du débat et du travail militant et s’en prennent à mon apparence, à mon corps et aux détails de ma vie privée. Les gens participent juste parce que tout le monde harcèle, alors “il faut que j’en sois”. »**

> ***— Entretien avec la participante (P2), journaliste d’investigation et cofondatrice d’un média.***

Ces témoignages concordants des défenseures confirment un changement profond dans la nature de l’autocensure : alors qu’elle était autrefois dictée par la crainte de la répression de l’appareil sécuritaire, elle est aujourd’hui motivée par le souci d’éviter l’incitation ouverte menée par des influenceurs disposant de larges plateformes. Si certains influenceurs ont participé à des réunions et à des événements publics avec des instances officielles et des autorités locales après décembre 2024, la preuve d’une couverture ou d’une orientation institutionnelle directe de leurs campagnes reste soumise aux niveaux de preuve stricts : l’enquête se limite à établir l’existence d’un intérêt politique et social mutuel, sans conclure à un lien de subordination organisationnelle en l’absence de preuves techniques et judiciaires décisives.

Le suivi de terrain montre également le croisement de la violence numérique avec le harcèlement et la traque dans le monde physique : lors d’un incident documenté à Damas, à la suite d’un rassemblement civique, une défenseure a fait l’objet d’un harcèlement physique filmé dans la rue. Dès que la vidéo a été publiée et relayée par des pages publiques, l’incident s’est transformé en une campagne de diffamation féroce : selon la veille du département Documentation et études de l’association Seen, sa page recevait, dès qu’elle publiait la moindre clarification, près de 100 commentaires injurieux et insultes à caractère sexuel par minute au plus fort de la première vague, ce qui l’a finalement contrainte à désactiver son compte pour protéger sa sécurité psychologique et physique.

Le résultat final de ces dynamiques est la production d’un **« effet dissuasif collectif » (The Chilling Effect)**[8] : le message d’intimidation ne s’arrête pas à la mise sous silence de la défenseure visée, mais parvient à intimider les voix féminines émergentes et à les pousser vers l’autocensure et le retrait de la vie publique.

[8] UNESCO, The Chilling: Global Trends in Online Violence Against Women Journalists, document de recherche.

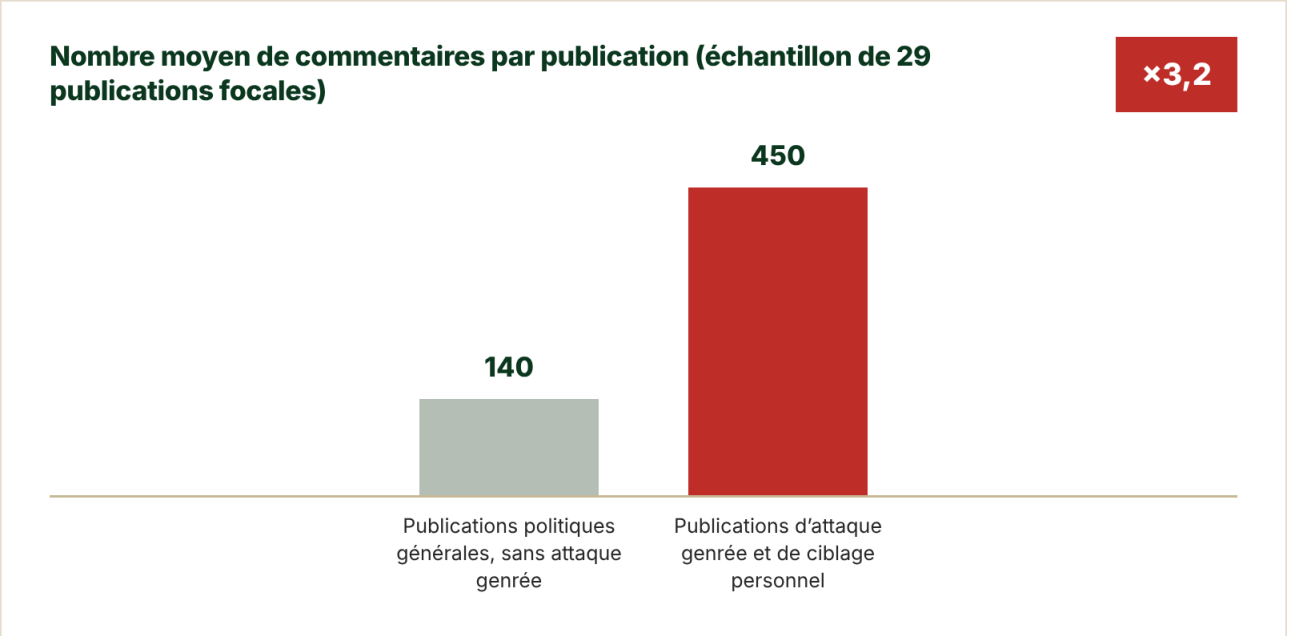


Figure (11) : multiplicateur de l'interaction relative des publications d'attaque genrée par rapport aux publications générales

Le graphique présente la comparaison statistique du volume d'interaction sur l'échantillon des 29 publications focales examinées : le nombre moyen de commentaires passe de 140 pour les publications politiques générales à 450 pour les publications comportant un ciblage genré et des attaques personnelles (soit 3,2 fois plus, une hausse de +221 %), ce que les acteurs exploitent pour provoquer un choc psychologique rapide imposant aux personnes ciblées un retrait contraint.

4.2 Au-delà d'Internet : convergence de la violence et exclusion physique

La violence numérique contre les défenseures syriennes n'est pas un simple « préjudice virtuel » : des cas documentés dans ce rapport montrent qu'elle peut précéder une violence physique ou l'accompagner. Le rapport documente un phénomène de « convergence » (Convergence), où l'incitation numérique rejoint les mesures punitives matérielles et les pressions sociales :

- **Mesures judiciaires et administratives :** le cas de la défenseure de premier plan (H1), directrice d'une association féministe (O1), constitue l'exemple le plus grave de cette convergence. En avril 2025, à la suite d'une féroce campagne de diffamation en ligne fondée sur un enregistrement audio tronqué et d'origine inconnue, la campagne s'est accompagnée de menaces d'incendier le bureau de l'association. En quelques jours, l'incitation numérique s'est muée en mesure officielle : les autorités locales ont déposé une plainte contre elle, la police a fermé l'unique bureau de l'association dans le nord-ouest de la Syrie, et sa famille a été poursuivie. Cette chronologie montre une concomitance frappante entre la campagne numérique et les mesures officielles, ce qui laisse penser que la campagne a servi de couverture à ces mesures.

- **Instrumentalisation de la pression familiale (censure par procuration) :** les assaillants ciblent les familles des défenseures comme levier pour les faire taire. Dans le cas de (P1), l'attaque s'est étendue à sa famille par des formules incitatives telles que « il n'y a pas d'homme à la maison ». Selon le témoignage de la participante (P6), ses proches vivant en Syrie ont même été menacés de mort et d'arrestation pour la contraindre au silence. Cette stigmatisation pousse les familles — par peur ou pour éviter le scandale dans les milieux conservateurs — à exercer des pressions coercitives sur les militantes afin qu'elles démissionnent ou mettent fin à leur présence publique.
- Menaces transnationales : les incidents documentés montrent que le fait de vivre hors du territoire syrien, dans les pays d'asile, ne met pas les défenseures à l'abri des menaces physiques. La participante Mme Fadwa Mahmoud (P3) a reçu, sur ses comptes privés, des messages de menace directs indiquant précisément la rue et le lieu de sa résidence dans une ville européenne, accompagnés de la phrase : « Tu es à (Berlin) et nous pouvons t'atteindre ». Ce schéma confirme le passage du terrorisme numérique des querelles verbales à des pratiques de répression transnationale visant à instiller une peur existentielle de l'élimination physique et à restreindre la liberté de mouvement et le travail militant.

4.3 Effets du ciblage sur l'activité numérique des défenseures

La comparaison de l'activité numérique publique de 12 défenseures ciblées — en mesurant le rythme hebdomadaire moyen de publication sur trois fenêtres temporelles : la période de référence (deux semaines avant le début de la campagne), la période de pic du ciblage et la période post-campagne (deux semaines après le reflux de la vague d'attaques) — montre que le rythme moyen de publication est passé de 14 publications par semaine avant la campagne à 3 pendant celle-ci, avant de remonter partiellement à 5 après. 4 défenseures ont temporairement fermé leurs comptes, 8 ont restreint leurs paramètres de confidentialité et de commentaires, et la part des contenus militants consacrés aux affaires publiques est passée de 78 % à 24 %. Les témoignages qualitatifs étayent le lien causal direct entre ces changements et les campagnes dans les cas étudiés, tandis que les autres changements sont présentés comme une corrélation temporelle.

05

Chapitre cinq

Lacunes de protection et défis

Si le système de ciblage numérique se nourrit de l'interaction sociale et technique pour se propager, sa persistance et son efficacité démesurée sont consacrées par l'environnement juridique et réglementaire qui l'abrite : l'analyse juridique appliquée des incidents documentés révèle un environnement propice à l'impunité, alimenté par un double déficit systémique. D'une part, le système législatif national souffre d'une application sélective, que les cas documentés dans ce chapitre mettent en évidence, et qui transforme la loi elle-même, protection présumée, en instrument d'intimidation et de dissuasion. D'autre part, les grandes plateformes numériques se dérobent à leurs obligations de diligence raisonnable et ignorent la spécificité du contexte syrien, laissant les défenseures face aux campagnes de mobilisation sans aucun garde-fou technique préventif.

5.1 Matrice des défaillances de protection (Applied Protection Matrix)

Plutôt que de passer les textes en revue de manière théorique, cette enquête pose une question pratique : quelle protection chaque défenseure aurait-elle dû recevoir dans les cas documentés, et où la défaillance s'est-elle produite ?

Cas documenté	Acte numérique observé	Protection juridique / réglementaire présumée	Point de défaillance effectif (lacune)
Participante (P1)	Falsification visuelle et production d'une vidéo truquée par intelligence artificielle, gestion d'une page sous identité usurpée, et incitation confessionnelle systématique.	**Au niveau national :** article 13(2) de la Déclaration constitutionnelle et dispositions relatives à l'usurpation d'identité et à la diffamation électronique dans la loi sur la cybercriminalité ; **Plateformes :** politiques d'interdiction et de retrait des hypertrucages non consentis (Non-Consensual Deepfakes).	**Au niveau national :** absence de mécanismes de plainte sûrs pour les femmes résidant hors des zones de contrôle direct, et défiance envers la justice ; **Plateformes :** lenteur des procédures de réponse, et simple désactivation de la page sans empêcher les instigateurs d'en créer immédiatement de nouvelles.

Cas documenté	Acte numérique observé	Protection juridique / réglementaire présumée	Point de défaillance effectif (lacune)
Défenseure (H1)	Campagne de diffamation composite, appels publics à incendier les locaux de l'organisation, et menaces d'élimination.	**Au niveau international :** article 20(2) du Pacte international relatif aux droits civils et politiques (interdiction de l'incitation à la violence) ; **Au niveau national :** lois protégeant les organisations de la société civile et le droit d'association.	**Convergence de la violence :** utilisation par les autorités locales de la campagne numérique comme couverture justificative et prétexte sécuritaire pour fermer l'unique bureau de l'organisation dans le nord-ouest de la Syrie et poursuivre la défenseure et sa famille, en adéquation directe avec les objectifs des auteurs.
Participante Mme Fadwa Mahmoud (P3)	Menace directe d'atteinte physique précisant l'adresse de résidence dans une ville européenne, accusations de trahison politique et de collaboration.	**Au niveau international :** lois pénales de l'État de résidence et règlement européen sur les services numériques (DSA), contraignant pour les plateformes au sein de l'Union européenne.	Charge de la documentation pénale reposant entièrement sur une victime psychologiquement épuisée, et absence d'un signaleur de confiance assurant le suivi des violations transnationales.
Participante (P2) / ciblage institutionnel	Ciblage d'un média féministe pour l'empêcher de révéler des affaires de violences sexuelles et de terrain.	**Au niveau national :** article 13(1) de la Déclaration constitutionnelle (garantie de la liberté de la presse, de publication et du travail médiatique).	Qualification judiciaire et sociale de la production de connaissances sur les crimes sexuels comme « atteinte aux bonnes mœurs », privant le journalisme d'investigation de toute protection professionnelle.
Participante (P6)	Menaces d'élimination physique, traque de ses proches en Syrie pour faire pression sur elle, et signalements publics aux forces de l'ordre en vue de son arrestation.	**Au niveau national :** dispositions incriminant la menace, le chantage moral et l'usage abusif du réseau ; **Plateformes :** protocoles de protection des défenseures contre la coordination transnationale et la traque familiale (Doxxing).	La plainte en justice devient elle-même une menace qui aggrave le risque d'arrestation et de représailles contre la famille restée en Syrie, tandis que les algorithmes des plateformes sont incapables de relier le ciblage de la défenseure aux menaces visant ses proches, ce qui l'a contrainte à l'autocensure et à suspendre son activité pendant de longues périodes.

5.2 La législation comme outil de répression : application sélective et absence de perspective de genre

L'examen de l'architecture législative syrienne régissant l'espace numérique — au premier rang de laquelle la loi n° 20 de 2022 sur la lutte contre la cybercriminalité — révèle une impasse structurelle qui s'aggrave pendant la phase de transition ouverte par le basculement des rapports de force en décembre 2024 : cette législation demeure en vigueur dans la pratique et conserve ses dispositions vagues, fondamentalement contraires aux principes de la Déclaration

constitutionnelle et aux normes internationales en matière de liberté d'expression et de justice de genre.

Bien que le ministère de la Justice ait annoncé en juin 2026 son intention de réviser la loi et de reformuler ses articles problématiques[9], et qu'il ait ensuite publié la circulaire n° 26[10] encadrant les avis de recherche et réservant la détention à des décisions du parquet fondées sur des preuves matérielles, ces mesures sont restées des correctifs procéduraux limités qui n'ont pas touché au fond de l'incrimination, et ne se sont pas accompagnées d'une suspension de l'application de la loi, comme le réclamaient les organisations indépendantes de défense des droits humains. En conséquence, l'application concrète est demeurée marquée par une forte dualité, qui se manifeste selon deux axes principaux :

- **Aveuglement structurel au genre (Gender Blindness) :** la loi ne contient aucune qualification ni reconnaissance juridique de la **« violence numérique fondée sur le genre et facilitée par la technologie » (TFGBV)**. Les tribunaux et les services de lutte contre la cybercriminalité traitent les affaires de ciblage des défenseures au moyen de catégories pénales traditionnelles inadaptées (comme les articles 23 et 24 relatifs à la diffamation et à l'injure simples, ou les articles 21 et 22 relatifs à l'atteinte à la vie privée). Cette qualification traditionnelle ignore :
  - le caractère collectif et réticulaire des campagnes (qu'il s'agisse de la mobilisation mécanique ou du comportement dirigé et des tactiques d'inondation par les interactions) ;
  - les menaces non consenties issues de l'intelligence artificielle et de l'hypertrucage (Deepfakes), conçues spécifiquement pour l'assassinat moral ;
  - l'absence totale de mécanismes de protection et de mesures conservatoires urgentes, telles que des ordonnances de protection judiciaire d'urgence, l'obligation pour les fournisseurs de services de procéder à un retrait immédiat, ou des voies de plainte confidentielles tenant compte de la sensibilité des questions d'honneur et de stigmatisation familiale dans les milieux locaux.
- **La persistance de l'application sélective (Selective Enforcement) comme arme de dissuasion :** malgré l'encadrement des procédures de détention provisoire, la réalité du terrain montre que les dispositions sur la diffamation et les qualifications extensibles sont mobilisées avec une souplesse et une rapidité exceptionnelles dès lors que les critiques visent des personnalités influentes de la nouvelle administration ou des acteurs locaux

[9] Déclaration du ministère syrien de la Justice, juin 2026 : réexamen de la loi sur la cybercriminalité (loi n° 20 de 2022).
[10] Circulaire judiciaire n° 26, publiée par le ministère de la Justice de la République arabe syrienne le 9 juin 2026, visant à encadrer les procédures applicables aux affaires de cybercriminalité et à préserver les droits et libertés des citoyens.

armés. À l'inverse, lorsque des pages et des réseaux très influents mènent contre les défenseures des campagnes explicites et documentées de diabolisation, de diffamation et d'accusations de trahison (atteignant le seuil de l'incitation à la violence physique selon les critères du « Plan d'action de Rabat » des Nations unies), les forces de l'ordre s'abstiennent d'intervenir d'office, invoquant « l'absence de plainte personnelle directe » ou « la difficulté de remonter aux comptes d'emprunt ».

Ce contraste transforme la loi et ses organes d'exécution, d'instruments de réparation, en moyens de reproduire l'intimidation et de consacrer l'impunité ; il explique la réticence des défenseures et des survivantes interrogées (comme les participantes P1 et P6) à saisir la justice : dans de telles conditions, porter plainte constitue un risque sécuritaire qui expose davantage leur sphère privée et met leurs familles en danger, dans un environnement judiciaire dépourvu de garanties d'indépendance.

5.3 La dérobade des plateformes : l'absence de « diligence raisonnable » dans le contexte syrien

Les entreprises de réseaux sociaux portent une responsabilité propre et directe en matière de droits humains au titre des « Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme ». La confrontation des politiques des plateformes avec les incidents relevés dans l'archive numérique met en évidence un manquement manifeste des plateformes à leurs obligations de diligence raisonnable dans le contexte syrien :

- **Le recul des politiques de Meta et le démantèlement de la protection proactive :** l'assouplissement par Meta, début [11]2025, de ses politiques de modération du discours de haine et sa tolérance envers les qualificatifs dégradants dans les espaces de débat politique ont fragilisé la sécurité numérique des femmes : la fin de la « modération proactive » s'est traduite en pratique par le transfert intégral de la charge du signalement et de la preuve sur la victime. À cela s'ajoute la faible capacité des algorithmes et des systèmes de contrôle automatisé à saisir la charge des dialectes syriens locaux et du langage codé sur le plan confessionnel et genré (métaphores voilées de l'honneur, sobriquets dévalorisants), si bien que les algorithmes sont incapables d'intervenir et contribuent à amplifier les contenus injurieux en fonction de l'intensité des interactions.
- **Refuge sûr et salles d'opérations sur Telegram :** malgré les annonces répétées de la plateforme sur le développement d'outils de modération, les chaînes et les fils de Telegram restent des « refuges sûrs » et ouverts pour la circulation de contenus falsifiés, la coordination des campagnes de diffamation et la désignation des cibles, avant que les

[11] Mises à jour de la politique de Meta sur les conduites haineuses, janvier 2025

contenus ne soient exportés et injectés sur des plateformes plus répandues comme Facebook. Ces campagnes tirent parti de la structure unidirectionnelle des chaînes et des groupes fermés, qui prive la victime de toute possibilité de répondre ou de réfuter les allégations.

La conjonction des carences législatives et de l'application sélective au niveau local avec la dérobade technique des plateformes internationales aboutit à une structure d'impunité systémique : la violence numérique y opère librement comme « élément moral de l'agression et préparation du terrain à la violence physique », au service de la mise sous silence et de l'isolement des défenseures, transformant progressivement l'espace public syrien en un environnement hostile à la présence active des femmes et confisquant leur droit fondamental à la participation civique et politique.

06

Chapitre six

Recommandations politiques et opérationnelles

Les résultats de cette enquête indiquent que le ciblage numérique des défenseures syriennes n'est ni un simple « préjudice virtuel » ni un dysfonctionnement technique passager, mais un mécanisme de répression composite où se conjuguent le système masculiniste, la polarisation politique et les algorithmes des plateformes. Les « lacunes de protection et de réparation » (gouvernementales et techniques) étant systémiques et généralisées, le démantèlement de cette machine à faire taire exige une intervention à plusieurs niveaux, qui dépasse les réactions individuelles des victimes pour bâtir des filets de sécurité institutionnels, techniques et juridiques solides.

Sur la base des preuves documentées, l'association « Seen pour la paix civile » et le « Centre for Information Resilience (CIR) » formulent les recommandations urgentes suivantes :

6.1 Au législateur et aux autorités syriennes (démanteler l'application sélective)

- **Une qualification genrée de la violence numérique :** inscrire une définition juridique autonome et précise de la **« violence numérique fondée sur le genre et facilitée par la technologie » (TFGBV)** dans le cadre des amendements globaux à la loi sur la lutte contre la cybercriminalité et au code pénal, en incriminant expressément : la production et la diffusion d'hypertrucages non consentis (Deepfakes), les campagnes de diffamation et de raids dirigés (Brigading), et les menaces visant les défenseures et leurs familles ; et créer des mesures conservatoires d'urgence (telles que des « ordonnances de protection judiciaire d'urgence ») garantissant le blocage immédiat des contenus et la protection des espaces des victimes.
- **Suspendre les dispositions vagues :** suspendre immédiatement les articles vagues à caractère politique de la loi sur la cybercriminalité (notamment ceux relatifs à « l'affaiblissement du sentiment national » ou à « l'atteinte au prestige de l'État »), qui servent dans la pratique d'arme de dissuasion contre les défenseures et les voix critiques. Il

convient également de transférer les actions en « diffamation et injure » de la compétence pénale à la compétence civile, et de cesser d'utiliser la détention provisoire et les avis de recherche comme outils de sanction préventive.

- **Contenir « l'élément moral » de l'agression :** mettre fin résolument aux pratiques consistant à fermer les yeux sur les campagnes d'incitation numérique au nom de la « liberté d'expression », dès lors qu'elles atteignent le seuil de l'incitation à la haine et à la violence selon le test en six points du « Plan d'action de Rabat » des Nations unies ; et activer l'obligation pour le ministère public d'agir d'office contre les comptes, les personnalités publiques et les créateurs de contenu qui mènent l'incitation, afin d'empêcher que l'espace virtuel ne serve de justification préalable à des agressions physiques (fermeture de locaux d'organisations civiles, menaces d'élimination physique).

6.2 Aux grandes plateformes numériques (rétablir la diligence raisonnable)

- **Une évaluation des risques propre au contexte syrien :** que Meta et Telegram s'engagent à mener une évaluation approfondie et continue des risques de l'espace numérique syrien en tant qu'« environnement de conflit hautement fragile et divisé », conformément aux Principes directeurs des Nations unies relatifs aux entreprises et aux droits de l'homme ; et qu'elles reviennent sur les politiques de modération assouplies adoptées début 2025, qui ont laissé les femmes sans aucune protection face aux vagues de mobilisation mécanique et de comportement dirigé.
- **Localiser les algorithmes et actualiser les lexiques :** investir sans délai dans les capacités de modération automatisée et humaine des dialectes syriens et du langage genré codé. Les plateformes doivent ouvrir des canaux de coopération technique avec les unités de veille locales afin de recevoir régulièrement les lexiques sémantiques et les listes de termes nouveaux utilisés pour contourner la modération.
- **Des canaux de réponse rapide (Escalation Channels) :** créer des voies de signalement d'urgence fiables, réservées aux défenseures des droits humains et aux organisations de la société civile agréées, garantissant une réponse rapide (sous 24 heures) aux demandes de retrait de contenus menaçant la vie et de vidéos falsifiées, avec un mécanisme efficace empêchant les auteurs de recréer les pages fermées.

6.3 Aux organisations de la société civile et aux mouvements féministes (passer au soutien proactif)

- **Une stratégie de « soutien proactif » (Proactive Support) :** passer d'un modèle de soutien qui attend la demande de la victime (ce qui arrive rarement en raison de l'autocensure et des pressions familiales) à des unités d'intervention précoce qui repèrent les campagnes dès leur apparition et prennent contact de manière confidentielle et proactive avec les personnes ciblées pour leur proposer des protocoles de sécurité numérique et un soutien psychologique avant que le préjudice ne s'aggrave.
- **Activer « l'archivage par procuration » (Proxy Archiving) :** mettre en place des réseaux de défense des droits chargés de documenter et d'archiver les contenus injurieux, les commentaires et les menaces au nom de la personne ciblée. Cette démarche préserve les preuves pénales en vue de poursuites futures et épargne à la survivante le fardeau psychologique de la lecture directe des insultes et des menaces qui la visent, elle et sa famille.
- **Contentieux stratégique et transnational :** constituer une coalition d'avocats bénévoles pour suivre les affaires de « répression transnationale », et activer le principe de compétence universelle pour documenter et poursuivre les auteurs de menaces visant des défenseures résidant dans les pays d'asile, ainsi que pour documenter les parcours des plaintes au niveau local afin de mettre en lumière l'application sélective, preuves juridiques à l'appui.

6.4 Aux bailleurs de fonds et à la communauté internationale (financer la riposte)

- **Soutenir l'infrastructure d'enquête numérique :** orienter les financements vers les projets qui détiennent et développent des outils d'analyse et d'évaluation propres à l'espace numérique syrien, afin d'accompagner le passage de la documentation descriptive traditionnelle à des enquêtes numériques capables de mettre au jour les réseaux de « comportement dirigé » et de fournir des preuves d'attribution solides.
- **Relier la justice transitionnelle à l'espace numérique :** faire du niveau de protection des défenseures dans l'espace numérique et de l'arrêt de la mise sous silence systématique un critère essentiel et contraignant pour évaluer le respect, par les autorités locales, des normes relatives aux droits humains et des processus de consolidation de la paix, et intégrer ces indicateurs dans les rapports de la Commission d'enquête internationale indépendante sur la Syrie (COI) et du Mécanisme international, impartial et indépendant (IIIM).

Conclusion

Cette enquête montre que le ciblage numérique des défenseures syriennes ne saurait être réduit à un effet secondaire d'une liberté d'expression débridée : il s'agit d'une stratégie systématique d'exclusion politique qui redessine les frontières de l'espace public dans la Syrie de l'après-Assad. Cette dynamique de ciblage suit un enchaînement précis. Elle s'ouvre par la féminisation de la sanction — diffamation du corps, mise en cause de la vertu et pression sur l'entourage familial — et s'accompagne du retrait de la protection et du soutien social. Elle culmine ensuite dans la stigmatisation politique et l'assignation confessionnelle qui délivrent le « permis d'abuser », jusqu'au retrait contraint, à la fermeture de locaux ou aux menaces d'élimination physique. Ses effets destructeurs dépassent largement la victime directe : ils produisent un « effet dissuasif collectif » (Chilling Effect) qui décourage toute femme d'entrer dans la vie publique ou d'ouvrir des dossiers de redevabilité et de justice transitionnelle.

Les mots de la participante (P1) — « On nous fait taire… d'une manière ou d'une autre, on nous fait taire » — restent une véritable sonnette d'alarme face aux menaces qui pèsent sur la transition. À l'inverse, ceux de la participante Mme Fadwa Mahmoud (P3) — « Je continuerai à écrire et je continuerai à parler » — témoignent que la machine à faire taire n'est pas une fatalité. Entre cette intimidation et cette résistance se dessine la responsabilité commune et directe des forces de l'ordre et des autorités locales, des grandes plateformes numériques, des organisations de la société civile et des acteurs internationaux engagés dans la consolidation de la paix : veiller à ce que l'expression d'une opinion et la participation publique ne deviennent pas un coût existentiel payé au prix de la sécurité et de la dignité des femmes syriennes.

Annexe A

Méthodologie et cadre de l'enquête (Investigative Methodology)

Ce rapport adopte une approche d'enquête hybride combinant une veille numérique étendue, le traitement automatique des textes au moyen du modèle linguistique spécialisé (SeenBERT), une vérification humaine rigoureuse (Human-in-the-Loop), l'enquête en sources ouvertes (OSINT), des entretiens qualitatifs approfondis et une analyse juridique appliquée, afin de décortiquer les schémas de violence numérique fondée sur le genre et facilitée par la technologie (TFGBV) et de discours de haine visant les défenseures et les militantes syriennes depuis le 8 décembre 2024, et de mesurer l'effet dissuasif direct sur leur participation civique et politique.

Le rapport ne traite pas les données extraites comme un recensement exhaustif de toutes les violations, mais comme un échantillon significatif et statistiquement vérifié de l'espace public accessible à la veille. L'enquête adopte une distinction méthodologique stricte fondée sur les **« quatre niveaux de preuve »** (exposés au point 7), afin de distinguer la réaction sociale stéréotypée (mobilisation mécanique) du comportement dirigé, tout en s'abstenant d'attribuer une quelconque activité à une entité politique ou sécuritaire précise en l'absence de preuves d'attribution concordantes atteignant un niveau de preuve suffisant pour étayer une attribution dans un cadre judiciaire.

Pour atteindre ces objectifs, le cadre méthodologique a été conçu selon les 14 étapes et paramètres suivants :

1 Questions fondamentales de l'enquête

L'enquête vise à répondre de manière méthodique aux questions centrales suivantes :

1. Quels sont les marqueurs discursifs et sémantiques des abus fondés sur le genre, et comment se combinent-ils avec la stigmatisation politique et l'*assignation confessionnelle* pour disqualifier les défenseures ?
2. Comment les attaques numériques évoluent-elles, et quelle est la nature de la participation des comptes et des utilisateurs ?
3. Quels indicateurs comportementaux et techniques distinguent l'interaction mécanique du public (mobilisation mécanique) du comportement dirigé et des tactiques de manipulation algorithmique ?

4. Quels changements mesurables affectent le rythme et la nature de l'activité numérique des défenseures au cours des trois périodes d'attaque (avant, pendant et après la campagne) ?
5. Quels sont les effets du ciblage sur la sécurité personnelle et professionnelle des défenseures et sur leur participation publique ?
6. Quelles sont les lacunes structurelles du système législatif national et des politiques de diligence raisonnable des plateformes numériques qui empêchent une protection et une réparation effectives pour les victimes ?

2 Périmètre des données et définition de « l'incident » (Data Scoping & Incident Definition)

Pour constituer la base de données numérique, l'enquête s'est appuyée sur deux axes parallèles de collecte et d'analyse des contenus publiquement accessibles sur Facebook et Telegram :

- **Premier axe (veille en réseau) :** suivi des pages d'information, des chaînes et des groupes syriens publics où apparaissaient des récits ou des débats liés aux défenseures et aux militantes, en se concentrant sur les contenus comportant des indices de violence genrée, de diffamation, d'incitation ou de menace.
- **Second axe (veille inverse) :** veille approfondie des comptes et des pages publiques de plusieurs défenseures et militantes ciblées, considérés comme des points centraux permettant de suivre les attaques reçues et leur évolution, ainsi que les changements dans l'activité des personnes ciblées avant, pendant et après le ciblage.

Définition de l'incident

Aux fins du présent rapport, chaque commentaire ou publication injurieux n'est pas traité comme un incident distinct. Une distinction a été opérée entre l'élément isolé et l'incident (Incident).

L'incident se définit comme un ensemble de publications, de commentaires ou de contenus numériques liés au ciblage d'une défenseure ou d'une militante déterminée, apparaissant dans une fenêtre de temps rapprochée et partageant la cible, le récit utilisé, un point de départ ou un mode de diffusion, de sorte qu'ils peuvent être analysés comme une seule et même vague d'attaques. Les éléments ne sont pas regroupés en un incident au seul motif que leur contenu est similaire : le rapprochement s'appuie, selon les cas, sur un ensemble d'indicateurs comprenant la proximité temporelle, la similitude des messages, les republications, les points de départ communs, la séquence de diffusion et tout autre indicateur vérifiable.

3 Veille, classification automatisée et construction du lexique sémantique (SeenBERT)

Pour traiter le volume important de données, le modèle linguistique développé localement, SeenBERT, a été utilisé comme outil de tri initial, de détection de schémas et d'identification des contenus nécessitant une révision humaine approfondie.

La veille s'est appuyée sur un lexique sémantique genré propre au contexte syrien, développé et mis à jour pour repérer les expressions idiomatiques, les dialectes locaux, les mots codés et les formules utilisées pour contourner les systèmes de modération automatisée des contenus.

Les contenus initialement repérés par le système ont été classés selon cinq axes d'analyse principaux (les deux derniers étant fusionnés dans la figure 3) :

1. objectification et humiliation sexuelles ;
2. stigmatisation et diffamation morales et genrées ;
3. mépris genré, dénigrement et disqualification ;
4. intimidation et menaces physiques ;
5. atteintes visuelles et hypertrucages.

Cette classification est utilisée comme outil d'analyse de la veille, et non comme un jugement automatique définitif sur la nature du contenu ou sur l'existence d'une violation. Les contenus repérés par le système font l'objet d'une révision humaine avant d'être intégrés à la base de données finale. Pour limiter les risques d'erreur automatique, les cas ambigus — notamment l'ironie, le double sens, les expressions locales et les contextes religieux ou sociaux à double usage — ont été traités comme des cas nécessitant une vérification humaine et ne sont pas tranchés par la seule classification automatique.

4 Vérification humaine (Human-in-the-Loop)

Les résultats préliminaires de l'analyse automatisée ont été soumis à une révision humaine visant à :

- vérifier l'applicabilité des critères pénaux et le contexte de circulation ;
- écarter les faux positifs (False Positives) et distinguer les polémiques purement politiques du ciblage genré ;
- décoder le double sens, l'ironie et les expressions locales ambivalentes ;
- consolider la séparation entre les violations genrées probables et le groupe témoin de la polémique politique générale ;
- déterminer si des contenus différents relèvent d'un même incident ou d'incidents distincts.

Lorsque les éléments de preuve ne permettaient pas de déterminer la classification avec un degré de confiance suffisant, les cas ont été traités comme non concluants et n'ont pas servi de base à des conclusions catégoriques.

5 Critères de sélection des cas et conduite de l'enquête

L'enquête est passée de la veille à grande échelle à l'étude de cas déterminés selon un ensemble de critères : gravité du ciblage, ampleur de la diffusion, clarté des preuves numériques disponibles, existence d'indices de simultanéité ou de coordination de l'activité, possibilité de suivre l'évolution chronologique de l'incident et possibilité de mesurer l'évolution de l'activité de la personne ciblée. La sélection des cas ne visait pas à représenter statistiquement l'ensemble des défenseures ou de toutes les formes de violence numérique en Syrie, mais à étudier des cas à forte valeur explicative et probatoire permettant d'examiner les mécanismes du ciblage, sa diffusion et ses effets.

Les 12 défenseures ont été retenues selon les critères suivants : engagement public visible après décembre 2024, exposition à une campagne documentée, existence d'un compte public pouvant faire l'objet d'une veille, et diversité des origines géographiques et professionnelles.

6 Analyse des acteurs : distinguer l'interaction du public et le comportement dirigé

L'enquête a distingué deux niveaux de comportement numérique, en évitant de présumer une coordination du seul fait de la similitude des discours :

- **a. L'interaction désorganisée du public (mobilisation mécanique) :** participation d'un certain nombre d'utilisateurs à l'attaque ou reproduction d'un discours injurieux de manière convergente et quasi automatique, en réaction à des stimuli masculinistes profondément ancrés, sans preuves suffisantes d'une direction centrale ou d'une coordination préalable entre eux.
- **b. Le comportement dirigé :** existence d'un ensemble d'indicateurs pouvant révéler une coordination délibérée entre des comptes ou des entités dans le but d'amplifier artificiellement un contenu, un récit ou une activité : simultanéité inhabituelle, similitude répétée des contenus, schémas de republication en chaîne ou recours à des comptes liés de façon anormale. Aucun indicateur isolé n'est considéré comme une preuve suffisante d'une attaque coordonnée ; l'analyse repose sur l'accumulation d'indicateurs indépendants. Le rapport distingue en outre la preuve de la coordination comportementale de son attribution à une entité déterminée.

7 Niveaux de preuve et d'attribution (Attribution Confidence Levels)

Pour éviter de tirer des conclusions non étayées, l'enquête s'en est tenue à une échelle de preuve à quatre degrés :

1. **Niveau 1 (description du comportement) :** établir l'existence du contenu abusif et en déterminer la charge sémantique et genrée.
2. **Niveau 2 :** relever les preuves comportementales, temporelles et réticulaires révélant l'existence d'un comportement coordonné.
3. **Niveau 3 (liens documentés) :** établir le recoupement entre les comptes engagés dans la campagne et des réseaux ou des plateformes influentes identifiés.
4. **Niveau 4 (attribution formelle) :** relier la campagne à une entité sécuritaire ou politique précise, ce que le rapport n'a établi que dans les limites où des preuves matérielles concordantes et irréfutables étaient disponibles.

8 Mesure des effets sur l'activité numérique

Pour étudier les effets des campagnes de ciblage, l'activité des comptes publics des personnes ciblées a été analysée sur trois périodes (résultats présentés à la section 4.3) :

1. Période de référence : activité enregistrée au cours des deux semaines précédant le début de la campagne ;
2. Période de ciblage : période couvrant l'escalade et le pic de la vague d'attaques ;
3. **Période post-campagne :** activité enregistrée au cours des deux semaines suivant le reflux ou l'arrêt de la vague d'attaques.

La mesure a porté sur : le rythme de publication, le type de contenu, la part des contenus personnels par rapport aux contenus publics, les interruptions temporaires, la suppression de contenus, le passage à des paramètres de confidentialité plus restrictifs et l'évolution des thèmes abordés. Un changement d'activité n'est pas automatiquement interprété comme un effet causal direct, sauf s'il est étayé par des témoignages et des preuves qualitatives indépendants ; à défaut, il est présenté comme une corrélation temporelle.

9 Documentation et conservation des preuves numériques

Les preuves numériques ont été conservées selon les protocoles d'enquête en sources ouvertes (OSINT) : documentation des liens permanents, conservation des horodatages (Timestamps) en UTC, captures d'écran complètes, extraction des métadonnées (Metadata) et conservation de copies archivées afin de protéger les contenus contre toute modification ou suppression ultérieure par les comptes instigateurs.

10 Entretiens qualitatifs (Qualitative Interviews)

Six entretiens qualitatifs approfondis ont été menés avec des défenseures des droits humains, des militantes, des journalistes et des responsables d'organisations civiles et féministes syriennes :

- **(P1) :** défenseure des droits humains et actrice de la consolidation de la paix.
- **(P2) :** journaliste d'investigation et cofondatrice d'un média indépendant.
- **(P3) Fadwa Mahmoud :** militante politique et des droits humains engagée sur le dossier des disparitions forcées, résidant à l'étranger.
- **(P4) Thuraya Hijazi :** directrice d'une organisation féministe et de défense des droits.
- **(P5) Afraa Hashem :** défenseure des droits humains et militante civique.
- **(P6) :** militante indépendante des droits humains et militante politique.

Les entretiens visaient à mettre au jour les violations inaccessibles par les sources ouvertes : messages privés, tentatives de chantage, menaces physiques et censure par procuration exercée par le biais de la famille. Ils ont été guidés par le « Code Murad » (Murad Code) et le principe « ne pas nuire » (Do No Harm), avec l'exigence d'un consentement éclairé, le chiffrement des enregistrements et leur conservation sur des supports sécurisés et isolés. Les témoignages des participantes sont soumis au principe de triangulation (Triangulation) par recoupement avec des preuves numériques indépendantes.

11 Analyse juridique appliquée (Applied Legal Analysis)

L'analyse juridique a porté sur le système législatif syrien (loi n° 20 de 2022 sur la cybercriminalité et circulaires judiciaires publiées en 2026), le Pacte international relatif aux droits civils et politiques et les conventions internationales pertinentes. Elle s'est concentrée sur les lacunes d'application, l'application sélective des dispositions pénales, l'absence institutionnelle de définitions genrées, et l'évaluation du respect, par les grandes plateformes numériques, de leurs responsabilités de diligence raisonnable au regard des normes des Nations unies relatives aux entreprises et aux droits humains.

12 Critères d'inclusion et d'exclusion

Les contenus et les incidents ont été intégrés à la base de données finale lorsque les conditions suivantes étaient réunies : activité survenue dans la période étudiée, rattachement à un compte ou à un espace vérifiable, clarté de la cible et de l'objet du ciblage, et présence d'indicateurs suffisants de l'une des formes de violence couvertes. Les contenus dont la source, le contexte ou la date n'ont pu être vérifiés ont été exclus.

13 Limites de la recherche (Limitations)

- **Espaces fermés et contenus supprimés :** le rapport porte sur les contenus publics ; les correspondances privées et les groupes fermés n'apparaissent donc qu'à travers ce que les participantes ont documenté, et la suppression de comptes peut entraîner la perte d'une partie de la chronologie.
- **Biais d'observabilité :** les données disponibles portent surtout sur les militantes et les défenseures les plus visibles dans l'espace public, ce qui ne reflète pas nécessairement l'ampleur du ciblage subi par les militantes locales moins présentes sur les plateformes.
- **Limites de la classification automatique :** possibilité d'erreurs d'interprétation de SeenBERT concernant les dialectes, le double sens et l'ironie ; c'est pourquoi le modèle a toujours été complété par une révision humaine.
- **Limites de l'attribution :** éviter de confondre similitude linguistique et coordination dirigée, et ne pas attribuer les attaques à une entité précise sans preuves indépendantes.
- **Limites de la mesure des effets :** tenir compte du fait que l'évolution de l'activité numérique peut être liée à de multiples facteurs autres que la seule campagne en ligne.
- **Limites de l'échantillon qualitatif :** le volet qualitatif repose sur six entretiens raisonnés et vise l'explication et l'approfondissement, non la représentativité statistique.

14 Principe de vérification et de triangulation (Triangulation)

L'enquête s'est appuyée sur la triangulation des preuves en recoupant les sources numériques, l'analyse des séries chronologiques, les données de réseau, les entretiens et les sources ouvertes. Le rapport distingue ainsi précisément trois niveaux de conclusion :

1. ce qui peut être établi directement par les preuves numériques ;
2. ce qu'indiquent des indicateurs concordants avec un degré élevé de probabilité ;
3. ce qui demeure une hypothèse analytique non tranchée, du fait des limites des données disponibles.

Annexe B

Classification statistique des violations probables

Sur la base des opérations de contrôle et de vérification humaine rigoureuse (Human-in-the-Loop) de l'ensemble des **données archivées**, la matrice de classification statistique a été mise à jour selon la répartition analytique cumulée retenue pour l'ensemble des violations genrées probables, soit un total de **2 295 éléments comportant une violation probable**, regroupés en 29 incidents distincts (extraits de **7 793 éléments interactifs examinés** dans l'archive, face auxquels figurent **5 429 éléments constituant le groupe témoin de la polémique politique générale non genrée**, et 69 éléments jugés non concluants).

Le tableau suivant détaille les axes sémantiques, la définition opérationnelle de chaque axe, son volume en nombre et en pourcentage, et son unité de mesure :

Axe de violation / catégorie	Définition opérationnelle	Nombre d'éléments probables	Part du total (N = 2 295)	Méthode de comptage et unité de mesure
1. Stigmatisation morale et familiale	Atteintes à l'honneur et à la réputation, incitation contre la famille et pression par l'intermédiaire des proches (« il n'y a pas d'homme à la maison »), et accusations de trahison identitaires associées à l'assignation confessionnelle ou politique pour lever la protection sociale.	**1 128**	**49,2 %**	Élément (publication ou commentaire) comportant une violation genrée probable et vérifiée en contexte
2. Mépris genré, dénigrement et disqualification	Dépouiller la défenseure de sa compétence intellectuelle et politique, recours à des diminutifs dévalorisants, âgisme et remarques sur le physique, et cantonnement forcé du rôle des femmes à la dépendance domestique et aux stéréotypes.	**577**	**25,1 %**	Élément (publication ou commentaire) comportant une violation genrée probable et vérifiée en contexte

Axe de violation / catégorie	Définition opérationnelle	Nombre d'éléments probables	Part du total (N = 2 295)	Méthode de comptage et unité de mesure
3. Objectification sexuelle et atteintes à la vertu	Réduction directe au corps, commentaires grossiers sur l'apparence et les parties intimes, allusions pornographiques, et usage de termes diffamatoires et d'obscénités sexuelles explicites.	**403**	**17,6 %**	Élément (publication ou commentaire) comportant une violation genrée probable et vérifiée en contexte
4. Intimidation et menaces physiques	Menaces directes d'élimination physique, de meurtre ou de lynchage, divulgation des lieux de résidence et des déplacements (Doxxing), et signalements publics aux services de sécurité pour poursuivre la défenseure et sa famille.	**133**	**5,8 %**	Élément (publication ou commentaire) comportant une violation genrée probable et vérifiée en contexte
5. Atteintes visuelles et hypertrucages	Recadrages dégradants de photos personnelles, montages et dessins diffamatoires, et vidéos composites ou hypertrucages (Deepfakes) non consentis générés par intelligence artificielle.	**54**	**2,4 %**	Contenu visuel ou incident de falsification composite vérifié
Total général	**Bilan final des éléments comportant des violations genrées pleinement caractérisées**	**2 295**	**100 %**	**Extraits par l'examen humain de 7 793 éléments archivés et vérifiés**

Note de contrôle méthodologique

- Les deux derniers axes sont fusionnés dans certains graphiques de synthèse (comme la figure 3) sous l'intitulé général **« Intimidation, menaces physiques et atteintes visuelles », soit un total de 187 cas représentant 8,1 %** de l'échantillon total, afin de montrer le lien fonctionnel entre la falsification visuelle et la menace d'agression physique.
- Ont été exclues de ce décompte toutes les polémiques purement politiques et les critiques générales dépourvues de dimension genrée (soit 5 429 éléments), recensées comme

« groupe témoin » permanent pour établir la ligne de base et la comparaison statistique de l'interaction.

Références et sources

1. Rapports et études institutionnels antérieurs

- **Centre syrien pour les médias et la liberté d'expression (SCM) :** *enquête « Campagnes de dénigrement des femmes : la menace est-elle passée ? Le ciblage des militantes comme outil de répression et d'exclusion politique »*, par Tamam Saymoua et Lara Aizouqi, publiée dans le cadre d'un projet commun (Khatt 30), 27 janvier 2025.
- **Association « Seen pour la paix civile », en partenariat avec le « Centre for Information Resilience (CIR) » :** *rapport « Le discours de haine dans l'espace numérique syrien (2025) : schémas, significations et effets dans un environnement de conflit »*, département Documentation et études, mars 2026.
- **Association « Seen pour la paix civile » :** *base de données actualisée et archive numérique des violations relevant de la violence numérique fondée sur le genre et facilitée par la technologie (TFGBV) en Syrie*, département Documentation et archivage, 2025–2026.
- **Commission d'enquête internationale indépendante sur la République arabe syrienne (COI) :** *rapports périodiques présentés au Conseil des droits de l'homme des Nations unies sur les violations de la liberté d'expression, la restriction de l'espace civique et les attaques contre les défenseurs des droits humains* (2024–2026).

2. Cadres internationaux et instruments relatifs aux droits humains

- **Nations unies :** *Déclaration universelle des droits de l'homme*, article 19 (liberté d'opinion et d'expression) et article 12 (protection contre les immixtions arbitraires dans la vie privée).
- **Assemblée générale des Nations unies :** *Pacte international relatif aux droits civils et politiques* (ICCPR), notamment les articles 19 et 20 (interdiction de l'incitation à la haine, à la discrimination ou à la violence).
- **Principes du « Code Murad » (Murad Code) :** normes mondiales pour la collecte et la documentation des preuves relatives aux violences sexuelles et fondées sur le genre dans les conflits et les contextes de transition politique, dans le respect du principe « ne pas nuire » (Do No Harm).

3. Politiques et normes des plateformes numériques

- **Meta :** *Standards de la communauté (Community Standards)*, politiques actualisées relatives au discours de haine, au harcèlement et à l'intimidation des personnalités publiques et des militantes (versions 2025 et 2026).
- **Telegram :** *conditions d'utilisation et lignes directrices de modération des contenus*, relatives au traitement des chaînes publiques, au signalement des contenus injurieux et aux infractions liées aux menaces et à la diffamation.
- **Union européenne :** *règlement sur les services numériques (Digital Services Act – DSA)*, cadre de référence pour les obligations de diligence raisonnable et d'évaluation des risques des très grandes plateformes dans les environnements fortement polarisés.

4. Méthodologies techniques et analyse de réseaux

- **Modèles de traitement automatique du langage naturel (NLP) :** littérature scientifique de référence sur le développement et l'évaluation de modèles linguistiques spécialisés et leurs applications aux dialectes locaux et à l'analyse sémantique (Devlin et al., 2019 - BERT ; Antoun et al., 2020 - AraBERT ; critères de calibrage sémantique du modèle SeenBERT).
- **Analyse des réseaux sociaux (Social Network Analysis - SNA) :** méthodologies scientifiques utilisées pour suivre la circulation de l'information, repérer les indicateurs de comportement dirigé et analyser les métadonnées (Metadata) des commentaires et des interactions simultanées (François, 2019 - ABC Framework ; protocoles du DFRLab pour l'analyse du comportement dirigé).

Un musellement systématique

Le ciblage numérique des défenseures des droits humains en Syrie et les tentatives de les exclure après la chute du régime

CIR CENTRE for INFORMATION RESILIENCE ×